



Faculty of Computer Science, Electrical Engineering and Mathematics
Paderborn University
Warburger Str. 100
33098 Paderborn

Dissertation submitted in partial fulfillment of the
requirements for the degree of
Doctor of Natural Sciences
(Dr. rer. nat.)

Between How It Works and What It's For

An Analysis of Everyday Explanations of a
Technical Artifact

Lutz Terfloth

Course of Study:	Computing Education Research
Examiner:	Prof. Dr. Carsten Schulte Prof. Dr. Heike Buhl
Supervisor:	Prof. Dr. Carsten Schulte

April, 2026

Between How It Works and What It's For

An Analysis of Everyday Explanations of a Technical Artifact

Lutz Terfloth

The thesis is typeset with $\text{\LaTeX}$, based on the Scientific Thesis Template maintained by Oliver Kopp available at <https://github.com/latextemplates/scientific-thesis-template>, and adapted to fit stylistic needs.

The following digital tools were used in the preparation of this dissertation: LanguageTool (<https://languagetool.org>) and Claude (<https://claude.ai>) for spell checking, linguistic revision, and grammar checking. GitHub Copilot for assistance with analysis scripts. The author bears sole responsibility for all content.

Danksagung

DIESE Arbeit entstand im Rahmen meiner Tätigkeit als wissenschaftlicher Mitarbeiter in der Fachgruppe Didaktik der Informatik unter der Leitung von Prof. Dr. Carsten Schulte. Während meiner Zeit an der Universität Paderborn durfte ich viele tolle Menschen kennenlernen, die diese Arbeit geprägt haben und ohne die diese Dissertation nicht entstanden wäre. Einigen möchte ich persönlich danken.

Im akademischen Kontext gilt mein tiefster Dank meinem Doktorvater, Prof. Dr. Carsten Schulte. Seine Unterstützung während meiner Zeit in seiner Arbeitsgruppe und sein Beharren darauf, dass es wichtig ist, nicht nur zu verstehen, *was* ein Artefakt ist, sondern auch, *warum* es so ist, haben diese Arbeit und meine Sicht auf die Welt stärker geprägt, als ihm vielleicht bewusst ist. Über das Fachliche hinaus möchte ich ihm dafür danken, dass er stets der Gegenpol war, den ich benötigte: optimistisch, wenn mir der eigene Antrieb fehlte, und unbequem ehrlich, wenn Selbstzufriedenheit drohte. Diese Art der Begleitung war immer herzlich, selten bequem, und sie hat mich geprägt — fachlich wie persönlich.

Ebenso dankbar bin ich Prof. Dr. Heike Buhl, die mir durch gut strukturiertes und konstruktiv formuliertes Feedback stets dabei half, meine Gedanken in einen kohärenten Text zu fassen und dabei immerzu hilfreiche Kommentare zur statistischen Auswertung äußerte. Ihre Rolle als Zeitwächterin in den Meetings und ihr offenes Ohr im Projekt habe ich aufrichtig geschätzt — ebenso wie ihre gesunde Balance aus Rigorosität und Herzlichkeit.

Prof. Dr. Erik Barendsen danke ich für die vielen inspirierenden Gespräche über Forschung, Informatikdidaktik und die qualitative Methodik im Rahmen der stets bereichernden Doktorandenkolloquien. Mein erstes Doktorandenkolloquium in Nijmegen war ein besonderer Moment, der einen schönen Start in meine wissenschaftliche Karriere markierte — umso mehr freut es mich, dass er nun als Kommissionsmitglied auch Teil dieses Meilensteins ist. Erik, allow me to switch to English: Thank you for your openness, your thoughtful feedback, and for making me feel welcome in the research community from the very beginning.

Prof. Dr. Christian Scheideler danke ich für seine unkomplizierte Bereitschaft, Teil der Kommission zu sein, und für sein aufrichtiges Interesse an informatikdidaktischen Fragestellungen.

Dr. Harald Selke danke ich für seine Mitarbeit in der Kommission, vielmehr aber auch für die vielen guten Gespräche während der Mittagspause, die ich in Zukunft vermissen werde.

Danke an Prof. Dr. Friederike Kern, nicht nur für die offenen Diskussionen, als ich potenzielle Querverbindungen zwischen meiner Arbeit und linguistischer Theorie sah, sondern auch für das gute Feedback zu meinen Artikeln. Ich habe die konstruktive Zusammenarbeit und Unterstützung während meiner gesamten Arbeit im Projekt sehr genossen.

Vielen Dank, Dr. Felix Winkelkemper, nicht nur dafür, dass du (vor einer gefühlten Ewigkeit) ein hervorragender Betreuer meiner Bachelorarbeit warst, sondern auch dafür, dass du wichtige Grundlagen vermittelt hast, die mir dabei halfen, wissenschaftlich zu arbeiten. Als Kollege habe ich außerdem dein offenes Ohr sehr geschätzt, welches du in schwierigen Zeiten stets hattest. Für dein Feedback zum, meiner damaligen Meinung nach, finalen Entwurf der Dissertation bin ich dir sehr dankbar.

Vielen Dank, Dr. Lukas Höper, dass du mich nie zum Schweigen gebracht hast, obwohl du als Büronachbar reichlich Gelegenheit und Grund dazu gehabt hättest. Unsere Gespräche haben mir geholfen, teils diffuse Gedanken und komplexe Themen zu klären, meine eigenen Analysen besser einzuordnen und herausfordernde Phasen zu meistern. Ich bin froh, dich Freund nennen zu dürfen.

Vielen Dank an Vivien Lohmer und Michael Schaffer für eure Zusammenarbeit und Unterstützung während des gesamten Projekts. Vivien, dir gebührt mein besonderer Dank, dass die Transkripte des letzten Arbeitspakets rechtzeitig fertig wurden. Michael, danke für die zahlreichen Fragebögen, die auch für meine Dissertation relevant waren, die anregenden Diskussionen, sowie die Auswertung des Video-Recalls.

Einen besonderen Dank richte ich an Lars Hoferichter und Claire Elisabeth Roberts: Ihr wart die besten studentischen Hilfskräfte, die ich mir hätte wünschen können, und die Zusammenarbeit mit euch war mir eine wahre Freude.

Zuletzt möchte ich der Arbeitsgruppe der Informatikdidaktik der Universität danken. Meine Zeit bei Euch war abwechslungsreich und die guten Diskussionen werden mir fehlen. Dies ist weniger ein Abschied, und mehr ein „Bis bald“.

Meiner Frau Corinna schulde ich am meisten Dank. Du bist die wichtigste Konstante in meinem Leben. Deine Geduld, deine ausgeglichene Natur, deine Fürsorge, sowie die unerschütterliche Unterstützung während unserer gesamten Partnerschaft haben die vorliegende Arbeit erst möglich gemacht. Du warst mir eine wichtige Stütze, die mich in schweren Momenten aufrechtgehalten hat. Danke, dass du unserem Sohn eine so wundervolle, liebevolle Mutter bist. Ich freue mich auf weniger Zeit am Schreibtisch und mehr Zeit mit euch. Ich liebe euch beide mehr, als ich in Worte fassen kann.

An meine Eltern: Erst als ich selbst Vater wurde, habe ich verstanden, was es wirklich *bedeutet*, Elternteil zu sein, und wie viel ich darüber noch lernen muss. Umso mehr danke ich euch für eure Geduld, euer Vertrauen und dafür, dass ihr die Hoffnung in mich nie verloren habt, auch wenn ich euch manchmal allen Grund dazu gegeben habe.

Abschließend möchte ich allen weiteren Familienmitgliedern und allen engen Freunden für die Unterstützung während des gesamten Prozesses danken. Ein besonderer Dank gilt meinen Schwiegereltern, die nicht nur liebevolle Großeltern sind, sondern auch eine enorme Stütze. Ich bin froh, in eurer Nähe leben zu dürfen.

Rückblickend begann ich diese Reise voller Unsicherheit darüber, ob ich sie zu Ende bringen könnte. Ich beende sie in der Gewissheit, dass ich es konnte. Für diese persönliche Einsicht, bin ich dankbar und nehme sie neben allem Fachlichen mit, das ich auf diesem Weg bereits lernen durfte.

Abstract

TECHNICAL artifacts have a dual nature: they possess a physical structure that enables them to perform functions associated with human intentions. Philosophy of technology holds that adequate understanding of such artifacts requires grasping both of these sides and their connection, yet little empirical work has examined whether and how people actually navigate this dual nature when explaining artifacts to one another. This dissertation investigates everyday face-to-face explanations of a technical artifact, the board game Quarto, through four cumulative Building Blocks (BB1–BB4). The analytical lens is the Co-Constructive Dual Nature Framework developed herein, which distinguishes two perspectives on the artifact, Architecture (how it works) and Relevance (what it is for). Their interplay is examined along three criteria: a dual-perspective criterion concerning the need to address both perspectives, a collaborative criterion concerning the co-constructive dynamics between explainer and explainee, and a cognitive criterion concerning the epistemic modes of construal that underlie perspective selection. BB1 establishes intonation-phrase segmentation as a reliable transcription basis for the qualitative coding of utterances addressing the dual nature in explanation content. BB2 documents a robust Architecture-dominant pattern in naturalistic explanations, with explainers producing approximately 60–70% Architecture content and communicating it as a prerequisite for Relevance. BB3 shows that this baseline persists even under systematic co-constructive pressure from explainees signalling, for example, Relevance interest, though partial adaptation occurs. BB4 experimentally tests the effect of explanatory approach on recipients' understanding: Architecture-focused and Relevance-focused explanations produce equivalent comprehension, but only integrated explanations that connect both yield significantly higher enabledness with the artifact. The central finding is a tension between explanatory default and explanatory need: the approach that yields enabledness and thus produces the deepest understanding demands the integration of both perspectives. In the naturalistic explanations (BB2), this integration was not reliably produced. This tension carries consequences for computing and technology education, where curricula that separate Architecture from Relevance risk failing to sufficiently foster integrative understanding, and thereby risk falling short of educational aims such as self-determination. Explainable AI is likewise affected: when, for instance, feature importances are used to generate counterfactual explanations, these may support comprehension but not the agency that people who interact with opaque but consequential systems require.

Kurzfassung

TECHNISCHE Artefakte besitzen eine duale Natur: Sie haben sowohl eine physische Struktur, als erfüllen sie auch Funktionen im Dienst menschlicher Absichten. Die Technikphilosophie argumentiert, dass ein angemessenes Verständnis solcher Artefakte die Erfassung dieser beiden Seiten und deren Verbindung erfordert. Bislang fehlt es an empirischer Forschung dazu, ob und wie Menschen dies in der Tat bewältigen, wenn sie einander Artefakte erklären. Die vorliegende Dissertation untersucht alltägliche Face-to-face-Erklärungen eines technischen Artefakts, des Brettspiels Quarto, anhand von vier kumulativen Building Blocks (BB₁–BB₄). Den analytischen Rahmen bildet das im Rahmentext entwickelte Co-Constructive Dual Nature Framework, das zwei Perspektiven auf das Artefakt unterscheidet: Architecture (wie es funktioniert) und Relevance (wozu es dient). Ihre Auswirkungen auf den Erklärungsprozess werden entlang dreier Kriterien theoretisch beschrieben: das Kriterium der dualen Perspektive untermauert die Notwendigkeit, beide Perspektiven zu adressieren; das kollaborative Kriterium untermauert die ko-konstruktiven Dynamiken zwischen Erklärendem und Adressaten; und das kognitive Kriterium untermauert, welche epistemischen Haltungen der Perspektivenwahl zugrunde liegen. BB₁ etabliert die Segmentierung nach Intensionsphrasen als reliable Transkriptionsgrundlage für die qualitative Kodierung von Äußerungen, die die duale Natur im Erklärungsinhalt adressieren. BB₂ dokumentiert ein robustes Architecture-dominantes Muster in naturalistischen Erklärungen, bei dem Erklärende etwa 60–70% Architecture-Inhalte produzieren und diese als Voraussetzung für Relevance auslegen. BB₃ zeigt, dass dieses Muster auch unter systematischem ko-konstruktivem Einfluss von Explainees bestehen bleibt, die etwa Relevance-Interesse signalisieren, jedoch findet teilweise eine Anpassung statt. BB₄ prüft letztlich den Effekt des Erkläransatzes auf das Verständnis der Rezipienten: Architecture-fokussierte und Relevance-fokussierte Erklärungen führen zu vergleichbarer Comprehension, doch nur integrierte Erklärungen, die beide Perspektiven verbinden, erzielen signifikant höhere Enabledness. Also genau jene Form des Verstehens, die in Einklang mit Comprehension dann auch kompetentes Handeln mit dem Artefakt ermöglicht. Der zentrale Befund ist eine Spannung zwischen erklärerischem Standard und erklärerischem Bedarf: Genau der Ansatz, der Enabledness erzeugt und damit das tiefste Verständnis ermöglicht, erfordert die Integration beider Perspektiven. In den naturalistischen Erklärungen (BB₂) ist dies aber nicht zuverlässig passiert. Diese Spannung hat Konsequenzen für die Informatik- und Technikdidaktik, in der Curricula, die Architecture von Relevance trennen, Gefahr laufen, das integrative Verständnis nicht hinreichend zu fördern und Bildungszielen wie Mündigkeit nicht gerecht zu werden. Aber auch der Bereich von Explainable AI ist betroffen, wenn beispielsweise Feature-Importances genutzt werden, um Erklärungen mit Counterfactuals

zu generieren. Diese unterstützen dann zwar die Comprehension, nicht aber die häufig so notwendige Handlungsfähigkeit der Personen, die mithilfe der Erklärung mit solchen intransparenten, aber folgenreichen Systemen interagieren.

Contents

Danksagung	5
1 Introduction	15
1.1 Three Challenges of Technical Artifact Explanation	17
1.2 Central Argument and Theoretical Innovation	19
1.3 The Broader Context	19
1.4 Research Approach and Motivation	20
1.5 Dissertation Structure	21
2 Foundations of Explaining Technical Artifacts	23
2.1 The Dual-Perspective Criterion	24
2.2 The Collaborative Criterion	35
2.3 The Cognitive Criterion	43
2.4 Synthesis: The Co-Constructive Dual Nature Framework	49
3 Empirical Contributions	53
3.1 Building Block 1: Methodological Foundation	57
3.2 Building Block 2: Dual Nature in Naturalistic Explanations	91
3.3 Building Block 3: Monitoring of and Adapting to Explanees	119
3.4 Building Block 4: Influence of Dual Nature Focus on Understanding	151
4 Interpretation and Discussion	187
4.1 The Dual-Perspective Criterion	188
4.2 The Collaborative Criterion	189
4.3 The Cognitive Criterion	192
4.4 Revisiting the Research Questions	193
4.5 Implications	195
4.6 Limitations	200
4.7 Closing	202
5 Conclusion	205
5.1 Summary of Findings	206
5.2 Theoretical Contributions	207
5.3 Directions for Future Research	208
5.4 Broader Implications	212
5.5 Closing Remarks	214

Contents

6 Publications	215
Bibliography	217
 Appendix	 233
A Data and Code Availability	233
B Coding Manual for Content Coding	235
C Transcript Examples	237
C.1 Naturalistic Setting — BB2 (WP2-P2-VP05)	238
C.2 A-Condition — BB4 (WP4-P2-VP31)	241
C.3 R-Condition — BB4 (WP4-P2-VP64)	245
C.4 AR-Condition — BB4 (WP4-P2-VP79)	257
D Understanding Instrument (Building Block 4)	265
D.1 Structured Interview Guide (Pilot, 2022)	265
D.2 Questionnaire Version 3 — Final Version Used in Study	268
E Supplementary Material by Building Block	275
E.1 BB1 — Transcription in Computing Education Research	275
E.2 BB2 — Adding Why to What	275
E.3 BB3 — Navigating the Dual Nature	279
E.4 BB4 — Bridging the Dual Nature	281
F Exploratory Analyses	291

1 Introduction

“Somewhere in his writings—and I regret having forgotten where—John von Neumann draws attention to what seemed to him a contrast. He remarked that for simple mechanisms, it is often easier to describe how they work than what they do, while for more complicated mechanisms, it is usually the other way around.”

Edsger W. Dijkstra in *Selected Writings on Computing: A personal Perspective* (1982, p. 212f)

ANYONE who has tried to explain a complex technology to someone knows the feeling: not knowing where to start, struggling with how much detail is appropriate, and wondering what the other person actually needs and wants to know. Explaining is one of the most common activities in everyday life: parents explain games to their children, computer science teachers explain algorithms to students, and engineers explain their inventions to novices. Yet despite the ubiquitous nature of explanations, explaining is surprisingly challenging in face-to-face settings (cf. Keil, 2006; Rohlfing et al., 2021). This difficulty is not simply a matter of expertise. The difficulty, I will show, reflects a fundamental challenge that arises whenever the object of explanation is a human-designed artifact.

We are surrounded by such artifacts—smartphones, algorithms, vehicles, household appliances—and we use them so effortlessly that we rarely think about how they work. Research on the “illusion of explanatory depth” shows that people systematically overestimate how well they understand everyday devices: when asked to actually explain how a zipper or a flush toilet works, they discover that their understanding is far shallower than they assumed (cf. Keil, 2006; Rozenblit and Keil, 2002). The difficulty goes beyond incomplete knowledge. When pressed to explain an artifact to another person, we face another challenge: should one describe *how* the artifact works, its mechanisms, components, and internal logic. Or should one focus on *what it does* and *why it matters*, its purpose, its effects, and its role in the lives of those who use it (cf. Kroes, 1998)?

Consider a coffee grinder as an example. One might explain it as a machine for quickly turning coffee beans into ground coffee; addressing what the artifact is for, its *Relevance* (Schulte, 2008b; Winkelkemper et al., 2024). Explaining how it works internally, however, requires a different kind of narrative: turning a lever activates gears that rotate grinding stones, which crush beans that fall through a chamber

by gravity; this addresses its *Architecture*¹ (Schulte, 2008b; Winkelnkemper et al., 2024). The challenge for any explainer, besides having understood the artifact, lies in navigating between these two perspectives by showing how the Architecture enables the intended Relevance and vice versa (cf. Kroes, 2006).

This tension is not unwarranted. It reflects something more fundamental about the nature of human-designed objects. Technical philosophy states technical artifacts² have a “dual nature” (Kroes, 1998, 2010). Technical artifacts are simultaneously physical objects that adhere to natural laws (i.e., have an Architecture) and carriers of intentions and purposes (i.e., carry Relevance). A smartphone, for instance, can be described through its silicon circuits and underlying electrical mechanisms, or through its role in enabling human communication. One may argue that technical artifacts, similar to other products created by humans, carry the “derived” intentionality of “their creators” (Rey, 2025). Some argue that, to really understand an artifact thoroughly, neither the structural (Architecture) nor the functional (Relevance) description alone captures what such an object fundamentally is (Kroes, 2010; de Ridder, 2007; Schulte, 2008b; Winkelnkemper et al., 2024).

This dual nature creates specific challenges for explanation (Kroes, 2006, 2010; de Ridder, 2006), as a comprehensive description of an artifact needs to address both perspectives while also making their connection understandable (Kroes, 2006; Vermaas and Houkes, 2006). Kroes (2006) describes a “logical gap” between structural (Architecture) and functional descriptions (Relevance) which “resists deductive closure”, highlighting the difficulty facing anyone who explains artifacts: the two perspectives cannot be derived from one another, but require deliberate bridging work to render the artifact intelligible. Yet we know surprisingly little about how and if people actually handle this task in practice. Do explainers—the person giving the explanation—intuitively address both dimensions? Do they adapt their emphasis depending on the situation or explanation goal? And does it matter for the understanding of the explainee—the person receiving the explanation—which perspective receives more attention? Ultimately, the question is whether the theoretical imperative to explain and connect both dimensions of the dual nature actually manifests itself in naturalistic explanatory behavior.

¹From hereon I will use Architecture and Relevance in their capitalized form as working terms for these two perspectives. Their theoretical foundation is developed in Chapter 2. Where I draw on sources that use different terminology, I will add (Architecture) or (Relevance) after the original terms.

²Henceforth I will use the terms technical artifacts and artifacts interchangeably.

1.1 Three Challenges of Technical Artifact Explanation

Given the dual nature of technical artifacts, it is crucial to think about what it means to fully understand one during an explanation. In the dual nature theory literature, one argument is that designers possess such understanding precisely because they can bridge the gap between the two sides. They grasp both how the artifact works (Architecture) and what it is for (Relevance), and crucially, how these connect (Kroes, 1998, 2006; de Ridder, 2006). Understanding thus entails considering both perspectives together rather than separately, a requirement that creates a fundamental tension in everyday explanatory practice. Responding to three fundamental challenges is necessary to clarify the manner in which this tension presents itself.

The Collaborative Challenge: How do explainers and explainees work together to construct understanding? When one explains version control to someone, the explanation might emphasize document versioning for an author of a book, branching workflows for a developer, or collaboration features for a project manager. The concept of co-constructive explanation (Reusser and Pauli, 2015; Rohlfing et al., 2021) provides useful concepts for describing how such adaptivity works in practice. Explaining requires explainers to continuously *monitor* which aspects the explainee currently understands and what is still unclear (Chi et al., 2004; Clark and Brennan, 1991), to provide adaptive *scaffolding* that helps explainees connect new information to what they already know (Mercer, 2002; Wood et al., 1976), and to establish and maintain *common ground* (e.g., shared knowledge, assumptions, and vocabulary) through ongoing interaction (Clark and Brennan, 1991). This interplay becomes particularly demanding when the object of explanation is a technical artifact, as explainer and explainee must build shared understanding not just about individual components but about how those components relate to one another, too.

The Cognitive Challenge: How do people shift between different modes of construal during explanations? Research in cognitive science suggests that people do not follow a universal approach to making sense of things. Instead, people adopt different epistemic stances, distinct ways of reasoning about an object, depending on what they are trying to grasp (Dennett, 1987; Keil, 1994). When explaining a touchscreen of a smartphone, for instance, one might adopt a mechanistic mode of construal to describe how capacitive sensing detects electrical changes, or a functional mode of construal to describe how the screen enables user input. These are not merely stylistic choices; they reflect different cognitive orientations toward the object being explained (Keil, 1994; Lombrozo and Wilkenfeld, 2019). People appear to produce qualitatively different utterances depending on which stance they adopt (Lombrozo and Carey, 2006), and coordinating between stances may require considerable cognitive effort (Kelemen et al., 2013). For technical artifacts,

this coordination becomes particularly consequential, as their dual nature means that both mechanistic reasoning about the Architecture and teleological reasoning about the Relevance may need to be maintained and connected throughout the explanation.

The Dual-Nature Challenge: What makes technical artifacts distinctive as objects of explanation? The issue most central to this dissertation underpins both the collaborative and the cognitive challenge in important ways: technical artifacts have a dual nature that resists reduction (Kroes, 1998, 2006; Vermaas et al., 2011). Unlike natural phenomena that can be explained by reference to physical laws (e.g., rain), or purely conceptual entities that exist in human minds (e.g., friendship), explaining technical artifacts means addressing their Architecture and their Relevance. This creates specific challenges for explanation content: explainers have to address each perspective separately and demonstrate how Architecture enables Relevance; how, for example, functional requirements (e.g., secure, encrypted messaging) constrain architectural choices (e.g., end-to-end encryption protocols). The content of an explanation thus becomes the site where the ontological complexity of artifacts plays out through the practical work of addressing and bridging both sides.

Addressing these challenges demands an interdisciplinary theoretical foundation, which Chapter 2 develops in detail. Each of the three challenges are treated as an analytical criterion that an adequate account of technical artifact explanation should satisfy. The dual nature of technical artifacts, when they serve as objects of explanation, necessitates grounding in the philosophy of technology, extended here by examining how dual nature theory has been applied in computing education research. The interactive nature of everyday explanations requires drawing on linguistic and conversation-analytic theory to understand how meaning is created jointly through dialogue. Finally, since explanations involve cognitive sense-making processes that shape how explainees reason about artifacts, theories from cognitive psychology provide important insights into how people adopt different modes of construal (Dennett, 1987; Keil, 2006; Lombrozo and Wilkenfeld, 2019).

Existing research has largely treated these three disciplinary perspectives in isolation. Computing education research has identified, for example, a tendency for programming instruction to foreground technical and structural aspects of programs while underrepresenting the broader dimensions of their nature, including their role as tools, their embeddedness in social contexts, and the values they encode (Lonati et al., 2022). Cognitive psychology has examined modes of construal in experimental settings (Kelemen, 1999b; Lombrozo and Gwynne, 2014), but rather disregards how people naturally coordinate different stances during the explanation of artifacts. Conversation analysis has elaborated on a notion of explanation as a social practice in which understanding is iteratively co-constructed by the participants (Miyake, 1986; Quasthoff et al., 2017; Rohlfing et al., 2021),

but has not specifically addressed the demand that technical artifacts place on this process. What remains unexplored is how explainers and explainees coherently navigate the dual nature tension—jointly navigating both Architecture and Relevance—when they co-construct understanding in everyday explanations of artifacts. This dissertation addresses precisely that concern.

1.2 Central Argument and Theoretical Innovation

This dissertation advances a fundamental argument: a comprehensive explanation of a technical artifact requires addressing its Architecture (how it works), its Relevance (what it is for and why it matters), and the integration of the two. Throughout the dissertation, *understanding* is used in the sense developed by Buschmeier et al. (2025), who distinguish *comprehension* (knowing-that) from *enabledness* (knowing-how); the term covers both dimensions unless one is named specifically. I assume that in everyday explanations of technical artifacts, the navigation of their dual nature leads to identifiable patterns, some of which are more effective at achieving understanding than others. If such patterns can be identified and understood, the findings can inform both educational practice and theory (cf. e.g., Winkelkemper et al., 2024), as well as the design of explanation agents and explainable AI (XAI) (cf. e.g., El-Assady et al., 2019; Sokol and Flach, 2020). The theoretical contribution of this dissertation is an integrated framework that brings together three research areas: adaptivity in explanation processes, dual nature artifact ontology, and cognitive modes of construal, to capture the distinctive challenges of technical artifact explanation. The empirical contribution is a set of studies demonstrating that people systematically address both sides of the dual nature and that the way they do so has consequences that affect understanding.

1.3 The Broader Context

The importance of understanding technical artifact explanation has grown with the complexity of the artifacts themselves. Modern technologies often function as “black boxes” (Pasquale, 2016), purposefully designed in a way that we can use them without comprehending their internal architecture. This opacity results in genuine educational and social challenges (cf. e.g., Pasquale, 2016): How do we help people develop meaningful technological literacy when systems become increasingly complex and opaque yet easy to use? How do we ensure that explanations of AI systems, algorithmic decision-making, and digital platforms serve genuine understanding rather than creating mere impressions of comprehension or unhealthy trust (Peters and Visser, 2023)?

Current challenges in XAI research illustrate this problem in a particularly consequential domain. Knowing that pixel [247, 189] contributed 0.73 to a classification score provides architectural insights without providing any Relevance detail. Saying an AI focused on the patient’s age and symptoms provides Relevance without architectural grounding, potentially creating false impressions of understanding or *unhealthy trust* (cf. Peters and Visser, 2023). The dual nature framework suggests that neither approach alone can yield genuine understanding and that the core difficulty lies in connecting how computational structures enable decision-making purposes. This is structurally the same challenge the framework identifies for any technical artifact: two logically disconnected perspectives that must be bridged for understanding to emerge. The “black box” problem in AI (Castelvecchi, 2016) may thus be interpreted as a dual nature problem. What explaining *actually* means in the context of XAI remains an open question (Miller, 2019; Rohlfing et al., 2021; Sokol and Flach, 2020), and understanding how humans naturally navigate dual nature tensions in simpler contexts may help illuminate this debate, too.

A similar challenge exists in education. Disciplines such as technology education and computing education have developed various literacy, competency, and awareness frameworks (see e.g., Höper and Schulte, 2023; Jia et al., 2025; Wing, 2008) aimed at preparing learners for an increasingly interconnected world (cf. Boix Mansilla and Schleicher, 2022). However, underlying these frameworks’ overarching goals is a fundamental problem: teachers themselves, regardless of experience, appear to struggle to connect the two sides of the dual nature (Frederik et al., 2011). This suggests that for designing better curricula or better XAI interfaces, we need a clearer empirical picture of how the dual nature is actually navigated in practice, especially if the claim that both sides need integration leads to expected effects on understanding.

1.4 Research Approach and Motivation

This dissertation investigates the three challenges outlined above by examining how everyday explanations of a technical artifact unfold in practice. Rather than normatively prescribing how artifacts should be explained, it empirically investigates how people navigate the dual nature tension and how this affects collaborative and cognitive dynamics, and understanding outcome. The empirical domain chosen for this purpose is a proxy artifact, the board game *Quarto*³, explained by one person to another in a controlled but naturalistic setting.

The selection of the proxy artifact Quarto as an explanandum, the object of explanation, represents a deliberate methodological choice. First, it addresses a persistent dilemma in explanation research: achieving experimental control while preserv-

³Henceforth I will just use Quarto, without italics and exclamation mark, to refer to the game.

ing naturalistic interaction (Kern et al., 2024). Board games are well-suited for this purpose for several reasons. Games exhibit clear dual nature characteristics at accessible complexity: they possess material components and rule structures (Architecture) while creating purposeful experiences and strategic interactions (Relevance), without requiring specialized technical knowledge that might confound explanatory patterns. A deliberate absence of physical game components in the experimental setting creates a controlled black box scenario that mirrors real-world explanation contexts where artifacts are not available for direct manipulation, encouraging explicit verbalization rather than demonstration. Games explanations also represent a familiar social practice and feel natural, reducing experimental artificiality while enabling systematic comparison across dyads. Finally, as human-designed artifacts with clear intentional purposes, board games provide a tractable domain for investigating dual nature integration principles. Since all technical artifacts inherently manifest this dual nature, insights gained from this controlled domain can be meaningful for new hypotheses for other, more complex artifacts.

This dissertation addresses four overarching research questions:

- **RQ1 – Dual Nature Expression:** How can the two perspectives on the dual nature be systematically distinguished through qualitative content analysis of explanations?
- **RQ2 – Explanatory Patterns:** What patterns emerge in how explainers sequence and integrate Architecture and Relevance perspectives across the course of an explanation?
- **RQ3 – Collaborative Navigation:** How do collaborative processes shape the navigation between dual nature perspectives during everyday explanations of technical artifacts?
- **RQ4 – Effect on Understanding:** What effect does the explanatory approach of describing the dual nature have on the comprehension and enabledness dimensions of explainee understanding?

1.5 Dissertation Structure

The dissertation is organized as follows. Chapter 2 develops the theoretical background by examining three complementary perspectives, each derived from one of the challenges outlined above: the dual-nature dimension through philosophy of technology and computing education research, the collaborative dimension through co-construction theory, and the cognitive dimension through research on explanatory stances. Finally, the chapter synthesizes these foundations into

the Co-Constructive Dual Nature Framework, an integrated analytical model that captures how collaborative processes, dual-perspective navigation, and cognitive stance flexibility operate together in explanatory practice.

Chapter 3 presents the empirical approach and reports the dissertation's four cumulative contributions (henceforth called building blocks, or short BB): (BB₁) is a methodological review establishing a theoretical foundation and an approach to transcription as an analytical tool, (BB₂) is a naturalistic study of Quarto explanations, (BB₃) is a confederate study investigating explainer adaptation to explainee interests, and (BB₄) an experiment examining the effect of explainers' explanatory approach on explainees' understanding. As these building blocks build on one another, each study inherits the analytical tools and empirical baseline of its predecessors.

Chapter 4 interprets the findings, reconnects them to the research questions, relates them to the broader challenges of technical artifact explanation, discusses implications for educational practice and XAI development, and examines the methodological, theoretical, and empirical boundaries of this research.

Chapter 5 concludes the dissertation by summarizing its central findings and outlining directions for future research.

2 Foundations of Explaining Technical Artifacts

“Humans are driven to acquire and provide explanations. Within months of uttering their first words, children ask ‘why.’ Preverbal infants explore phenomena that puzzle them in an attempt to uncover an explanation of why an effect occurred. As adults, we must frequently choose between explanations of why politicians lost, why the economy is failing, or why a war is not winnable. [...] Explanations are therefore ubiquitous and diverse in nature.”

Frank C. Keil in *Explanation and Understanding* (2006, p. 228)

As Keil observes, explanations are ubiquitous and diverse. Not all explananda are alike, and different explananda call for different kinds of reasoning.¹ When we explain why rain falls, we draw on causal reasoning (Hempel and Oppenheim, 1948). When we explain why people behave in certain ways, we draw on intentional reasoning (Dennett, 1987). Technical artifacts are neither purely physical phenomena governed by natural laws nor purely intentional entities. They are deliberately engineered objects that embody human purposes within physical structures, creating what philosophers of technology term their “dual nature” (Kroes, 1998). This has consequences not only for how one explains artifacts, but also for what it means to understand them. Throughout this dissertation, *understanding* is used in the sense developed by Buschmeier et al. (2025), who distinguish two dimensions: *comprehension*, referring to the capacity to reproduce or recognize how an artifact works and what it is for (knowing-that), and *enabledness*, the capacity to act competently with and upon the artifact (knowing-how), and both need to be pronounced for someone to have agency. The distinction matters for this chapter because the two dimensions may not stand in the same relation to the two sides of the dual nature, a question the framework developed here is designed to make tractable and that Chapter 3 and Chapter 4 return to empirically.

¹This dissertation uses standard terminology from philosophy of science: the *explanandum* (plural: *explananda*) is the object or phenomenon to be explained; the *explanans* is the content offered in explanation of it (cf. Hempel and Oppenheim, 1948).

From a linguistic perspective, everyday explanations are not unified but vary systematically in their epistemic targets: Klein (2008) distinguishes How-, Why-, and What-explanations, noting that everyday explanations draw on all three types rather than privileging any single one. Following Fisher et al. (2023), who apply this typology to game explanations, such explanations can be classified as procedural (Kotthoff, 2009), aligning them primarily with Klein’s How-type. Yet precisely because the artifact has a dual nature, explaining it may require foregrounding different characteristics of the artifact depending on the explanatory moment, due to different epistemic targets.

This chapter develops the theoretical foundations needed to address the three challenges contextualized in Chapter 1, treating each challenge as an analytical criterion. Section 2.1 draws on dual nature theory (Kroes, 2010) to capture what makes artifacts distinctive as objects of explanation, and establishes the distinction between Architecture and Relevance as the central conceptual vocabulary of this dissertation. Section 2.2 engages co-constructive explanation theory to understand how everyday explanation functions as an adaptive social practice rather than information transfer, and Section 2.3 draws on research into epistemic modes of construal that people use when making sense of artifacts. The idea across these sections is that compatible theories recur across a wide range of research traditions in their own terminology. Tracing these similarities is not meant to suggest that these traditions say the same thing, but to demonstrate that the dual nature vocabulary offers a productive way to establish connections between theories and thus broaden empirical foundations. Where such analogies are drawn, the text first develops each tradition’s concepts in their original terms before mapping them onto the dual-nature distinction. Section 2.4 then synthesizes these three perspectives into the Co-Constructive Dual Nature Framework, an integrated analytical model that provides the theoretical infrastructure for the empirical investigations reported in the later chapters.

2.1 The Dual-Perspective Criterion

WHAT ONTOLOGICAL FEATURES MAKE TECHNICAL ARTIFACTS DISTINCTIVE AS OBJECTS OF EXPLANATION?

Technical artifacts are constitutive of human existence: “without artifacts, there would be no recognizable human life” (Baker, 2004). From the rudimentary hand truck designed to alleviate manual labor to complex cellular networks underpinning global communication, throughout human history, people have engineered artifacts to fulfill specific purposes. Unlike natural objects such as trees or mountains, artifacts are infused with purpose by their creators. The dual nature theory

(Kroes and Meijers, 2002) posits that this purposiveness gives artifacts an irreducibly twofold character: both tangible *physical structure* and *function shaped by human intent*. In explanation theory, the explanandum constrains the explanans, “the way in which the explainer conceptualizes the object of explanation” (Rohlfing et al., 2021), so the dual nature of artifacts shapes what can and should be said when explaining them. Neither material composition nor functional purpose alone suffices; understanding an artifact requires grasping both aspects and their connection. The remainder of this section develops this claim theoretically, illustrates through a historical example, and examines against empirical evidence.

Kroes (2010) develops this point by showing that each perspective alone fails in a characteristic way. A purely physical description (of the Architecture) may account for the causal processes through which an artifact operates, but without reference to function (its Relevance) the object loses its status as a technical artifact altogether. Conversely, a purely functional description may account for what the artifact is for and how it serves human ends, but it “effectively black boxes its physical structure” (Kroes, 2010, p. 54). The two types of description are not merely different in emphasis but logically decoupled: Relevance claims cannot be deduced from Architecture descriptions, nor the other way around, creating what Kroes (2006) terms a “logical gap” that requires non-deductive inference patterns to bridge. The dual nature thus manifests in two distinct yet complementary explanatory approaches, each irreducible to the other.

But the width and breadth of the definition of the term function in dual nature program’s conception has been criticized as too narrow. Vaesen (2011) argues the program suffers from a functional bias: an artifact’s functional nature is not exhausted by what it is designed to perform, since many non-functional properties, such as marketability, ease of manufacture, or aesthetic qualities, equally reflect the intentions of designers and users. This point becomes even more pressing for complex digital artifacts. A messaging application does more than transmit text, it reshapes expectations about communication, alters social norms around availability (Winkelkemper et al., 2024), and raises questions about privacy and data ownership.

The functional side of an artifact, in other words, extends well beyond its designed functions for end users. It encompasses the goals, values, and societal structures associated with its use. Within computing education, Budde (2021) developed this further by distinguishing between an artifact’s technical features (its designed functionality in an idealized context) and its secondary functions (the individual and social meanings that emerge through use), drawing on Engbring’s (2003) distinction between artifacts, cognifacts, and sociofacts². This differentiation demonstrated that the functional side of digital artifacts resists reduction to designed function and requires analytical categories that capture the subjective

²in German: Artefakte, Soziefakte, Kognifakte

and socially situated dimensions of artifact meaning. Recognizing this, Schulte and Budde (2018) reframed the functional side as an *interpretive perspective*, distinguishing it from the *analytical perspective* on the artifact's structure and emphasizing that this broader view inherently involves subjective interpretation rather than objective description. They subsequently defined the term *Relevance* to capture this expanded scope, encompassing not only function but also purposes, values, and societal implications. It is this broader terminology and conceptualization that this dissertation follows. From hereon, I will adapt the terminology as well as the conceptualization as perspectives, following the work of (Budde, 2021; Winkelnkemper et al., 2024).

The Architecture Perspective (Winkelnkemper et al., 2024) describe the artifact's physical construction and internal workings. It encompasses material composition, structural relationships, causal mechanisms, and observable behaviors. As Kroes (1998) notes, such descriptions “make use of concepts from physical laws and theories and are free of any reference to the function of the object”. From this perspective, a smartphone is an assembly of semiconductors, circuits, batteries, and displays operating according to how they are programmed. For digital artifacts, this perspective must extend beyond hardware to include computational structures. Winkelnkemper et al. (2024) provide a definition:

“The **architecture** (or structure) perspective describes how an artefact is constructed physically, and how it works internally. An architectural description of an instant messenger, for example, encompasses hardware components like servers on the Internet, the internal data structures which constitute messages and contact lists, as well as formal structures like the algorithms for encryption and decryption of messages or for network communication. Although structures like these are often described in abstract form and are sometimes even declared to be non-physical, they are, of course, implemented physically within the artefacts and are therefore physically effective. The perspective of architecture is that perspective of explanation which engineers—including computer scientists—typically deal with extensively.”

This reformulation importantly acknowledges that Architecture for digital artifacts must include algorithms, data structures, and information flows, not merely hardware components. These abstract structures, while often described formally, are ultimately physically implemented and causally effective.

The Relevance Perspective (Winkelnkemper et al., 2024) address the complementary aspect: the artifact's connection to human purposes, designer intentions, and social contexts. It encompasses intended functions, purposes, values, and social meanings. It cannot be derived from physical description alone, as the

same physical structure can serve different functions in different contexts, and the same function can be realized through different physical structures. From this perspective, a smartphone transforms from a collection of components into a communication device, information portal, and social connector. Winkelnkemper et al. (2024) define the perspective as follows:

“The **relevance** (or meaning, usage) of a digital artefact is that perspective of explanation covering the relevant aspects of public discourse regarding the artefact. It includes what in the duality program is called function but also extends to the goals associated with its usage, the purposes and values ascribed to it, as well as societal structures which have a relation to the artefact (such as laws, rules or associations). Much of what is subject to public discourse and therefore is covered in the relevance perspective, goes beyond what technicians and engineers typically talk about professionally. One aspect of the relevance considerations of instant messaging services, for example, would be the observation that their existence is changing people’s expectations regarding communication, so that today, among many people, it is expected that friends and acquaintances respond quickly to messages which have been sent to them.”

This expanded conception proves particularly crucial for contemporary technical artifacts like AI systems, whose societal impact extends far beyond their originally intended functions. To illustrate how the integration of both perspectives manifests in an epistemic sense, the next subsection examines an artifact deliberately chosen for its unfamiliarity.

2.1.1 Illustrative Example: A Mystery Artifact

To demonstrate how these perspectives integrate in practice, I examine an artifact deliberately chosen for its unfamiliarity to most readers. While this mechanical example predates the digital era, it clearly illustrates the fundamental dual nature that extends to contemporary digital systems. This unfamiliarity allows the reader to actively experience the epistemic process of encountering both perspectives and their connection when making sense of an unknown object. The artifact depicted in Figure 2.1 is described as follows:

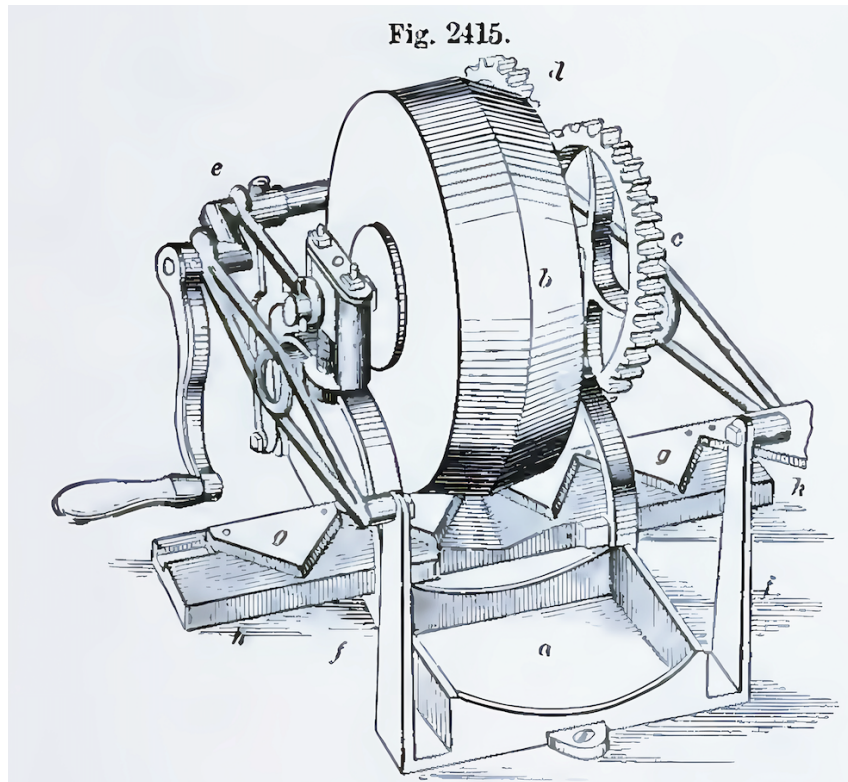


Figure 2.1: A Mystery Artifact (Knight, 1882)

“In Fig. 2415 a bed *a* supports a sliding carriage on which is the doubly beveled grindstone *b*, having a gear-wheel *c*, which engages with a wheel *d* on the shaft *e*. From the uprights *f* pivoted connections run to a crank at each end of the driving-shaft, and from each end of this shaft similar connections are made with the axis of the grindstone. The knives *g* are secured to the guide *h* fixed to the frame at a suitable angle. On turning the crank, the stone is caused to rotate as usual, and is reciprocated back and forth [...]” (Knight, 1882)

This description lies entirely within the Architecture perspective: it details the artifact as a physical object with specific components arranged in particular spatial relationships, and includes causal connections between them (e.g., “the stone is caused to rotate” and connections “run to a crank”). It can be objectively verified through direct observation.

However, I deliberately removed the first sentence of the original description, which identifies the artifact as “[a] machine adapted to the grinding of the section knives of harvesters, which are riveted to the knife-bar” (Knight, 1882). This sentence immediately places the artifact in relation to human intentionality and purpose, establishing its Relevance and connecting it to the domain of agriculture.

Such Relevance cannot, without additional effort, be derived from the physical description or the depiction alone (Kroes, 2006, cf. “logical gap”,).³ It requires reference to human intent and the context of use—the purpose of grinding harvester knives in agriculture—to be genuinely understandable. Through this reference, we can also connect the artifact to whatever pre-existing knowledge we have about its domain.

The original description further demonstrates this integration by explaining how the physical structure enables “insuring the even grinding of the teeth from root to point, and preserving the shape of the stone” (Knight, 1882). Here the Architecture and Relevance are explicitly connected: the specific arrangement of components is designed deliberately to achieve this intended function. The physical components work together (“On turning the crank, the stone is caused to rotate...”) to realize the intended purpose (“insuring the even grinding of the teeth”). A mere description of the Architecture would leave readers oblivious to what the artifact is for; a description of its Relevance alone would leave them oblivious to its inner workings, and without any imagination how it may look.

The harvester-cutter grinder thus exemplifies how the dual nature is not merely conceptual but ontological: the artifact’s very existence *as* such a device depends on both its Architecture and Relevance. This observation carries forward into the remainder of this dissertation: effective artifact explanation requires bridging between Architecture and Relevance rather than treating them in isolation, suggesting that explainers and explainees may need to navigate these relationships together.

2.1.2 Empirical Support

The dual perspectives are theoretical constructs. While systematic empirical validation of the dual nature theory’s claims remains limited, research in computing education provides some related evidence of how this duality creates practical challenges in learning contexts. This section applies the dual-nature lens to existing empirical research, as only a small set of studies address the dual nature theory directly, but ideas in various articles provide important context for its claims. They help to contextualize how learners navigate the dual nature when understanding technical artifacts.

³Designers begin conceptualizing an artifact from its intended Relevance before creating appropriate Architecture. Conversely, archaeologists finding unknown artifacts must deduce Relevance from physical Architecture (Perlman, 2004, p. 37), a process that succeeds only when sufficient contextual information about human needs and cultural practices is available. Both paths ultimately require grasping the connections between Architecture and Relevance.

The distinction between *mechanisms* and *explanations* by Soloway (1986) offers an early articulation of a related insight. Soloway defines mechanisms as the chains of actions through which an artifact operates, a concept that maps straightforwardly onto Architecture. His notion of explanation, however, resists simple alignment with Relevance. For Soloway, an explanation provides “a trail of how and why the artifact was designed the way it was” (Soloway, 1986, p. 855), necessarily encompassing both the mechanism itself and its design rationale. In this sense, Soloway’s “explanation” is itself an integrative concept: it requires connecting architectural structure to purposive reasoning. Soloway found that novices often grasp mechanisms without being able to provide such explanations, meaning they learn how code works without grasping why it was designed that way.

Similarly, Spohrer et al.’s (1985) research on debugging identified *plan* and *goal* mismatches, meaning students implement *plans* (the Architecture) that fail to realize their intended goals (the Relevance). This suggests that the process of designing an artifact, thereby creating its dual nature, creates specific challenges when learning programming, too. Learners must simultaneously maintain mental representations of both what they want to achieve and how their implementation achieves it, thus cross-relating both sides of the dual nature. But this connection is often not established correctly by learners (cf. Spohrer et al., 1985). Importantly, Spohrer et al.’s framework also implies that the relationship between goals and plans is not one-to-one: a given goal (Relevance) can be realized through multiple plans (Architecture), and a given plan may serve several subgoals. Budde (2021, p. 112ff) consolidates this point across multiple research traditions and argues that this non-bijective relationship is a defining characteristic of the dual nature itself. For the present dissertation, this is consequential: it is precisely because Architecture does not determine Relevance (nor vice versa) that their integration requires deliberate explanatory effort rather than arising intuitively from the presentation of either side alone (Kroes, 2006).

Source code is, in a sense, text, and research on how readers construct meaning from text has directly informed models of program comprehension. Kintsch’s (1991) Construction-Integration model distinguishes between the *text base* and the *situation model*. The *text base* is a representation of the explicit textual content, whereas the *situation model* is a mental representation that integrates this information with prior knowledge to construct a coherent understanding of what the text is about. This distinction offers a suggestive parallel to the dual nature framework: understanding requires both processing the surface information (analogous to Architecture) and constructing deeper meaning by connecting it to prior knowledge about the topic and its context (analogous to Relevance). When reading technical documentations, for instance, readers must simultaneously build a text-base of the technical specifications while constructing a situation model of how the system functions in practice. This suggests a hierarchical dependency: without adequate processing of the text-base, the situation model cannot be reliably constructed (cf. Kintsch, 1991). Van Dijk and Kintsch (1983) demonstrated a similar point in ear-

lier text comprehension research, showing that effective comprehension requires readers to move beyond the text-base to construct situation models that capture the referential meaning; what the text is actually describing in the world. Applied to programmed technical artifacts, this demand is particularly acute: readers may not achieve genuine understanding by just processing source code but should actively construct situation models that integrate how the artifact's Architecture enables its Relevance.

Pennington's (1987) distinction between *program model* and *domain model* highlights similar processes in the context of program comprehension. The program model contains information about the surface features of programs such as syntax and control flow. The domain model integrates these with domain knowledge about the problem being solved. Importantly, Pennington's model implies that effective comprehension is not achieved through either representation alone but requires cross-referencing between them, with the program model (Architecture) typically constructed before the domain model (Relevance) (cf. Pennington, 1987, p. 102). This sequential relationship suggests that deep understanding of technical artifacts requires not just knowing both representations but developing the ability to connect them flexibly.

Beyond moving and integrating these perspectives, there is the question of granularity: at what level of detail can an artifact's Architecture and Relevance be addressed? The Block Model (BM) provides a framework that speaks to this question (Schulte, 2008a; Schulte et al., 2010). It organizes program understanding into three dimensions: text surface (the actual code), program execution (how the code runs), and function (the goals and purposes of the program). These three dimensions are separated into two sides: structure (encompassing text surface and execution) and function. Their analysis of multiple program comprehension models reveals that expert programmers develop the ability to cross-reference between these dimensions (Schulte et al., 2010, cf. p. 80), while novices often become trapped in isolated understanding of structural elements (Architecture) without connecting them to functional purposes (Relevance). The educational implication, as the authors note, is that teachers should avoid teaching about function detached from structure, and equally avoid teaching structure without bridging to function (Schulte et al., 2010, cf. p. 70). This reinforces the findings discussed above and suggests that effective technical artifact understanding also requires connection across different levels of detail, not just between the two sides of the dual nature.

A distantly related line of work examines how such navigation into the details and between the two sides may unfold collaboratively. By studying student dyads making sense of sewing machine mechanics, Miyake (1986) developed the constructive-interaction framework. The study reveals how deep understanding emerges through iterative cycling between phases of understanding and non-understanding: identifying function, questioning that function, searching for mechanisms, proposing explanations, critiquing those proposals, and confirming

mechanisms before the cycle repeats at a deeper level. Miyake's key contribution is recognizing that feelings of non-understanding are not obstacles but essential markers of readiness to deepen comprehension, and that collaborative dialogue provides the engine for this process. While her framework focuses primarily on navigating mechanistic complexity through interaction, it demonstrates how collaborative processes may be essential for integrating architectural understanding (how systems work) with relevance understanding (why they work and what purposes they serve). This collaborative dimension will be developed further in Section 2.2.

Neither the Block Model nor the constructive-interaction framework fully addresses how learners integrate Architecture knowledge with Relevance knowledge in the sense defined by dual nature theory. Both implicitly recognize that technical understanding cannot be reduced to purely mechanistic knowledge, but neither explicitly theorizes the challenge of bridging between the two perspectives as distinct ontological dimensions of the artifact.

Recent empirical work provides more direct evidence for these dual-perspective challenges. Kallia (2023)'s comparative case study of inferential strategies during solved programming problem comprehension reveals that successful program comprehension requires students to navigate between the "problem level" and the "program level," a distinction that corresponds roughly to the Relevance and Architecture perspectives. Proficient students systematically employ what the authors call "Global Backward Explanations" that establish connections between these levels by linking mechanistic code behaviors (Architecture) to functional problem goals (Relevance). Struggling students, in contrast, concentrate predominantly on mechanistic associations without constructing interpretive bridges to functional purpose. Students who fail to engage in this dual-perspective reasoning experience systematic difficulties with both comprehension and transfer to similar problem contexts, providing concrete evidence that effective programming pedagogy should explicitly scaffold learners' capacity for connecting Architecture and Relevance.

Moving beyond computer science education into broader technology education, Attisano et al. (2021) investigated children's discussions of a mechanical artifact (a coffee grinder) in museum settings. Children naturally talked about multiple aspects of the machine including its overall appearance and component parts, but mechanistic talk was relatively rare, occurring in only 18 of 61 participants during the learning phase. The study also revealed age effects, with older children (4–8 years) showing better recall across all categories. Rather than demonstrating intuitive dual-nature competence across all age groups, the findings suggest that children at certain ages may need additional scaffolding to effectively integrate structural and mechanistic understanding when making sense of technical artifacts. This is particularly noteworthy given newer evidence that adults systematically underestimate children's capacity for understanding mechanistic explanations:

both experienced teachers and lay adults tend to deprioritize teaching the Architecture when working with young children, perceiving it as too challenging despite evidence that children benefit from such explanations (Betz et al., 2021).

In conclusion, the empirical evidence consistently implies: learners who fail to connect Architecture and Relevance experience systematic comprehension difficulties, while proficient understanding is characterized by the ability to cross-reference between both perspectives at appropriate levels of abstraction. What these findings leave open is what this means for how artifacts should be explained, a question the following implications address.

2.1.3 Implications for Understanding Artifact Explanation

The dual nature theory establishes important principles about how technical artifacts are distinctive objects of explanation. First, explaining artifacts requires integrating both perspectives rather than treating them separately. As the evidence reviewed above indicates, this integration is itself an achievement rather than an automatic byproduct of presenting either side.

Second, bridging the two perspectives requires deliberate effort because they are logically independent: Architecture does not entail Relevance, nor does Relevance entail Architecture. Kroes (2006) describes this as a “logical gap” between the two sides. De Ridder (2006) specifies what bridging it involves in practice, namely showing how the physical structure enables the behavior associated with the function, or in the terminology of this dissertation, how Architecture enables Relevance.

Third, the level of detail required in each perspective depends on the explanation’s purpose, audience, and context. Successful artifact explanation requires not only navigating between perspectives but also calibrating granularity to participants’ needs and goals. The Block Model (Schulte et al., 2010) makes this granularity dimension explicit, and collaborative sense-making enacts it visibly in interaction (Miyake, 1986).

Finally, the two perspectives are not properties intrinsic to the artifact alone. Engineers and designers bridge both aspects precisely because they understand how one enables the other (Kroes, 2002), and de Ridder (2007, p. 239) suggest that explaining a technical artifact is “more or less the reverse” of designing it. Where designers move from intended function to physical realization, explainers must reconstruct that relationship by identifying the artifact’s Relevance and showing how its Architecture realizes it. While some accounts treat the duality as a property of the artifact itself (cf. Kroes, 2010), the framework adopted here follows Budde (2021) and Schulte and Budde (2018) in holding that the two sides, in the epistemic context of making sense of an artifact, are also constituted through the perspective the human adopts. What Kroes (1998) formalized as structural and

functional description modes are, on this account, not neutral reflections of an artifact's properties but active construals through which the artifact is interpreted. What counts as Architecture or Relevance content is therefore not fixed by the artifact but negotiated between explainer, explainee, and the interactional context, a point taken up in Section 2.2.

Taken together, these implications give concrete shape to what the introduction termed the *dual nature tension*: the persistent explanatory challenge of having to address two logically disconnected yet ontologically inseparable perspectives and bridge them in a way that produces genuine understanding.

The Dual Nature Tension >>

Definition: The *dual nature tension* refers to the persistent explanatory challenge of having to address and integrate two logically disconnected yet ontologically inseparable perspectives on a technical artifact:

1. **Architecture** – physical structures, mechanisms, and causal relationships
2. **Relevance** – functions, purposes, and human meanings

The tension surfaces along three interconnected criteria that structure this chapter:

- **Dual-perspective:** An explanation must address both perspectives, since the artifact cannot be fully understood through either alone (Section 2.1).
- **Collaborative:** Their integration is not given in advance but negotiated through interaction as explainers and explainees jointly navigate between the two perspectives (Section 2.2).
- **Cognitive:** Understanding requires adopting multiple explanatory stances: mechanistic for Architecture, teleological for Relevance. Neither stance alone suffices (Section 2.3).

The tension arises not from the fact that artifacts *have* two natures, but from the requirement that explaining them demands continuous *bridging*. In everyday explanatory conversations this admits no general solution and instead calls for adaptive, cooperative work.

Miyake's (1986) dyads making sense of sewing machines offered an early glimpse of what such cooperative work looks like. The next section develops the theoretical foundations behind it.

2.2 The Collaborative Criterion

HOW DO EVERYDAY EXPLANATIONS OF TECHNICAL ARTIFACTS UNFOLD AS A COLLABORATIVE PROCESS?

Having established that technical artifacts require integrated explanations of both Architecture and Relevance, and that this integration proves challenging rather than intuitive, the question becomes: how might this integration actually occur in practice? As outlined at the start of this chapter, the dual-nature lens is now applied to a second body of literature: research on collaborative explanation. The goal is not to re-derive the Architecture–Relevance distinction but to examine whether and how the co-constructive processes through which everyday explanations unfold through monitoring, scaffolding, and grounding interact with the dual nature of the explanandum. To uncover why everyday explanation requires a fundamentally different theoretical lens than formal accounts of explanation, it is useful to consider the contrast between two traditions.

Traditional philosophy of science characterizes explanation as logical deduction from general laws to specific phenomena. Popper (1935) argued that to causally explain an event means to deductively derive a statement describing it from laws and initial conditions.⁴ The Deductive-Nomological Model (DNM) formalized this approach. From a DNM perspective, explanations answer why-questions by showing how phenomena follow logically from general laws applied to particular contexts (Hempel and Oppenheim, 1948). This model has influenced philosophical and scientific approaches to explanation, promoting the notion that an explanation is effective (and complete) if it shows how phenomena necessarily result from general laws and certain antecedent conditions. While philosophy of science has developed various alternatives to the Deductive-Nomological Model, the fundamental tension between formal logical approaches and everyday explanatory practices remains⁵.

In contrast to views in the philosophy of science tradition, empirical research on everyday explanation reveals a notably different picture. People routinely favor explanations based on purposes, functions, and intentions rather than deducing outcomes from general laws, and their preferences vary systematically across domains (Keil, 2006). In domains involving intentional agents or designed artifacts, teleological reasoning predominates (cf. Kelemen, 1999c), and people tend to

⁴“[E]inen Vorgang ‘kausal erklären’ heißt, einen Satz, der ihn beschreibt, aus Gesetzen und Randbedingungen deduktiv ableiten” (Popper, 1935, p. 31).

⁵Well-written and comprehensive reviews of different concepts of explanation have been published in the past. These works capture the breadth of the topic and do so more effectively than I could ever do. Thus, I would much rather refer interested readers to Miller (2019) and Salmon (1990)

value explanatory features such as simplicity and utility over strict logical deduction (Lombrozo, 2007). Everyday explanations emphasize understanding motives, reasons, and social context (Miller, 2019), reflecting a fundamentally different approach from scientific explanations with their emphasis on universal laws and explanatory completeness (cf. Rohlfling et al., 2021).

When people explain in everyday situations, they “depart radically from the image of people methodically considering a set of axioms and running through a deductive chain” (Keil, 2006, p. 228). Instead, everyday explanations unfold as inherently dynamic and collaborative processes where participants actively construct understanding through dialogue. This shift from knowledge transfer to knowledge co-construction represents more than a minor modification; it reconceptualizes what explanation is and how it works (Rohlfling et al., 2021). Rather than evaluating explanations against formal standards, the co-constructive view foregrounds the processes through which understanding develops: monitoring, scaffolding, and grounding. Because this view focuses on the dynamics of the process rather than the specific characteristics of what is being explained, it complements the content-focused dual-perspective criterion developed in the previous section.

Key principles of co-constructivism as described by Reusser and Pauli (2015) are: (1) knowledge is actively constructed rather than passively received; (2) learning is situated in specific contexts and communities of practice; (3) meaning-making occurs through social negotiation. Similar to how teaching, albeit jokingly, was once interpreted as mechanically funneling knowledge into passive recipients (the “Nürnberger Trichter”), the co-constructive view rejects this one-way transfer model in favor of “a joint process of negotiating meaning through dialogue” (Reusser and Pauli, 2015).

In such situated, collaborative activity (Reusser and Pauli, 2015), explanation⁶ is not a linear deduction but a dynamic process involving questioning, elaboration, and adaptation to the emerging context. This process demands sophisticated coordination between cognitive and linguistic resources (Kern, 2011): explainers must mentally decompose complex phenomena into constituent elements while making them linguistically accessible to others, drawing on and reorganizing existing domain knowledge and mental representations of the subject matter (Chi, 2000; Keil, 2006). The quality of the resulting explanation is validated not through (formal) correctness but through (pragmatic) coherence and the degree to which it supports understanding in the specific situation (Reusser and Pauli, 2015; Rohlfling et al., 2021).

⁶Henceforth I will use the terms “everyday explanation” and “explanation” interchangeably to refer to everyday explanations. If other types of explanations are considered, it is clearly stated in the text.

In summary, this understanding of co-constructed explanations differs fundamentally from scientific explanations. Scientific explanations pursue completeness and exactness, whereas everyday explanatory dialogues respond to knowledge gaps that emerge during the interaction itself. Miller (2019) captures this contrast: “[E]xplanations are not just the presentation of associations and causes (causal attribution), they are contextual. While an event may have many causes, often the explainee cares only about a small subset (relevant to the context), the explainer selects a subset of this subset (based on several criteria), and explainer and explainee may interact and argue about this explanation.” This processual character transforms the explanandum into a “moving target”: the object of explanation shifts as understanding develops and new questions arise (Rohlfing et al., 2021).

Several features of this dynamic are particularly relevant to everyday explanations of technical artifacts. They build incrementally on established *common ground*, with each utterance depending on what has been mutually established or is considered shared knowledge. As the explanandum moves, participants continuously revise their models of each other’s knowledge and understanding, and explanatory strategies adapt in real time to comprehension signals or emerging signs of confusion.

2.2.1 Co-Constructive Processes: Monitoring, Scaffolding, and Grounding

Co-construction is not taken for granted. As Reusser and Pauli (2015) emphasize: it “does not just happen because individuals are co-present: individuals must make a conscious, continued effort to coordinate their language and activity with respect to shared knowledge.” The dynamics of co-constructive explanation can be theoretically captured through three interrelated processes: *monitoring*, *scaffolding*, and *grounding*; processes that describe how participants coordinate their contributions in practice.

Monitoring involves the continuous assessment of an explainee’s understanding through verbal and non-verbal cues, as well as also the self-monitoring of an explainers’ “own knowledge and understanding of the material” (Roscoe, 2014) as they explain. Clark and Krych (2004) demonstrate how speakers adjust their utterances in real-time based on the explainee’s signals of understanding, for instance through reformulations triggered by signs of confusion. From a conversation-analytic perspective, monitoring extends beyond explicit comprehension checks to include the micro-details of timing, overlap, repair sequences, and multimodal cues that indicate understanding states (Jefferson, 1984; Schegloff, 2007). Monitoring thus influences both how a conversation unfolds pragmatically (how something is explained) and semantically (what is explained). What explainers monitor is represented mentally as what is commonly referred to as a partner model Buhl

et al. (cf. 2024) and Schaffer et al. (2024): a developing mental representation of the explainee's, for example, knowledge and interests. Their qualitative investigation of video recall interviews revealed that explainers track what explainees know and want to know, with these mental representations developing from vague initial conceptions toward well-defined partner models as the explanation progresses (see also Lazarov et al., 2026, for a methodological reflection on the video recall method).

Scaffolding is situative support provided by explainers that enables explainees to advance towards an understanding they could likely not reach on their own (Wood et al., 1976). It involves offering temporary assistance that is gradually withdrawn as the explainee develops understanding. The theoretical foundations of scaffolding are closely tied to Vygotsky's concept of the Zone of Proximal Development (ZPD): "the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers" (Vygotsky, 1978, p. 86). As Mermelshtine (2017) notes, "scaffolding could be thought of as the strategies employed within the ZPD to promote development." A common misconception is that scaffolding casts learners in passive roles. Instead, in co-constructive explanation, explainees are not mere recipients of knowledge but actively shape the explanation through their contributions, questions, and demonstrations of understanding. For technical artifact explanation, scaffolding is relevant because the dual nature creates specific zones of difficulty: an explainee may grasp the Architecture of an artifact yet require targeted support to understand its Relevance, or vice versa, and the explainer must calibrate their assistance to whichever perspective currently is beyond the explainee's independent reach, to, for example, bridge the "logical gap" (Kroes, 2006).

Grounding is the process through which the shared knowledge necessary for mutual understanding is established and maintained (Clark and Brennan, 1991). It ensures that participants share sufficiently similar interpretations of the information being communicated, creating *common ground*: shared knowledge, beliefs, and assumptions that enable successful interaction (Clark and Brennan, 1991; Wilkes-Gibbs and Clark, 1992). What grounding produces is not a static foundation but a dynamic, jointly constructed communicative space. As Mercer (2002) describes:

"For a teacher to teach and a learner to learn, they must use talk and joint activity to create a shared communicative space [...] which is reconstituted constantly as the dialogue continues, the teacher and learner negotiate their way through the activity in which they are involved." (Mercer, 2002, p. 141)

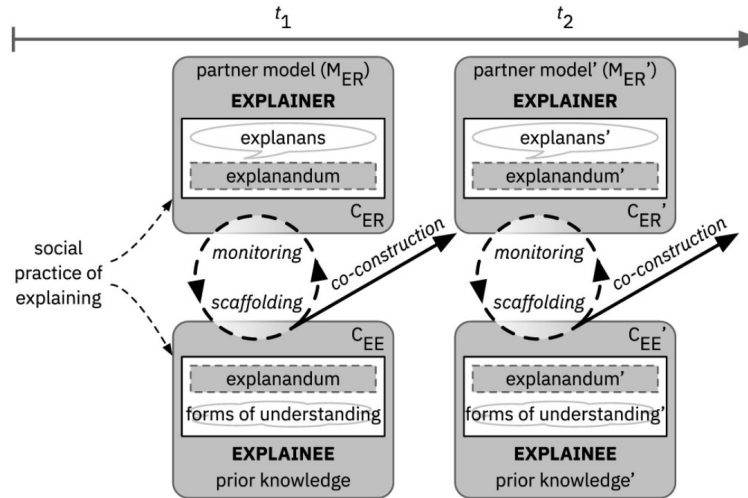


Figure 2.2: The dynamics of the co-construction process (Rohlfing et al., 2021). During the social practice of explaining, there is a continuous cycle of monitoring and scaffolding that influences the co-construction of understanding over time.

The relationship between these three processes of monitoring, scaffolding, and grounding is cyclical and iterative. Grounding establishes the foundational shared understanding, while monitoring continuously observes its state. When monitoring reveals gaps or misunderstandings, scaffolding strategies are employed to bridge them and re-establish common ground. The communicative space described by Mercer (2002) is thus maintained through this constant interplay of establishing shared understanding (grounding), assessing its effectiveness (monitoring), and providing adaptive support when needed (scaffolding).

Importantly, these are not unidirectional processes in which explainers monitor and scaffold explainees. The co-constructive perspective recognizes that explainees actively shape explanations through their responses, questions, and demonstrations of understanding, which serve as essential resources for the explainer (Rohlfing et al., 2021). These processes operate within what Rohlfing et al. (2021) conceptualize as social practices: established patterns of interaction that provide macro-level structure while remaining flexible enough for micro-level adaptation (cf. Figure 2.2). In technical contexts, this might involve recognized roles (expert/novice, teacher/student) that create expectations about who should monitor, when scaffolding is appropriate, and what constitutes adequate grounding.

A central idea is that explanation is most productive when calibrated to the explainee's current understanding, operating in the Zone of Proximal Development (Vygotsky, 1978) and elaborated in scaffolding research on parent-child (Mermelsh-
tine, 2017) and classroom interaction (Sentance and Waite, 2021). Rather than aiming for exhaustive accuracy, a successful explanation requires that the explainer

and explainee arrive at a shared sense of completion. In contrast, simply aiming for completeness can even be detrimental, as too much detail may overwhelm rather than support understanding (Keil, 2006). All in all, the quality of an everyday explanation does reside in the effectiveness of the dynamic, fragile co-constructive (Reusser and Pauli, 2015) interplay of monitoring, scaffolding, and grounding through which understanding is jointly achieved until both parties are satisfied.

2.2.2 Empirical Support

This section examines empirical evidence for how co-constructive processes manifest in explanatory interactions. The studies considered here reveal how monitoring, scaffolding, and grounding transform explanations into collaborative achievements rather than one-way transmissions of knowledge, even when researchers do not explicitly position their work within a co-construction framework. The wide range of contexts in which explanations occur has produced a body of evidence that is extensive yet fragmented. The following discussion synthesizes these varied findings through the lens of co-construction.

One area of research closely related to the dynamics of co-constructive explanation is teacher education in which a key concept is Pedagogical Content Knowledge (PCK) (Shulman, 1986): “that special amalgam of content and pedagogy that is uniquely the province of teachers, their own special form of professional understanding” (Shulman, 1987, p. 8). Coe et al. (2014) found strong evidence for a link between PCK and learning outcomes, noting that “[t]eachers must [...] understand the ways students think about the content, be able to evaluate the thinking behind students’ own methods, and identify students’ common misconceptions.” PCK, in other words, enables teachers to fulfill a critical function in the classroom: monitoring understanding.

Research on teaching effectiveness provides substantial empirical support for the coordinated use of these three strategies. Hattie (2012) demonstrates that when teachers establish clear learning intentions (grounding), provide effective feedback (monitoring and scaffolding), and use formative assessment to guide instruction, student learning improves significantly. As Hattie puts it, “[w]hen teachers see learning occurring or not occurring, they intervene in calculated and meaningful ways to alter the direction of learning to attain various shared, specific, and challenging goals” (Hattie, 2012).

However, accurately assessing whether learning is occurring is a skill in its own right. Research on student tutors shows that tutors can assess normative knowledge but fail to diagnose alternative conceptions (Chi et al., 2004). Hattie (2012) similarly found that teachers often overestimate their ability to assess student understanding. One systematic source of such misjudgment is egocentrism (cf.

Keysar et al., 1998; Nickerson, 1999): explainers tend to overestimate the common ground between themselves and their interlocutors, which disrupts the negotiation of shared understanding (Keil, 2006).

These findings raise a further question: is it always most important to provide the best possible explanation, or can engaging learners in the *process* of explaining improve their reasoning and understanding, even when the resulting explanations are imperfect? Wilkenfeld and Lombrozo (2015) address this by contrasting two perspectives. Inference to the Best Explanation (IBE) treats having the correct explanation as the key to improved understanding. Explaining for the Best Inference (EBI), by contrast, emphasizes the cognitive benefits of the explaining process itself, shifting the focus from the product to the activity of co-construction. In line with the findings of Hattie and Chi, EBI suggests that the activity of explaining can lead to better learning and inference even when the explanations generated are incomplete (Wilkenfeld and Lombrozo, 2015). For the present discussion, this is significant because it implies that the co-constructive process itself, especially through the active participation of the explainees, contributes to understanding and not just alters the content it produces.

An analysis of teacher-student interactions demonstrated that successful scaffolding depends on the teachers' ability to adjust support based on real-time assessments of student understanding through dialogue (van de Pol, 2012). This adaptive responsiveness exemplifies co-construction in action: explanations are not pre-determined but emerge through the interaction itself. However, "as observational studies show, coconstructive learning and dialogue are hardly homogeneous but inherently fragile processes in the service of convergence and mutual intelligibility" (Reusser and Pauli, 2015). This fragility underscores an important characteristic of co-constructive explanations: their success depends on the balance between explainer knowledge and explainee needs, a shared motivation on both sides, and is achieved through ongoing negotiation rather than preconstructed scripts.

Research on learning activities adds further support. Versteeg et al. (2019) compared peer instruction with self-explanation in medical physiology education and found that peer instruction significantly improved comprehension over the self-explanation condition. The authors attribute this advantage to collaborative discourse that enables the correction of misconceptions through co-construction of knowledge. This aligns with the ICAP framework (Chi and Wylie, 2014), which categorizes learning engagement into interactive, constructive, active, and passive modes, with interactive engagement yielding the deepest learning. Peer instruction fits within this highest tier, but its efficacy depends on both partners contributing constructively (cf. Reusser and Pauli, 2015).

The evidence for co-constructive processes extends beyond formal educational settings. Clark and Krych (2004) investigated speaker-addressee interactions during a Lego assembly task and found that effective communication requires speakers to monitor not only their understanding but also that of their addressees. When visual

feedback was eliminated, communication became slower and more error-prone, highlighting how co-construction depends on the availability of different monitoring channels. In a similar vein, Leung et al. (2025) studied parent-child interactions during a director-matcher task and found that parents intuitively scaffold what they term “conversational pact formation” the process by which interlocutors establish shared linguistic conventions. Parents spontaneously adjusted their support based on real-time monitoring of their children’s communicative needs, demonstrating how co-construction emerges naturally in explanatory interactions. These findings are consistent with broader reviews showing that scaffolding plays a significant role in children’s developmental outcomes across domains such as executive function, language acquisition, and cognitive abilities (Mermelshtine, 2017). In a related developmental context, Kern et al. (2024) operationalized the GLOBE model to assess discourse competence in preschool children’s game explanations, finding evidence of a fine-tuned interactive support system in which scaffolding adapts to the competence the child displays rather than following a fixed script.

Across these diverse contexts, successful explanatory interactions share a common characteristic: responsiveness to the dynamic flow of understanding. Evidence from different settings as varied as expert–layperson advice (Nückles et al., 2006), spatial route descriptions (Buhl, 2001), and classroom activities (Hattie, 2012) indicates that adapting to the addressee supports effective communication and learning outcomes. More directly, Buhl et al. (2024) show that partner models change over the course of explanations. The effectiveness of explanation thus depends not only on appropriate knowledge of the explanandum but also on the explainer’s ability to adapt to the explainee’s emerging understanding.

2.2.3 Implications for Understanding Everyday Explanation

The empirical findings reviewed above converge on several key insights. First, the criteria for evaluating co-constructive explanations differ fundamentally from those applied to scientific explanations. Rather than measuring explanations against formal logical standards or content completeness, explanatory success in everyday contexts emerges through the achievement of mutual understanding during interaction (Rohlfing et al., 2021). As the evidence on Explaining for the Best Inference suggests (Wilkenfeld and Lombrozo, 2015), the co-constructive process itself might even contribute to understanding, not just through the content it produces. Pragmatic coherence and joint sense-making constitute more relevant success criteria than adherence to formal models.

Second, co-constructive explanation is cognitively demanding, particularly for explainers who must simultaneously structure their knowledge, orchestrate the dynamics of interaction, and adapt to the explainee’s emerging needs. The coordination of monitoring, scaffolding, and grounding requires metacognitive awareness that goes well beyond domain expertise. The systematic obstacles identified in

the literature, including the curse of knowledge (Keil, 2006), egocentric bias (cf. Keysar et al., 1998; Nickerson, 1999), and the tendency to overestimate monitoring accuracy (Chi et al., 2004; Hattie, 2012), underscore the inherent fragility of co-constructive processes (Reusser and Pauli, 2015) while highlighting their necessity for promoting understanding.

The social dimensions of co-construction operate through established practices, but also maintain responsiveness to demands emergent in the dialogue. Van de Pol's (2012) analysis demonstrates how successful scaffolding emerges from teachers' ability to adjust support based on real-time assessments through dialogue, showing how explanations are collaboratively constructed rather than delivered from one to another.

These insights establish that successful explanation depends on alignment between explainer and explainee regarding knowledge states, goals, and contextual constraints. The collaborative processes of monitoring, scaffolding, and grounding describe how this alignment is achieved and maintained in interaction. These processes are domain-general in principle, but when the explanandum has a dual nature, they may acquire a specific burden: monitoring would need to track which perspective the explainee currently engages, scaffolding may need to support transitions between perspectives, and grounding would need to establish shared vocabulary for both. What the collaborative criterion does not address, however, is the cognitive question underlying these demands: what enables people to shift between mechanistic reasoning about how something works and teleological reasoning about what it is for? This is the third and final criterion needed before the three can be synthesized into the Co-Constructive Dual Nature Framework in Section 2.4.

2.3 The Cognitive Criterion

WHAT COGNITIVE PROCESSES ENABLE PEOPLE TO REASON ABOUT BOTH SIDES OF AN
ARTIFACT'S DUAL NATURE?

The previous two criteria addressed artifacts as distinctive as objects of explanation, and how explanations unfold as collaborative processes. This third criterion extends the dual-nature distinction to a further domain: cognitive science research on epistemic modes of construal. Where the preceding sections encountered the duality as a content-level and process-level challenge, this section examines it as a cognitive challenge, asking what enables people to shift between reasoning about Architecture and reasoning about Relevance.

This question gains additional force from an argument developed in the preceding sections. If the dual nature arises through the perspective the human adopts within the interaction (cf. Budde, 2021; Schulte and Budde, 2018), then the cognitive

processes that govern perspective adoption become constitutive of how the dual nature manifests in explanation. Architecture and Relevance are, on this account, not neutral reflections of artifact properties but active construals through which the artifact is interpreted. What drives a person toward one perspective, and what enables flexible movement between them, is therefore a necessary complement to the Dual-Perspective (Section 2.1) and Collaborative Criterion (Section 2.2).

The dual-perspective criterion established that for understanding technical artifacts, an integrated grasp of Architecture and Relevance is important. The cognitive question is what happens on the side of the person: what draws someone toward one description rather than the other, and what enables flexible movement between them? When confronted with an unfamiliar device, we might ask what it is *for*, or we might ask how it *works*—two questions that lead to qualitatively different kinds of reasoning.

Research demonstrates that people shift between different epistemic approaches when interpreting the world, sometimes emphasizing causal mechanisms, sometimes functions and purposes (e.g., Keil, 2006; Kelemen, 1999c; Lombrozo and Carey, 2006). Cognitive psychologists describe these shifts as *modes of construal* (Keil, 1994), philosophers of mind as *stances* (Dennett, 1987). Developed to account for reasoning across biological, physical, and intentional domains, these traditions converge on an observation directly relevant to this dissertation: when applied to designed objects, the orientations they describe align with the dual nature perspectives: mechanistic reasoning with Architecture, teleological reasoning with Relevance.

Keil (1994) identified distinct “modes of construal” that people use to understand different domains. He proposed that from early childhood, people alternate between at least two fundamental explanatory templates: *the mechanistic (physical) mode* and the *functional/teleological mode of construal*. The mechanistic mode focuses on components, parts, and causal chains, leading a person to reason about how technical artifacts work through their internal components and interactions. The teleological mode is used to make inferences about purposes served by things, emphasizing what technical artifacts are designed to accomplish rather than how they mechanistically operate. Each mode emerges early in development but becomes domain-specialized later in life. For non-living or non-agentive objects (such as rocks), people tend to provide mechanistic explanations of *what* something is made of. For living things or purposefully designed objects (such as animals and tools), they tend to provide teleological explanations of *what it’s for* or *what it intends to do*.

Dennett (1987), a philosopher of mind, proposed that people employ different *stances* as strategies for explaining and predicting the behavior of entities. The *physical stance* treats objects as following the laws of physics, with predictions based on their physical state and scientific laws (as, for example, applied in ballistics or chemistry). Applied to technical artifacts, this involves understanding

how electronic components, mechanical parts, and physical processes interact to produce observable outcomes. The *design stance* assumes entities are designed for specific purposes, with explanations drawing on knowledge about that design. As Dennett explains, when categorizing an object as an alarm clock, one can reason about its future behavior by assuming “it has a particular design—the design we call an alarm clock—and that it will function properly, as designed” (Dennett, 2009, p. 2f). The *intentional stance* treats entities as rational agents with beliefs and desires, as typically applied to humans, animals, and intelligent machines. For artifacts, this involves understanding them through the intentions of their creators. Each stance represents an increasing level of abstraction, from basic physical processes to purposeful design to mental states, and each leads to a distinct type of explanation (Amerongen, 2008, p. 82).

Regarding explaining the technical artifacts’ dual nature, the question how the intentional stance and the design stance relate to each other arises. Amerongen (2008) provides a critical reflection of the design stance and a comprehensive discussion of the stances’ relation to and usefulness for artifact interpretation that is especially helpful for disentangling the three stances and their relation to the dual perspectives. Dennett’s motivation for the design stance was, among other things, that interpreting the function of designed artifacts should not require the interpretation of intentions: in his view, the function of an artifact is simply what it can do best, rather than what the designer intended. However, as Amerongen (2008) argued thoroughly, there is a considerable overlap between the intentional and the design stance when it comes to artifact interpretation. She notes that “cognitive psychological research strongly suggests that the intentionalistic account is the better description of our ordinary way of ascribing functions to artifacts [...]. Our everyday conceptualization of artifacts is largely driven by what we believe the designer of the artifact intended the artifact to be” (Amerongen, 2008, p. 214). She ultimately concludes (Amerongen, 2008):

“Starting with rough hypotheses about intended purpose, capacity, and assumptions of rationality, we slowly might get to a better understanding of what an artifact could have been made for [...]. Putting the intentional stance back in line with the design stance would put the intentional stance at the core of the theory of the stances.”

Interestingly, a year after Amerongen published her Dissertation, Dennett—although not citing her work—added that the intentional stance is “a subspecies of the design stance” (Dennett, 2009, p. 3). Amerongen (2008) concludes from her review of existing empirical evidence that the intentional stance performs well concerning empirical reality, while the design stance does not. According to Kelemen (1999c), both the design and intentional stance derive from the more primitive teleological stance, leading to her claim that children are “promiscuous teleologists”.

In conclusion, empirical research in cognitive psychology indicates the modes of construal shape how people interpret the world, while Dennett proposed these three stances as predictive strategies⁷. Although the number of modes or stances differs across the two frameworks, they parallel each other closely: Keil's mechanistic mode corresponds to Dennett's physical stance, while the teleological mode can be understood as a compound of the intentional and the design stance (cf. Amerongen, 2008, thorough argumentation for this position). For the purposes of this dissertation, the key implication is that when these modes are applied to technical artifacts, the mechanistic mode aligns with the Architecture perspective, as both concern parts, processes, and causal mechanisms. The relationship between the teleological mode⁸ and the Relevance perspective is more layered. At its core, functional reasoning captures what an artifact is designed to do, which corresponds to the functional heart of Relevance. But as Amerongen's analysis suggests, interpreting artifacts in practice draws heavily on the intentional stance, which involves reasoning about designer intentions, user goals, and the broader context of human purposes (Amerongen, 2008). For complex, socially embedded artifacts, this intentional reasoning naturally extends to the wider aspects of Relevance: the values, social structures, and transformed practices associated with the artifact's use. The mode of construal remains teleological and intentional in character; what changes with artifact complexity is the range of human purposes and social meanings it must encompass. It is precisely the extension that the concept of Relevance as in (Budde, 2021; Winkelnkemper et al., 2024) was designed to capture.

2.3.1 Empirical Support

Empirical research suggests that these modes of construal are not fixed cognitive structures but flexible tools that people switch between dynamically based on context and explanandum (Lombrozo, 2009). Rather than adopting a mechanistic mode for some kinds of objects and a functional mode for others, people flexibly adopt one mode or the other depending on the entity in question and their inferential aims (Lombrozo and Wilkenfeld, 2019). There is supporting evidence that people naturally generate different kinds of explanations depending on context: they tend to offer particular explanations in response to specific features of the explanandum (Lombrozo and Carey, 2006), value mechanistic explanations more when the focus is on causal processes, and prefer functional explanations when reasoning about purposes or roles (Vasilyeva et al., 2015).

A further question is whether these different modes lead to different types of understanding. Lombrozo and Wilkenfeld (2019) argue that mechanistic and functional understanding are distinct, though they debate how deep the distinction runs

⁷Henceforth I will use stances and modes interchangeably.

⁸Henceforth I will use teleological and functional mode interchangeably

cognitively. At minimum, the two modes involve different causal commitments, privilege different relationships as a basis for generalization, and serve different inferential purposes. Notably, functional understanding operates with a degree of mechanism-independence implying one can understand what an artifact is for without knowing in detail how it works.

If the two modes of construal serve different epistemic purposes, a follow-up question is whether the order in which they are engaged matters. McCarthy and Keil (2023) found that adults show a systematic preference for functional information to precede mechanistic information, reflecting a broader “whole-before-parts” orientation. Across 160 participants, explanations presenting function before mechanism were rated higher than the reverse order, and participants also preferred a top-down traversal through hierarchical levels of abstraction. For this dissertation, this finding is notable: it suggests that a Relevance-first ordering may be cognitively more intuitive for recipients, even if—as the naturalistic data in this dissertation will later show—explainers tend to adopt an Architecture-first strategy.

Regarding the developmental origins of these stances, teleological reasoning about goals and intentions emerges remarkably early in development (Csibra, 2003; Gredebäck and Melinder, 2011; Hernik and Csibra, 2015), and there is evidence that such explanations may be psychologically privileged and often favored over mechanistic ones (Lombrozo and Wilkenfeld, 2019). “We humans have purpose in the brain. We find it hard to look at anything without wondering what it is ‘for’, what the motive for it is, or the purpose behind it” (Dawkins, 1995). Evidence suggests that teleological reasoning represents a default cognitive mode, a claim supported by evidence across the lifespan: young children give promiscuously teleological explanations of the natural world (Kelemen, 1999a), adults show increased acceptance of teleological explanations under time pressure (Kelemen and Rosset, 2009), and, in the strongest test of this claim, even professional physical scientists, whose disciplinary norms explicitly reject teleological reasoning, revert to teleological explanations when their processing resources are constrained (Kelemen et al., 2013). However, the developmental picture is not one of simple succession from one explanatory mode to another. Rather, different explanatory schemes appear to coexist from early on, with children and adults flexibly deploying teleological and mechanistic explanations depending on context and domain (Keil, 2006). What develops is not so much the availability of these modes but the sophistication with which they are selected and coordinated.

One question remains: can modes of construal, which are typically experimentally measured through explanation preferences or generation tasks and have often been characterised as relatively stable orientations, be plausibly aligned with perspective adaptation observable within a single explanation episode? Two clarifications are in order. A growing body of empirical work treats modes as context-sensitive orientations that can be cued by task demands, inferential goals, or the content

of the explanation itself (Liquin and Lombrozo, 2018; Vasilyeva et al., 2017). Even the Kelemen et al. (2013) “cognitive default” account describes functional mode of construal as something that can be suppressed and *re-emerge* under varying cognitive conditions, a claim about dynamic activation rather than trait-like stability. The inferential claim made here is as follows: perspective shifts in explanations are not treated as one-to-one markers of mode transitions in the cognitive sense, but as *surface evidence* of what is currently, in an epistemic sense, foregrounded. Architecture and Relevance are features of the artifact and the discourse about it reflects it; mechanistic and functional modes are features of the construer’s cognitive orientation. These are related but not identical. Any inferences about construal-relevant orientation therefore rest not on individual sentences in isolation, but on *patterns* of perspective distribution across larger stretches of dialogue, for example, dialogue thirds, phases, or episodes, which is the level at which the empirical analyses reported in the following chapters are ultimately conducted.

In sum, the empirical evidence establishes three points relevant to this dissertation. First, mechanistic and functional modes are flexible cognitive tools whose deployment depends on context, explanandum, and explanatory goal. Second, the two modes involve different causal commitments and serve different inferential purposes, and functional understanding can operate with a degree of mechanism-independence, providing cognitive grounding for the claim that Architecture and Relevance cannot substitute for each other in explanation. Third, functional reasoning appears to be the developmentally earlier and cognitively less demanding mode, which suggests a natural asymmetry between the two perspectives.

2.3.2 Implications for Understanding Artifact Explanation

The research on modes of construal carries several implications for how technical artifacts are explained in educational and everyday contexts. First, because technical artifacts possess a dual nature, their explanation requires flexible adoption of such stances rather than privileging a single explanatory approach. Different stances serve different epistemic functions, and effective explanation depends on matching the stance to the current understanding goal. Crucially, this flexibility must be coordinated between explainer and explainee: if the two parties are reasoning from different modes—one mechanistically, the other teleologically—the explanation risks failing not because of missing content but because of misaligned interpretation. This alignment challenge is particularly relevant when an explainee’s age or background knowledge makes certain stances less accessible, suggesting that scaffolding toward a particular mode, for example supporting children to engage mechanistically with an artifact’s Architecture, can be helpful for more integrated understanding (cf. Attisano et al., 2021).

Second, the developmental and cognitive evidence that functional reasoning is the earlier-emerging and less demanding mode has a direct implication for explanation sequencing. If Relevance-oriented reasoning is cognitively more natural—both developmentally and under ordinary processing conditions—then explanations that begin with Relevance may provide a more natural starting point for explainees. This supports educational approaches that establish functional context before introducing architectural detail, for example by anchoring explanations in the artifact’s role in everyday life before unpacking its internal mechanisms (cf. the idea of using personal instead of arbitrary data in Register and Ko, 2020). The empirical finding by McCarthy and Keil (2023) that recipients prefer function-before-mechanism ordering reinforces this implication normatively: the cognitively natural order for recipients may not be the order that explainers naturally adopt.

Third, because the functional mode develops earlier and might be cognitively less demanding than the mechanistic mode, explainees may naturally orient toward Relevance-focused understanding even when both perspectives are presented. Such an asymmetry between the two modes would imply that integrated understanding does not necessarily emerge spontaneously from content alone, but requires active coordination from the explainer—for example, through monitoring and deliberate scaffolding to bridge the two perspectives.

Together, these implications elucidate the core argument that propels this dissertation forward: explanatory effectiveness might depend less on the amount of Architecture or Relevance content than on how the two are coordinated—with each other, with the explainee’s current stance, and with the demands of the explanatory context. This sets the stage for a synthesis in the form of an integrated framework that will be developed in the following section.

2.4 Synthesis: The Co-Constructive Dual Nature Framework

The three criteria developed in this chapter do not merely describe parallel aspects of technical artifact explanation; they identify interdependent conditions that must be coordinated for explanation to succeed. This section synthesizes them into the Co-Constructive Dual Nature Framework, an integrated analytical model that captures the distinctive challenges of explaining technical artifacts and provides the theoretical infrastructure for the empirical investigations reported in later chapters.

The dual-perspective criterion established that technical artifacts occupy a distinctive ontological position that resists explanation from just a single perspective. Neither an account of Architecture alone nor of Relevance alone suffices for genuine understanding, which consists of comprehension (knowing that) and enabledness

Three-Layer Framework

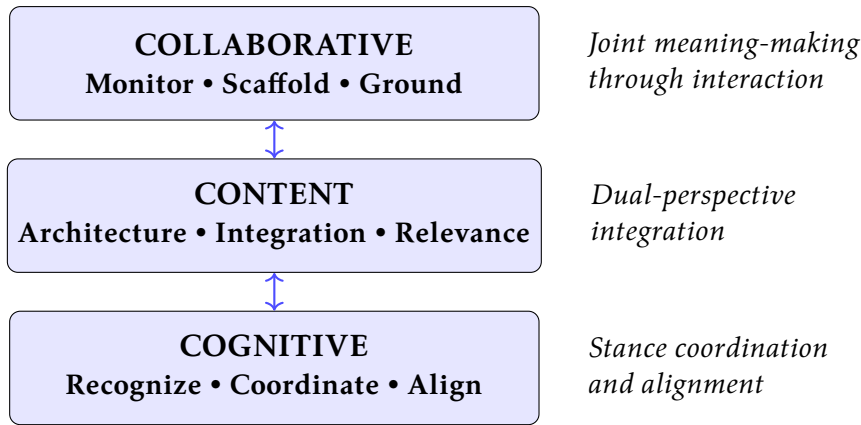


Figure 2.3: The three-layer structure of the Co-Constructive Dual Nature Framework, illustrating the collaborative, content, and cognitive layers and their interdependencies.

(knowing how) (Buschmeier et al., 2025). Specified as the *dual nature tension* earlier (see p. 34), what matters is their integration. Especially the empirical evidence presented in this chapter reinforced this theoretical claim. Different sources demonstrated that learners who fail to connect the two perspectives encounter systematic difficulties with both comprehension and transfer (e.g., goal-plan mismatches in Spohrer et al. (1985)).

The collaborative criterion established that this does occur through joint meaning-making between explainer and explainee. The co-constructive processes of monitoring, scaffolding, and grounding are the mechanisms through which dual-perspective navigation becomes possible in practice: monitoring reveals which perspective the explainee currently engages, scaffolding supports transitions between them, and grounding maintains the shared interpretive base from which both participants can operate. Explanatory success, on this account, is reinforced through adaptive interaction rather than achieved through content transmission alone.

The cognitive criterion revealed that people bring distinct but flexible epistemic tools to artifact interpretation. The mechanistic and teleological modes of construal align with the Architecture and Relevance perspectives, and the evidence suggests these modes are dynamically deployed depending on context, explanandum, and goal. The finding that the teleological stance can be observed earlier in development of children, and as it seems to be cognitively less demanding, introduces a natural asymmetry between the two perspectives, one with direct consequences for how explanation sequencing affects understanding.

Rather than treating collaborative process, dual-nature content, and cognitive stance as separate concerns, the framework demonstrates their synergistic operation. It is organized into three interdependent layers, illustrated in Figure 2.3.

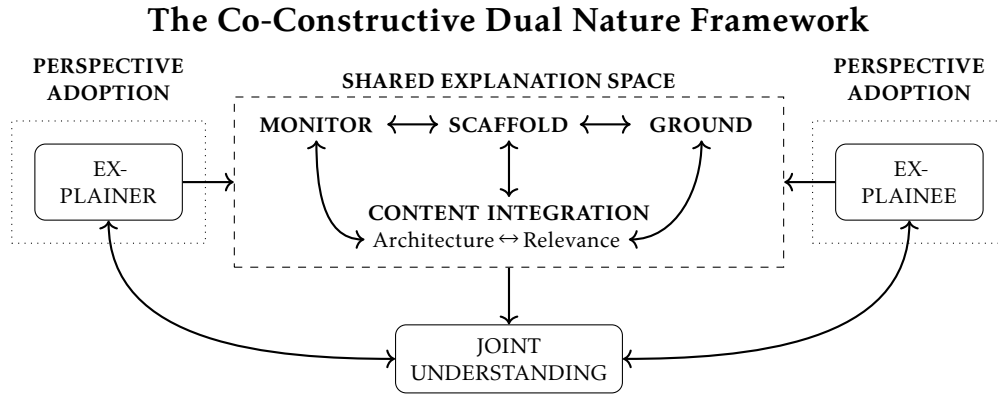


Figure 2.4: The Co-Constructive Dual Nature Framework. Explainer and explainee each bring stance adoption processes to a shared explanation space (cf. Mercer, 2002, idea of a shared communicative space) in which collaborative processes and content integration operate interdependently, oriented toward joint understanding. The collaborative layer captures joint meaning-making through monitoring, scaffolding, and grounding. The content layer covers how Architecture and Relevance perspective are used to address the artifact integrate the two sides in explanatory utterances. The cognitive layer captures the mode of construal adoption through which explainer and explainee orient themselves toward one or both perspectives.

The interdependence between layers is the framework's central theoretical claim. Content selection depends on monitoring what the explainee currently understands and which stance they are taking. Scaffolding strategies must be calibrated to support the stance transitions that the content structure requires. Grounding negotiates not just shared vocabulary but shared interpretive frameworks for connecting the two perspectives. It is not sufficient to address Architecture and Relevance in the content layer if the collaborative processes do not support the stance transitions required to move between them. Equally, monitoring is only effective insofar as the observer can identify which perspective the explainee is currently engaging, a judgment that requires sensitivity to both content and cognitive cues (e.g., the form of questions asked, or mechanistic versus functional vocabulary) simultaneously. The full dynamics of this interdependence are captured in Figure 2.4, which situates the shared explanation space (see Mercer, 2002, for the idea of a shared communicative space, from which this name was inspired) between the stance adoption processes of explainer and explainee, with joint understanding as the outcome toward which the interaction is oriented.

Because the three layers are interdependent, focusing analysis on one layer generates principled expectations about the others. Observable content shifts, in particular, can serve as indicators of underlying stance transitions and of the collaborative moves that prompted them, making the content layer a viable empirical entry point from which collaborative and cognitive dynamics can be inferred. The empirical chapters of this dissertation utilize this property directly in Chapter 3 which develops an appropriate analytical strategy.

3 Empirical Contributions

THE Co-Constructive Dual Nature Framework developed in the previous chapter is an idealized model, describing how explanations unfold as collaborative processes in which explainers and explainees jointly navigate Architecture and Relevance through the modes of construal. Translating such a model into empirical research necessitates strategic decisions regarding the initial step. Not all dimensions of the framework are equally viable as starting points, and attempting to operationalize all of them simultaneously would risk producing shallow measurements across too many layers at the cost of any one being reliable. The contributions of this dissertation are organized into four *Building Blocks* (BB1–BB4), each addressing a different empirical dimension of the framework in a cumulative sequence. In the process, the empirical work also developed analytical tools. In particular, a temporal normalization and visualization techniques for coded dialogues proved central to the pattern-finding reported below and that are reusable beyond the specific content categories of this dissertation and are extendable for finding patterns algorithmically (see Appendix).

The strategy this dissertation adopts, as argued for in Section 2.4, is to establish the *content layer* as the empirical entry point. The explanatory utterances are the most directly observable displays of the framework’s theoretical constructs: when an explainer says something about the artifact, that utterance is either addressed toward the Architecture of the artifact or toward its Relevance, or integrating both. That coding decision is intersubjectively verifiable in a way that qualitative inferences about underlying cognitive states might not necessarily be. This approach puts into practice the verbal analysis tradition described by Chi (1997), which provides a principled basis for transforming qualitative observations about the content of verbal utterances into quantifiable, interrater-verifiable codings without reducing the richness of the data to surface-level counting. The content layer is the appropriate layer from which to build upward from a computing education perspective: identifying whether an utterance addresses Architecture or Relevance provides the interpretive anchor that other analytical layers require. Without knowing what perspective is being communicated, analyses of how the partner model develops (Schaffer et al., 2024), how gestures support the explanation (Lohmer and Kern, 2024), or how collaborative activities structure the dialogue (cf. Fisher et al.,

2023) lack a reference point for interpreting their findings in dual-nature terms¹. The trade-off of this strategy is that the collaborative and cognitive layers of the framework, while theoretically motivated, remain empirically underdetermined within the present work. Independently operationalizing them would require coding process indicators (monitoring, scaffolding, grounding moves) and tracing stance transitions between participants, neither of which the present design does directly. Content patterns are used as proxies for these processes rather than operationalized alongside them. The extent to which this inferential step is warranted is discussed in Section 4.6.

Choosing explanation content as the entry point for qualitative research, however, raises a follow-up question: at what level of detail should the utterances in the data be segmented in the transcripts? This unit of segmentation determines which transitions become visible in the data. The transcripts analyzed across the three empirical studies are segmented into intonation phrases following the conventions established in BB1. Intonation phrases are not arbitrary transcription units: the empirical literature on prosodic segmentation identifies them as reflecting the temporal chunking of cognitive processing, with phrase boundaries corresponding to planning and packaging operations in speech production (Chafe, 1994; Park, 2002; Simpson, 2016). If this correspondence holds, then a transition from Architecture to Relevance content at an intonation-phrase boundary is not merely a change in topic but could be a candidate marker for a shift between mechanistic and teleological modes of construal. Content analysis of intonation-phrase-segmented transcripts thus provides structured empirical access not only to what is said but, indirectly, to the cognitive and collaborative processes that the framework predicts. BB1 establishes the viability of this approach by demonstrating that dual-nature coding at the intonation-phrase level achieves acceptable intercoder reliability, confirming that the segmentation grain is appropriate for the theoretical distinctions the framework requires. The empirical *Building Blocks* BB2–BB4 implement this strategy.

The main contributions of the dissertation are: (1) a synthesized theory, labeled the Co-Constructive Dual Nature Framework, (2) one methodological review paper elaborating on the transcription segmentation, and (3) three empirical study papers investigating different hypotheses addressing the research questions presented in Chapter 1. Figure 3.1 provides an overview of the relationship. All

¹I worked with Michael E. Schaffer and Vivien Lohmer within the TRR318-Ao4 project. Josephine Fisher was a colleague in a related project (TRR318-Ao1). Throughout my research, we shared ideas and had many discussions how our research and findings relate. As they all researched on explanations with the same explanandum Quarto, their findings are highly relevant to this dissertation.

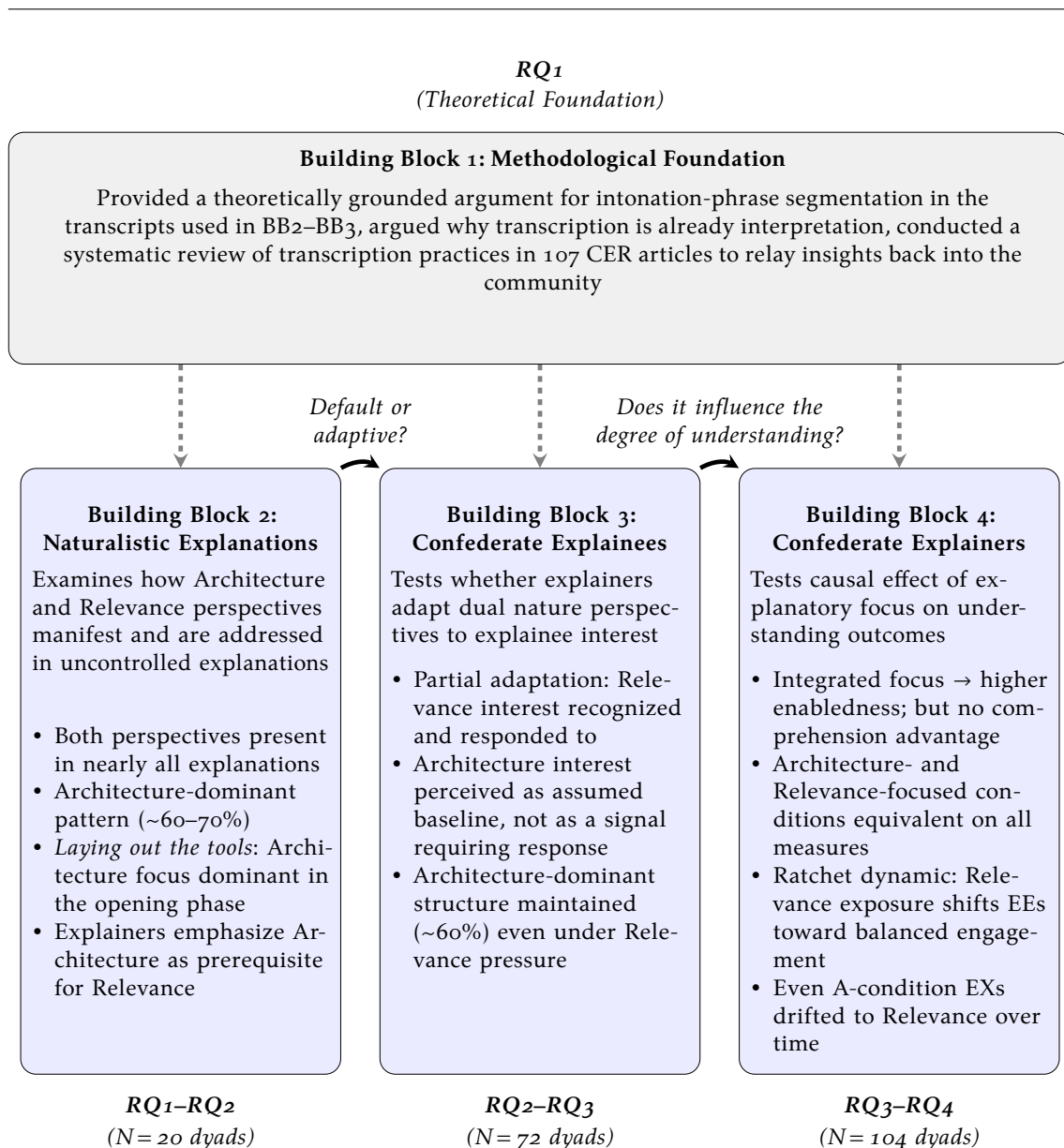


Figure 3.1: Overview of the four building blocks, their interrelations, and their relation to the research question posed in Chapter 1.

studies and contributions were conducted as part of project Ao4 within the TRR 318 *Constructing Explainability*². The Co-Constructive Dual Nature Framework provides the theoretical spine across all four Building Blocks.

The central explanandum is the board game Quarto. The choice of a board game rather than a familiar digital artifact was motivated by pilot observations: investigations using artifacts known to both participants, such as messaging applications,

²<https://trr318.uni-paderborn.de/en/projects/a04>

produced reluctant dialogue in which participants appeared to assume they already understood the artifact or were reluctant to admit they did not. Board games unknown to the explainee eliminated this effect, creating authentic knowledge gaps that motivated sustained co-constructive engagement.

A note on the cumulative format is warranted. The four Building Blocks were conceived as part of a coherent research program but written and submitted at different stages of the project, and the theoretical framework matured over this period: the Co-Constructive Dual Nature Framework reached its current form only after the empirical findings of BB2 and BB3 had been analyzed, and the comprehension/enabledness distinction central to BB4 emerged from interdisciplinary work that postdated BB2's publication. Consequently, the published papers do not always use identical vocabulary or foreground the same aspects of the framework. However, the chapters framing the papers (Chapters 2,4–5) situate each contribution within the framework as it now stands.

The four Building Blocks form a progressive sequence, moving from naturalistic observation to experimental investigation on a common methodological foundation, with each block building on the analytical tools and empirical baseline of its predecessor.

3.1 Building Block 1: Methodological Foundation

Building Block 1

L. Terfloth et al. (Apr. 2025). "Transcription in Computing Education Research: A Review and Recommendations". en. In: *Informatics in Education*. ISSN: 1648-5831, 2335-8971. DOI: [10.15388/infedu.2025.09](https://doi.org/10.15388/infedu.2025.09)

3.1.1 Summary

Transcription is rarely treated as a theoretically consequential step in qualitative computing education research (CER). This paper argues it should be. Drawing on Ethnomethodological Conversation Analysis (EMCA), it develops the position that transcription is not a neutral preparatory act but an interpretive process that actively shapes what phenomena become analytically visible. To empirically ground this argument, the paper presents a systematic review of 107 qualitative CER articles, finding that no article provided full disclosure of its transcription practices with 71% disclosing insufficient details. A case example from the Quarto study illustrates the practical stakes: initial coding using verbatim transcripts produced unacceptably low intercoder reliability ($k_{pre}=0.22$, $SD_{pre}=.17$), a problem that was resolved by switching to the GAT2 transcription system (Selting et al., 2009), which segments speech into intonation phrases rather than sentences. This change yielded a substantial and stable improvement ($k_{post}=0.76$, $SD_{post}=.03$). Based on these theoretical and empirical findings, the paper proposes three steps for selecting, applying, and communicating transcription standards in CER.

3.1.2 Authors' Contributions

Lutz Terfloth conceptualized the paper for the computing education research community, developed the theoretical argument connecting transcription theory to the study's research interests, designed and conducted the systematic literature review, led the case example analysis, synthesized the findings into a three-step framework for selecting, applying, and documenting transcription standards, and wrote and finalized the majority of the manuscript, and revised the paper in response to peer review. Vivien Lohmer contributed expertise in EMCA and interactional linguistics, wrote background sections on spoken language and transcription systems, and provided critical feedback throughout. Friederike Kern contributed expertise in EMCA and linguistic theory, supported and extended Vivien Lohmers background sections, and provided critical feedback on the theoretical framing. Carsten Schulte contributed to the framing of the paper within computing education research and provided feedback on drafts.

Lutz Terfloth did greater than or equal to 75% of the work.

3.1.3 Author's Commentary

The methodological choices underlying the empirical studies in this dissertation are not self-evident, so this paper explains and justifies them. This building block, in a sense, can be considered the methodological umbrella floating above the empirical experiments.

Before the empirical studies could be conducted, a foundational decision had to be made: which transcription system would allow the dual nature content of explanatory utterances to be coded reliably? This question turned out to be non-trivial. Initial pilot work with verbatim transcripts produced intercoder reliability scores far below acceptable thresholds ($k_{pre}=0.22$, $SD_{pre}=0.17$), not because the coding manual was poorly designed but because the segmentation units that verbatim transcription imposes—sentences—do not align with the cognitive and prosodic chunks speakers actually produce. Switching to GAT2, a transcription system rooted in Ethnomethodological Conversation Analysis (EMCA), resolved the issue by segmenting speech into *intonation phrases*. These are prosodically defined units whose boundaries have been argued to reflect cognitive chunking in speech production (Chafe, 1994; Park, 2002), and reception (Simpson, 2016). This change of segmentation in the transcripts produced a substantial and stable improvement in intercoder reliability ($k_{post}=0.76$, $SD_{post}=0.03$).

Beyond solving a practical coding problem, the shift to the GAT2 system carries a theoretical hypothesis that is central to this dissertation's methodological argument. If intonation phrases reflect cognitive processes (cf. Chafe, 1994; Park, 2002; Simpson, 2016) — specifically, the units of attention that speakers assemble and listeners process — then segmenting transcripts into these units does not merely improve coding precision. It means that the coded segments are more likely to correspond to genuine cognitive events: moments at which an explainer shifts focus from one aspect of an artifact to another, transitions from addressing its Architecture to its Relevance, or pauses to reformulate in response to an explaineé's confusion. This is the core claim that motivates treating content-level analysis as an empirical entry point into cognitive and collaborative processes: observable content shifts in intonation-phrase-segmented transcripts provide structured empirical access to transitions in the mode of construal that would otherwise require intrusive measurement techniques lessening naturalness of the setting.

The paper includes a systematic literature review of $n = 107$ qualitative computer science education research articles documenting that no article in the sample provided sufficient disclosure of its transcription practices. To recommend a better approach for future computing education articles, a three-step framework for

handling transcripts was proposed. This broader contribution situates the dissertation's methodological choices within a community context: the absence of reflective transcription standards in CER is not a niche problem but a widespread one that needs addressing, and the decisions made here are deliberately positioned as an example for how the community could respond to that gap. For the purposes of this dissertation, the central contribution of the paper is the theoretical and empirical justification of intonation-phrase segmentation as the coding unit for all subsequent studies. Beyond this methodological output, the interdisciplinary work with interactional linguists on the theoretical foundations of transcription deepened my understanding of prosodic segmentation in speech. That understanding informed the design decisions across BB2–BB4 and, more broadly, the theoretical argument developed in Section 2.3 that content patterns can serve as reliable indicators of the cognitive and collaborative dynamics predicted by the Co-Constructive Dual Nature Framework.

Transcription in Computing Education Research: A Review and Recommendations

Lutz TERFLOTH¹, Vivien LOHMER², Friederike KERN²,
Carsten SCHULTE¹

¹Paderborn University, Faculty of Computer Science, Germany

²Bielefeld University, Faculty of Linguistics and Literature, Germany

e-mail: {lutz.terfloth, carsten.schulte}@uni-paderborn.de,

{vivien.lohmer, friederike.kern}@uni-bielefeld.de

Received: February 2025

Abstract. Transcripts play a crucial role in qualitative research in computing education, with significant implications for the credibility and reproducibility of findings. However, unreflective and inconsistent transcription standards may unintentionally introduce biases, potentially undermining the validity of research outcomes and the collective progress of the field. In this article, we introduce transcription as a theoretically guided process rather than a mere preparatory step, illustrating its role using a case example. Additionally, through a systematic review of 107 qualitative research articles in computing education, we identify widespread shortcomings in the reporting and implementation of transcription practices, revealing a need for greater intentionality and transparency. To address these challenges, we propose a three-step framework for selecting, applying, and documenting transcription standards that align with the specific context and goals of a study. Rather than advocating for overly complex, one-size-fits-all transcription strategies, we emphasize the importance of a context-appropriate approach that is clearly communicated to foster trust and reproducibility. By advancing a more robust transcription culture, this work aims to support computing education researchers in adopting standards that enhance the quality and reliability of qualitative research in the field.

Keywords: literature review, transcription systems, qualitative research, conversation analysis.

1. Introduction

Qualitative research is a crucial methodology regularly used by researchers in computing education (CER). Although not always the case, the empirical data are transcribed to be analyzed in subsequent steps. However, *transcripts are not neutral representations of spoken language* because transcribing is neither simple nor objective (Davidson, 2009); they can even lead to misinterpretation of the data (O’Connell and Kowal, 1995b). The phrase “I never said she stole my money”, for example, has seven differ-

ent interpretations, depending on which word is stressed (Rudzicz, 2016). In qualitative social sciences, transcribing oral data is not just a preparatory step but also already part of the analysis. For that, numerous transcription systems for different analytic purposes were developed (cf. overview of systems by O’Connell and Kowal (2009), a system for transcribing talk-in-interaction – GAT2, by Selting *et al.* (2011)). The underlying motivation for the development of each of the respective systems differs, yet all strive towards the same overarching goal: improving the reliability and quality of the data analysis through the definition of transcription systems or standards. In contrast to such rigor, other fields consider transcription merely as the transformation of spoken into written data. Often, the choice of a transcription system is not critically evaluated in relation to the research questions, and detailed information about the chosen system is regularly missing from the methods sections of papers (Point and Baruch, 2023).

In CER, qualitative research is used, for example, to assess beliefs of computer science (CS) teachers (Bender *et al.*, 2016) or to analyze dialogues during pair-debugging (Murphy *et al.*, 2010). These transcripts are typically created before the analysis begins. In this instance, verbatim transcription remains the most prevalent approach, yet it is rarely specified with sufficient precision. Its typical implementation focuses mainly on semantic content and may overlook critical elements in computing education contexts. This shortcoming becomes especially pronounced in contexts where technology-mediated interactions (e.g., cursor movements in coding environments) or multimodal learning behaviors (e.g., non-verbal communication, such as gestures during pair programming) should be analyzed.

Ethnomethodological Conversation Analysis – an approach to the study of social interaction employed both in (interactional and anthropological) linguistics and sociology – has a long tradition of putting conversational interactions in writing. EMCA also has a long-standing tradition of discussing different transcription systems and adapting them to varying scientific research interests. As a research field, EMCA is primarily interested in the *how* of social interactions (Goodwin and Heritage, 1990), analyzing everyday interactions between people (Sidnell, 2012). For example, EMCA might explore how two participants initiate an explanatory interaction or *how* and when they use minimal feedback such as audible ‘*mhm*’s, hand gestures, head movements, or eye gaze for signalling or displaying understandings and misunderstandings to one another. For this kind of analysis, established transcription standards within the community are important because it is critical that transcriptions can be reproduced and that transcript-based interpretations can be understood by everyone in the community.

Drawing on ethnomethodological conversation analysis (EMCA) theory, we emphasize that transcription constitutes an inherently interpretive process that actively constructs analytical possibilities. Within CER, the choice of transcription system needs to be reflected and discussed. We experienced firsthand how habitually choosing a verbatim transcription system without reflection involves the risk of negatively impacting results. We argue that a shared understanding of transcription theory is beneficial for the community. Moreover, the development of standardized transcription strategies to improve qualitative research in our field is desirable.

This paper addresses that gap. Its main contributions are (1) a theoretical introduction to EMCA transcription theory; (2) a case example from a recent study to outline how the change of a transcription system improved the accuracy of the analysis; (3) a review of current transcription practices in our community; and (4) a practical three-step guideline for choosing, using, and communicating the transcription system in one's own research. We include a literature review of qualitative research articles from our community to empirically support our argument that CER needs better transcription standards. Readers may choose, depending on their interest, to read the theoretical foundations or to skip to the literature review, depending on their interests. This paper follows a recent trend of publications that review community practices (Sanders *et al.*, 2023; Oleson *et al.*, 2022). Importantly, we do not advocate for always choosing the most complex transcription system. Instead, we recommend a context-sensitive, reflective selection process and transparent communication of transcription choices within CER articles. The proposed three-step framework offers a structured, practical guide to support researchers in our field.

2. Background: A Theory of Transcription

In qualitative research, transcripts and transcribing are crucial, as it allows researchers to analyze and interpret the content of spoken interactions efficiently. However, oral conversations contain richer information, and much information is usually lost when a conversation is transferred into a written form. Generally, transcripts reduce the available data (such as video footage) in terms of complexity and are thus not to be understood as neutral representations of the original material. Instead, “the process of transcription generates the data upon which the analysis is built.” (Ayaß, 2015, p. 510) Transcribing is not just transforming (video-)recorded data into written text. In verbatim transcripts, for example, information about pauses, rhythm, intonation, dialect, or about gaze, gestures, or posture (for video data) is omitted. Therefore, as such information might be useful later during the analysis and interpretation of the data, the choice of whether to include it should be (a) carefully considered and (b) transparently documented.

To be able to preserve such information in written form, great effort has been put into the development of different transcription systems, especially since Ochs (1979) introduced the idea of *transcription as theory*. Some systems define a notation system for gestures, eye movements, and prosodic information (Davidson, 2009). More complex systems also include video frames. All efforts share the same goal: to systematically include paraverbal (e.g., pitch, volume, intonation) and nonverbal (e.g., gestures, gaze) elements in a standardized manner. In the 1990s, interest in these systems began to arise among the EMCA research community (O’Connell and Kowal, 2009, p. 241). Today, various standardized transcription systems are in use (Romero *et al.*, 2002, p. 620).

In the CER community, qualitative research methodologies are enjoying growing popularity (Fitzgerald *et al.*, 2011), referencing and reflecting on the choice of transcription system has yet to become standard practice (cf. Section 4). In the next section, we will therefore first elaborate on the theoretical background of the transcription of oral

Table 1
An overview of commonly used terms in linguistics and EMCA.
These are used throughout the paper and provided here for a quick reference

Word	Definition
GAT2	An abbreviation for 'Gesprächsanalytisches Transkriptionssystem' (translated: discourse and conversation-analytic transcription system); a system for transcribing speech in conversational interaction developed by (Selting <i>et al.</i> , 2011).
Intonation	The variation in pitch that conveys information beyond the literal meaning of words, including aspects such as sentence type (e.g., statement, question) and emotional expression.
Intonation Phrase	A segment of speech characterized by a single intonation contour, typically corresponding to a clause or phrase.
Notation	The symbols usable in a transcript to represent aspects of speech, for example, timing, pitch, and intonation.
Prosodic Information	An umbrella term used to describe information that is contained in pitch changes, intonation changes, pauses, or rhythm.
Pitch	The perceived frequency of a sound, often associated with the highness or lowness of a tone. In speech, pitch variations convey prosodic information such as stress, intonation, and emotion.
Transcription System	A set of conventions and notation signs used to represent spoken language in written form, which includes conventions for how, for example, pauses or simultaneous speech are represented in the transcript. Typically, transcription systems also include notations for phonetic, orthographic, and prosodic elements.
Paraverbal Elements	Vocal characteristics separate from the actual words, including tone of voice, volume, speed, and rhythm. These elements can convey additional meaning, emotion, or intent besides the literal content of the words (Mandal, 2014).
Nonverbal Elements	Communication cues (excluding words), such as facial expressions, body language, gestures, eye contact, and physical distance. These visual and tactile elements provide context and can significantly alter the interpretation of verbal communication (Frank <i>et al.</i> , 2015).

speech. To discuss why creating transcripts is always a *first step of data analysis*, the difference between spoken and written language is discussed. Subsequently, different forms of representing oral speech in written transcripts (sentences, utterances, intonation phrases) and their implications for data analysis are explained. Finally, we provide a practical overview of three different transcription systems that may be useful for various research interests in the CER field. Table 1 provides an overview of important terms used throughout the section to support readers unfamiliar with the vocabulary of EMCA.

2.1. *Spoken and Written Language: A Perspective from Linguistics*

Even disciplines that deal with language in its various forms, such as (interactional) linguistics or linguistic anthropology, have long neglected the differences between spoken and written language, even though they have important implications for transcription theory and researchers using transcripts. The differences become more obvious when

comparing spontaneously produced speech (e.g., during a conversation) versus speech produced by reading from a teleprompter. Imagine people speaking spontaneously. They rarely produce what would be considered a syntactically complete sentence in written language. In contrast, a person reading from a teleprompter will produce syntactically complex sentences, sometimes even with many subordinate clauses, when they are in the written template. But individuals reading out loud – at least when they are not professionally trained – sound rather artificial as prosodic features (e.g., rhythm, stress, pitch) differ considerably from speech produced spontaneously. In some instances, these differences may even lead to understanding problems when trying to follow the argumentation or content (Carlson, 2009).

The reason for this artificiality lies in the differences in how *spoken* and *written* language are produced and perceived. Spoken language is volatile, dialogic, and prototypically embedded in an environment of face-to-face situations. In such situations, all participants share time and (physical) space, which enables them to perceive each other through various sensory modalities (Stein, 2018; Roberts and Street, 2017). The dialogic nature of spoken language blurs categories of speaker and listener because they constantly change during an exchange; they become even more blurred when participants speak simultaneously, interrupt each other, or provide continuous verbal and non-verbal feedback. Additionally, spoken language features more variation, for example, across regions, social classes, or speech situations. Written language, on the other hand, is codified and rather standardized. It is also persistent and able to be archived, monologic, and relatively context-free. In written language, the modalities that are missing (intonation, stress, gesture) are compensated for by higher verbal explicitness. Differences between spoken and written language can also be found in the area of grammar and syntax, with a tendency for spoken language to feature less complex and often incomplete clause structures. There is plenty of empirical evidence that the grammatical structures of spoken language have their origin in its volatile and dialogic nature (Couper-Kuhlen and Selting, 2018; Auer, 1992).

In summary, creating transcripts requires more than simply converting spoken language into text. Instead, transcription systems are tools for preserving the volatile information within conversations and for capturing their transient nature. They provide guidelines for the transfer of audible (intonation, stress, rhythm, speech rate) and visible (gestures, facial expression, body posture) linguistic features into a written transcript. Depending on the research questions, preserving certain volatile information can be beneficial. For example, when analyzing classroom discourse or learners' retrospection during think-aloud techniques, often not only the content but especially the process of comprehension and the formation of ideas and beliefs are of interest. To support coders in being accurate, objective, and reliable in their interpretations, a suitable and well-considered choice of transcription system – one that preserves crucial information within the recording – is used.

In the next section, we will reflect on the segmentation of spoken language for the purpose of offering alternatives to the linguistic category *sentence*, which is basically a unit of written language. There are other units that are more useful when transcribing dialogic spoken language.

2.2. Segmentation of Spoken Language in Transcripts

As spoken language rarely contains grammatically and semantically complete sentences, transforming spoken language into sentence-based written language is a challenge that is seldom reflected upon outside of linguistics or EMCA. However, during transcription, the decision of when a sentence starts or stops is challenging, as the syntactic structures are systematically adapted to the needs and specifics of dialogic interactions (Auer, 1992). Indeed, research has compellingly shown that people segment speech differently from written language. Therefore, different approaches for segmenting spoken language in transcripts have been developed. Instead of using sentences, speakers naturally chunk their speech into so-called *intonation phrases* by grouping interpretable units together using mainly prosodic means (such as stressed syllables, intonation contour, and pitch) (Speer and Ito, 2009; Selting et al., 2011). Du Bois et al. (1993) define an intonation phrase as “[r]oughly speaking, [...] a stretch of speech uttered under a single coherent intonation contour. It tends to be marked by cues such as a pause and a shift upward in overall pitch level at its beginning, and a lengthening of its final syllable.” Speakers use intonation phrases to express meaning, highlight and chunk information, and implement and signal discourse structure (Selting, 2000; Selting et al., 2011). Additionally, intonation phrases often align with syntactic units within discourse, which are generally shorter than sentences. In doing so, they convey linguistic information such as focus, signals of completion or continuation, in addition to semantic information (Chafe, 1994; Selting et al., 2011; Bergmann and Mertzlufft, 2009).

Understanding intonation phrases is therefore of significant importance in EMCA-grounded linguistics, given their pivotal role in conveying subtle layers of meaning, outlining speech structures, and enriching the overall communicative effectiveness of spoken language. For computing education studies, using intonation phrases for the segmentation of speech has certain implications and offers various potentials. Empirical evidence indicates that intonation phrases reflect cognitive processes to a certain degree. According to Chafe (1994), the amount of information within an intonation phrase is limited not only due to physical constraints (such as breathing) but also cognitive limitations. In a thorough study of intonation phrases and their relation to cognition, Chafe (1994) argued that human physiology and cognition are interrelated. Park (2002, p. 639) provided an in-depth overview of arguments for why intonation phrases may also serve as cognitive units. In her dissertation, Simpson (2016) tested this hypothesis and found empirical evidence supporting it. She also concluded that intonation phrases are ways to “break up the continuous speech stream into processable portions” and are therefore important not only for the *production*, but also the *comprehension and processing* of language by recipients (Simpson, 2016).

In summary, there is overwhelming empirical evidence that intonation phrases reflect the focus of attention of the speaker and are also important for the processing of information by listeners. These findings can be beneficial for computing education researchers, particularly if their research is interested in, for example, how an understanding of a digital artifact develops throughout an explanation. When researchers ask questions that go beyond *what* was said (content) and are instead interested in the *cognitive and interac-*

tive processing during, for example, explanations, transcripts that segment speech into intonation phrases can improve the analysis. These units provide analysts with semantic chunks that align with cognitive processes and serve interactive functions, which can increase the accuracy of the analysis. In Section 6, we describe a method to easily identify prosodic units. Besides considering segmentation in transcripts, choosing the appropriate transcription system also requires attention.

2.3. Transcription Systems

Section 2 elaborates on how different transcription systems support different research interests. The systems vary mostly in their complexity, depending on how much of the additional information carried by spoken language they want to capture or emphasize (O'Connell and Kowal, 2009, cf.). The most important aspect for selecting a system is the level and types of details required for analysis. Clearly, there is no all-in-one solution when it comes to choosing a transcription system; more detail is not always better. However, to provide an overview, we introduce three transcription systems, ordered by increasing level of detail: (1) standard orthography (verbatim) or content-based semantic transcription using standard orthography, (2) GAT2 or Jefferson, and (3) multimodal transcription.

Verbatim transcripts use standard orthography and follow a rather simple set of rules (see Fig. 1 for an example); Kuckartz and Rädiker (2019, p. 42) provide a comprehensive overview of the most common rules. The goal of verbatim transcripts is to primarily preserve the content or semantic meaning of the spoken dialogue. In their simplest form, verbatim transcripts smooth dialects, remove laughs and affirmations, and include neither pauses nor parts of simultaneous speech. Punctuation marks are used to indicate the end of an idea or aspect. Each speaker's contribution is placed in a separate paragraph, preceded by an abbreviation indicating who spoke (e.g., I: for interviewer). Syntactical errors, discontinuations, or interruptions are smoothed over, resulting in a certain – and often unreflected – bias based on written language segmentation ('clauses' or 'sentences') and norms ('syntactic completeness'). The aim of this tran-

001	EX	For example, small, dark, hollow and round.
002	EE	Mhm.
003	EX	Exist only once. And then I start, for example, and give you any of these pieces. And you have to decide, where on the board you want to put it.
004	EE	Yes.
005	EX	And it always goes back and forth like this.

Fig. 1. In this transcript, Examinee (EE) and Examiner (EX) are engaged in an explanatory interaction. This transcript is a typical example of the verbatim transcription system that uses sentence structures for the segmentation of information.

scription system is easy readability and preservation of content-semantic aspects of the recorded interaction. It is the simplest form, easy to learn and easy to apply. The choice of verbatim transcripts is justified if the research is interested in the content or semantics of the empirical data. Therefore, verbatim transcripts are a suitable choice for methods of qualitative content analysis as described by, for example, Kuckartz (Kuckartz, 2014). However, “[i]n any case in which a speaker deviates from standard pronunciation, the transcription will clearly have a loss of information if that deviation cannot be represented in standard orthography” (O’Connell and Kowal, 2009), potentially hindering accurate interpretation of data.

Therefore, if research interests go beyond *what* was said at the surface, and the goals of interpretation require more detail, different systems may be necessary (see Fig. 2 for an example). Transcription systems such as **GAT2** (Selting *et al.*, 2011) or **Jefferson** (Jefferson, 2004) offer ways to include prosodic features such as intonation or stress¹. They work based on the principle that a transcript can be extended by various levels of detail. Necessary categories of a transcript are the sequential structure of the interaction, pauses, breathing (when communicatively relevant), elision of words, interruptions, and prosodic features like pitch and stress. Instead of using sentences for chunking information, intonation phrases are used.

Each line in the transcript represents one intonation phrase (see Section 2.2). Everything in this transcription system is written in lowercase by default. Exceptions are the prosodically marked or stressed syllables, which are represented by capital letters. Punctuation is used to indicate pitch movement at the end of each intonation phrase (rising (.),

```

001 EX   for exAMPLE,
002      SMALL dark (.) hollow and
003      round;
003 EE   hm_hm
004 EX   exist only ONCE.
005      and THEN;
006      I start for example,
007      and give you ONE,
008      anyONE of these pieces,
009      and YOU have to decide,
010      WHERE on the board;
011      you [(.)      ] WANT to
011      place it.
012 EE      [ye_es,      ]
013 EX   and it always GOES back and
013      forth like this;

```

Fig. 2. A GAT2 transcript of the same segment as before. The inclusion of pauses, simultaneous speaker contributions and intonations creates a different impression of the content.

¹ We will not distinguish between the two here. The Jefferson transcription system includes prosodic features (pitch, stress, intonation), though it segments speech differently. The system is widely used in the international EMCA community. Within the German community of interactional linguistics and EMCA, the GAT2 transcription system is more commonly used. GAT2 is especially interesting because it includes the segmentation of spoken language into intonation phrases. Our example uses GAT2 (see Table 2).

strong rising (?), falling (;), strong falling (.), or level (-)). Pauses are marked according to their length (.) for micro-pauses up to a length of .2 seconds, (..) for pauses up to a length of .3 seconds; pauses longer than .3 seconds are described in numbers).

The third, and most detailed, transcription system is a **multimodal transcription system** (see Fig. 3 for an example). The transcripts include all notations mentioned for GAT2 but additionally include multiple layers of multimodal behavior, such as gestures, facial expressions, or body movements. The amount of additional information about multimodal behavior included depends on the research interest. Transcripts can be a combination of GAT2 or Jefferson's transcription rules and notations of multimodal behavior following the conventions introduced by Mondada (2018, 2019, 2011). Recently, it has become more common to insert timestamped video frames to avoid long verbal descriptions of multimodal behavior. Such systems are suitable for research focused on

001	EX	for exAMPLE,
002		SMALL dark (.) hollow and round;
003	EE	hm_hm
004	EX	exist only ONCE#.
005		and THEN;
006		I start for example,#
007		and give you #ONE,
008		anyONE of these pieces,
009		and YOU have to decide,
010		WHERE on the board;
011		you [(.)] WANT to pla#ce it.
012	EE	[ye_es,]
013	EX	and that always GOES
	EX-ges	#####
		prep-D
014	EX	back and #forth; like this
		#0:01:15.036
	EX-ges	@EX - - - @EE #####
		stroke-D retr-D

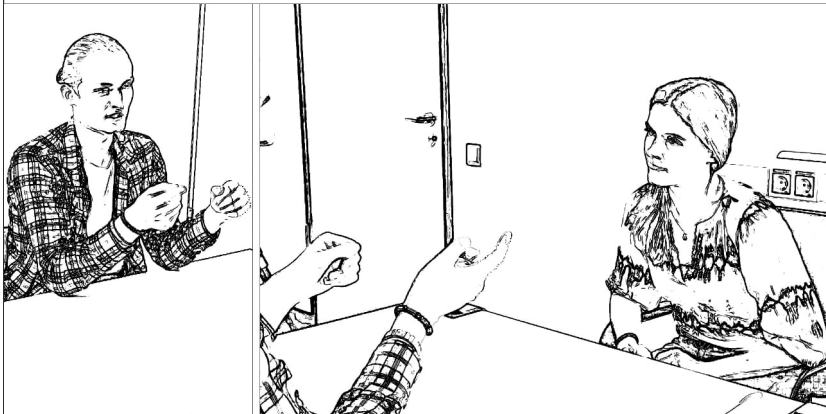


Fig. 3. A multimodal transcript of the same segment as before. Besides the content, stress and intonation, it also has some instances in which gestures are annotated. All points in the transcript that contain a # are moments at which a video frame was extracted to be used in the analysis. Here, only one video frame is provided to serve as an example of how this would look.

the fine details of multimodal behavior in interactions. However, the transcripts can be rather difficult to read because they contain a lot of information. Indeed, finding a compromise between readability and level of detail is an ongoing challenge when transcribing a video sequence.

In conclusion, regardless of the theoretical complexities, it is always sensible to reflect the value of including para- and nonverbal elements in transcripts. Certain paralinguistic elements are helpful for interpreting emotions, while others are rather ambiguous. For example, rapidly flowing speech is a marker of positive feelings, whereas stuttering is a sign of negative feelings. Laughed words, for example, are rather ambiguous: they may hint at a person feeling shame but can also simply be a sign of joy (for more interpretations, see Table 1 and Table 2 by Bloch (1996)).

Concerning prosody, “[p]rosodic elements such as intensity, pitch accent (i.e., the pattern of low and high tones used in a stressed word), and intonation, have been suggested to aid in conveying emotional affect (e.g., happiness) in acted speech.” (Olsen, 2019). Pause duration and pause occurrence have been found to “consistently mark narrative section boundaries, thus suggesting that pause is a very important structuring device in oral narratives” (Oliveira, 2002). Therefore, these elements of information are often important for *truly* understanding the data beyond its content. It can already be helpful if STRESSED syllables are capitalized in verbatim transcripts to improve the quality of the analysis. Ultimately, transcripts are tools that support us as researchers and can thus be adapted to specific needs. To connect all these theoretical aspects with practice, we will elaborate on the practical experiences and issues we have faced in the following section.

3. Case Example: Researching Naturally Occurring Explanations

Our example and experiences originate from a research project that focuses on analyzing how people engage in a naturally occurring explanation of a technical artifact (Terfloth et al., 2023). The overarching research interest was to find explanation patterns and strategies useful for computing education, with the aim of being able to construct understandable synthetic explanations in eXplainable Artificial Intelligence (XAI). The experiments that elicited close-to-natural explanations in a dialogic setting (explainer and explainee) formed the foundation for analyzing how understanding is monitored and scaffolded by the explainer to ensure the explainee’s understanding. The aim was to trace how human explainers dynamically and interactively address human learning when explaining a technological artifact.

Twenty explanatory interactions involving a technical artefact – the board game *Quarto!*² – were analyzed. Choosing a simple artifact was sensible to gain initial insights that provide a useful basis for further research into how more complex digital artifacts are explained in everyday settings. The explainers-already familiar with the

² What follows is a rather compact description of the study and theoretical foundation. For more details regarding the theory, as well as the complete study, see Terfloth et al. (2023).

game-were instructed to explain the game such that the other person would have a realistic chance to win if they played. The final video dataset includes recordings from 20 laboratory studies, in which 20 EX explained the game to 20 EE (19 male, 18 female, and 1 non-binary). Participants' ages ranged from 18 to 39 years ($M = 24.92$, $SD = 4.42$). Most (36) had an academic background, with 35 identifying as students across various disciplines such as engineering, education, economics, law, computer science, media studies, and linguistics, while two were full-time employees. Among the EX, seven had prior experience explaining the game. Gameplay experience among EX varied between 0 and 18 rounds ($M = 5.46$, $SD = 5.18$), and ten reported having general experience in providing explanations (e.g., from tutoring). Each study session was designed to last between two and three hours, including all pre-and post-assessments. The explanations lasted between 02:23 and 16:17 (mm:ss; $M = 07:24$, $SD = 03:22$). As is typical in CER, we transcribed the data using standard orthography in a smoothed, verbatim format, followed by coding the transcripts using content analysis. The transcripts included only semantic information and omitted stress or paralinguistic components (such as pauses and laughter).

The coding manual was based on a deductive code system derived from the dual nature theory of artifacts, according to which artifacts can be described either by their architecture and/or their relevance (Kroes, 1998; Schulte, 2008). In explanations, either aspect can be addressed, as artifacts are designed to be *means* to certain *ends*. Therefore, when coding, we identified at which points in the explanation each of the two sides was addressed by the participants (explainer and explainee). This allowed us to assess whether one side of the dual nature was addressed more frequently, and which of the two sides was addressed first. According to dual nature theory, for a holistic understanding of the artifact, both sides need to be addressed and understood.

3.1. Issues with Verbatim Transcripts

Initially, we did not reflect on whether verbatim transcripts were the right choice for our research questions. However, our intended analysis was affected in such a major way that (a) a reliable analysis was not possible, (b) we therefore had to change the transcription system, and (c) we were subsequently motivated to assess our community's practices and reflect on the implications.

Throughout our studies, we tested different iterations of the coding manual in pilots. In these pilots, two independent coders (a student assistant and the first author of this paper) coded a set of 10 verbatim transcripts from the pilot studies. See the example in Fig. 1 in section 2.3 for the verbatim transcript excerpt. However, the intercoder reliability was too low ($k_{pre} = 0.22$, $SD_{pre} = .17$). Following typical conventions, we tried improving the coding manual to address the issue by including better examples, elaborating on corner cases more clearly, and using clearer wording. However, all measures taken were unsuccessful, as the intercoder reliability remained unsatisfactory.

To gain a more profound understanding of the origin of these issues, both coders compared all coded segments in an intercoder session and identified some root causes.

The main cause leading to the highest number of disagreements was that coders defined the boundaries of a coded segment (i.e., architecture or relevance) differently. In the coding manual, the minimal coding unit was specified as “at least one word.” A typical example of disagreement was an instance in which one coder decided to include a word or phrase in a coded segment that the other coder chose not to include. During the intercoder sessions, many of these disagreements could not be resolved, as different boundaries for a code were due to different plausible interpretations of specific parts of the transcript. Ultimately, even though large parts of these coded segments contained nearly identical semantic information, they could not be counted towards agreement due to a lack of overlap percentage. In summary, in most instances, consensus was not reached.³

The switch from verbatim to the GAT2 transcription system resolved this issue. See the example in Fig. 2 in section 2.3 for the GAT2 transcript excerpt. It ultimately allowed us to follow the study’s initial research interest instead of prematurely rejecting it. As mentioned above, GAT2 transcripts segment speech into intonation phrases, which are smaller units than sentences in writing. The coding manual was altered so that the minimal coding unit was defined as *at least one intonation phrase* (i.e., one line in the transcript). The same two independent coders coded the final set of 20 GAT2 transcripts using the refined manual. This improved the intercoder reliability significantly and especially reduced the standard deviations (from $k_{pre} = 0.22$, $SD_{pre} = .17$ to $k_{post} = 0.76$, $SD_{post} = .03$). After finishing coding, both coders reported that, due to clearer rules for identifying the boundaries of a code – enabled by intonation phrase segmentation – boundary identification was now more precise and less ambiguous than before.

Premature rejection of our research interest due to an incorrect choice of, and lack of reflection on, the transcription system would have resulted in a type II error (beta error): we would have failed to detect an effect or a relationship when there actually was one.

3.2. Theory Guided Interpretation

From a theoretical perspective, the improved interpretation based on GAT2 transcripts can be attributed to three main points: (1) inclusion of stress information, (2) segmentation into smaller, prosodic units, and (3) the link between intonation phrases and cognitive and interactive processes.

First, GAT2 includes information about stressed (capitalized) syllables within each intonation phrase, which helps to differentiate certain cases better. For example, the phrase “I never said she stole my money” has seven different interpretations, depending on word stress (Rudzicz, 2016). In GAT2 transcripts, stress can aid in pointing out which side of the dual nature is being addressed in certain moments of the explanations. For example, “and then you decide where to put the piece” would address the process and rules of the game (architecture). “And then you decide WHERE to put” would address that there are different positions that one can use to place the piece strategically (relevance).

³ We coded using MaxQDA. The default setting in MaxQDA counts for agreement if 90% of the segment overlaps.

Second, GAT2 segments information into intonation phrases based on the prosodic characteristics of speech, which improved annotation accuracy for us. Coding the verbatim transcripts reliably was problematic, as there were multiple options for coding single sentences. In some instances, using one, two, or even three codes was sensible. In the case of compound sentences or other complex utterance structures, the boundaries of ideas that receive one code were difficult to assess. This changed with GAT2 due to chunking information into intonation phrases. Because these are based on the prosodic information in the utterances, the boundaries of the semantic chunks were predetermined. In the coding manual, we defined the rule that one intonation phrase always receives one code. Coders thus only had to focus on coding the content, as determining the boundaries of a coded segment was no longer necessary. In comparison with the verbatim transcripts, GAT2's segmentation provided clearer boundaries for coding, making content coding more straightforward and less ambiguous.

Third, empirical evidence suggests that intonation phrases reflect cognitive processes (Chafe, 1994; Park, 2002; Simpson, 2016). During language acquisition, people learn to segment speech into these units by using stress, intonation, and pitch to convey meaning. As our cognitive capacity is limited both for the speaker and the listener; however, there are boundaries to how much one can think ahead while speaking, and there are also boundaries as to how much information a listener can process in one go. For example, organizing – on the fly – which aspect an explainer addresses during an explanation of a game is a challenging task. Thus, if the explainers had reformulated what was said or paused, they might have spotted an inaccuracy or ambiguity in parts of their explanation. Especially these instances often signaled a shift in focus regarding which side of the dual nature was addressed and therefore contained important information for us.

In conclusion, GAT2 transcripts drastically enhanced intercoder reliability in our coding process. The improvements can be attributed to (1) the inclusion of stress information helpful for interpretation, (2) the segmentation based on prosody, and (3) the potential link to cognition. These features facilitated more explicit interpretations, reduced ambiguity, and improved coding reliability, offering a promising approach for analyzing the shifts in explanations. Potentially, due to the intonation-based segmentation, we were able to retrace cognitive processes more accurately when coding GAT2 transcripts, especially in moments when speakers shifted from addressing one side of the dual nature to the other. Regarding other (computing) education research endeavors, this aspect of cognitive chunking could underpin future interpretations in interesting ways.

To assess the existing risk of other researchers running into similar issues, a systematic literature review of articles from our community assesses the community's transcription practices in the next section. Afterward, we discuss whether current practices may lead to similar risks and elaborate on how much detail transcripts need.

4. A Bigger Problem? A Systematic Literature Review

What we experienced may be symptomatic of our community. Phrases such as *sessions were recorded and transcribed*, *both researchers read the transcribed*, *audio-taped in-*

terviews, or the retrospective interviews were recorded with the participants' consent and transcribed in full are epitomes of typical sentences found in the methods sections of papers in our field⁴. To solidify our argument empirically – that CER needs better transcription standards – this section contains a systematic literature review of our community's transcription standards. To this end, we conducted a review of 107 articles from our field. The literature review is guided by the question: to what extent transcription standards are reported in research articles in the CER field (RQ1)?

We acquired literature from five popular CER outlets: ICER, ITiCSE, TOCE, KOLI, and SIGCSE, using the keywords: transcript, transcribe, and transcription⁵. Using the software Publish or Perish, we combined the results for each of the three keywords and

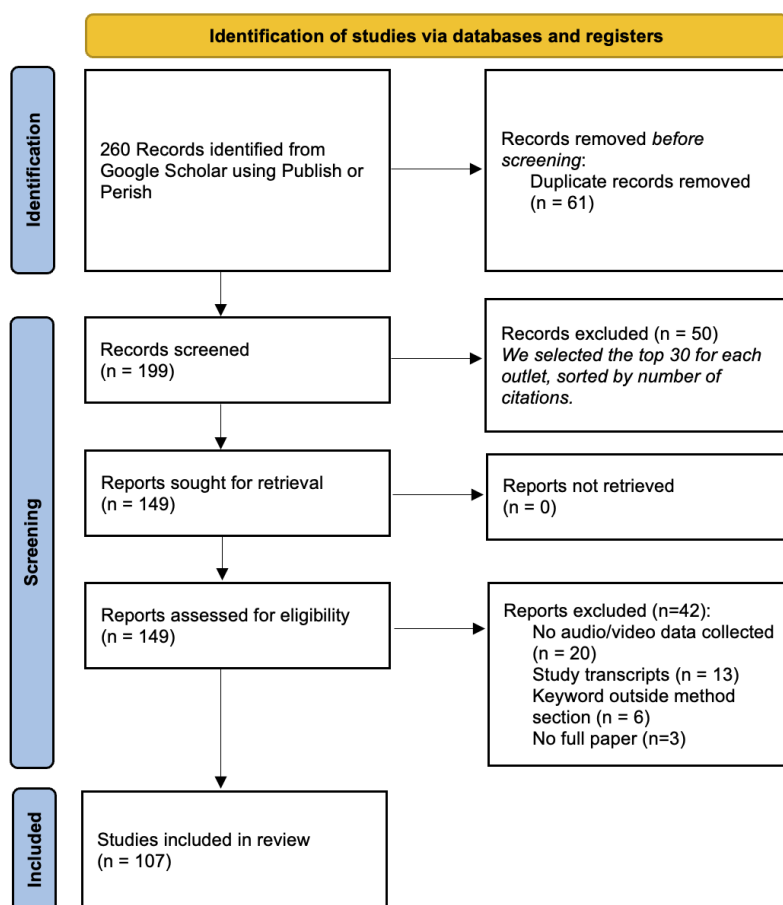


Fig. 4. PRISMA 2020 flow diagram for systematic reviews provides insights into the selection process.

⁴ These sentences are not direct quotes but paraphrases of popular phrases. We believe pointing fingers is not needed and is by no means helpful towards the more general point this paper tries to make.

⁵ See https://osf.io/8ejcu/?view_only=15343ef09ff24e3e98571994294dd484 for the article database and queries. It contains the final list of the reviewed articles. Additionally, it contains quotes of the parts where details were shared about the transcripts.

sorted them by the number of citations. From each of the five outlets, we selected – if possible – 30 unique articles. We screened the articles’ abstracts to ensure that the articles included in the review were qualitative studies that used transcripts. See Section 4.1 for more details on the criteria. This resulted in a total of 107 articles from five popular CER outlets (see Fig. 4).

4.1. Inclusion & Exclusion Criteria

We analyzed research papers from 2015 to 2023 and journal articles published in one of the five outlets and only included papers that were qualitative research *and* included at least one of the words *transcript*, *transcribe*, or *transcription* in the full text. Forty-two articles were excluded from the review for different reasons (see Table 2 for details). For example, 20 articles were excluded because they contained no transcribed data. Additionally, 13 papers mentioned “study transcripts,” which referred to a different type of transcript irrelevant to this article. Six papers only mentioned the keywords outside the methods section, such as in background sections, and did not work with empirical data in the form of transcripts. Finally, three papers were excluded because they were not full articles, such as posters or works in progress. This systematic exclusion process ensured that only relevant studies meeting the inclusion criteria were analyzed.

4.1.1. Threats and Limitations

Regarding researcher bias, the system for categorizing the different articles by the amount of detail shared about transcript development, we would claim that the bias is

Table 2
Summary of exclusion reasons by source and their relative percentages

Exclusion Reasons	Total	Outlet				
		ICER	KOLI	SIGCSE	TOCE	ITiCSE
No data transcribed	20	4 (20%)	5 (25%)	3 (15%)	2 (10%)	6 (30%)
Mentioned “study transcripts”, which are a different types of transcripts and irrelevant for this article	13	2 (15.4%)	2 (15.4%)	1 (7.7%)	4 (30.8%)	4 (30.8%)
The three keywords (transcribe, transcript, transcription) were only mentioned outside the method section of the papers, for example in the background sections. These papers were not working with empirical data in the form of transcripts.	6	0	0	2 (33.3%)	0	4 (66.7%)
Not a full paper (e.g., a poster or work-inprogress)	3	0	2 (66.7%)	0	1 (33.3%)	0
Total	42	6 (14.3%)	9 (21.4%)	6 (14.3%)	7 (16.7%)	14 (33.3%)

mitigated. The category system is fairly simple, and the application of the categories was straightforward. Additionally, two researchers *independently* categorized the papers and subsequently discussed the categorization of the different papers to reduce the risk of human oversight. Both researchers closely read the sections of all articles that provided insights into the transcription strategy.

The representativeness of the sample could be questioned, especially concerning whether the sample can reflect the popular practices of the community. To mitigate this risk, we purposely chose the articles that were well received (i.e., articles were sorted by the number of citations). Additionally, we chose articles from popular venues within the ACM Computing Education community. Moreover, we chose articles from 2015 until June 2024, to reflect current rather than past practices, while leaving some room for potentially identifying improvements and changes throughout these nine years.

4.2. Data Analysis and Categorization

We reviewed each of the articles to assess the details shared about the transcription process. The review focused on whether the article provided details about how the transcription was conducted, including the methods, procedures, and any potential biases or limitations associated with the transcription process. We did **not** review whether the transcription system, if identified in the articles, was suitable for the research questions, as this was beyond the scope of our article. Each article was assigned to one of three categories based on the level of disclosure provided:

- **Full Disclosure:** Articles that provided comprehensive details about the transcription process, including the methods used, the procedures followed, and any potential issues or biases addressed. More precisely, the articles reflected on their choice of system, stated which system was used, and argued how that supported their research objectives or shared the details of the system in, for example, the appendix.
- **Insufficient Disclosure:** Articles that mentioned the transcription process but lacked comprehensive details, leaving some aspects of the transcription process untransparent. For example, they used “verbatim” or “word-per-word” to communicate the types of transcript. However, the choice of system is not reflected, and the details shared are insufficient to reproduce their transcripts. It is unclear what information is included and what information is excluded.
- **No Disclosure:** Articles that did not provide any information about the transcription process and lack transparency of their methods. For example, the articles stated that transcripts were developed without any further details, or even withheld information about the creation of transcripts, even though, for example, transcripts were coded in later steps.

As stated before, each article was reviewed independently by two researchers (the first author and a student assistant) to ensure reliability and consistency in the categorization process. Discrepancies in categorization were resolved through discussion and

Table 3
Summary of articles reviewed and categorized according to the degree of disclosure

Articles	#	Degree of Disclosure		
		Full Disclosure of Transcription Strategies (comprehensive, transparent, detailed, reflective)	Insufficient Disclosure of Transcription Strategies (partial, vague, limited, incomplete)	No Disclosure of Transcription Strategies (opaque, absent, undisclosed, withheld)
ICER	29	0 (0%)	13 (44.8%)	16 (55.2%)
KOLI	22	0 (0%)	4 (18.2%)	18 (81.8%)
SIGCSE*	5	0 (0%)	0 (0%)	5 (100%)
TOCE	27	0 (0%)	6 (22.2%)	21 (77.8%)
ITiCSE	24	0 (0%)	8 (33.3%)	16 (66.7%)
Total	107	0 (0%)	31 (28.9%)	76 (71.0%)

* *Publish or Perish* just listed articles from 2017, even though we searched from 2015 and 2023. We tried to address the issue by changing the outlet name in the queries without success. Addressing the issue would have induced more complexity in our query logic. Thus, instead of further complicating the process, we used these 5 articles from 2017 instead.

consensus. Table 3 lists the absolute and relative numbers of articles within each of the three categories (full disclosure, insufficient disclosure, no disclosure) for each venue and across all venues combined.⁶

For a more profound understanding of each article's practices within the three categories, the wording used to describe the transcription process was assessed. However, out of 107 articles, no article provided information that, according to the theoretical discussion and our experiences, can be considered **sufficient**. Thus, the category **Full Disclosure** contained no articles.

The majority of articles in the category **No Disclosure** – at minimum – inform the reader about the existence of transcripts, for example, by stating that transcripts were coded. Some articles in this category fail to mention that transcripts were created, even though it was implicitly clear that transcription of recorded material was required. In all 76 articles, no details regarding the standards applied were shared. The articles did not clarify which information was preserved in the transcripts (e.g., stress, content, prosody, paraverbal or nonverbal components such as laughter).

The 31 articles in the category **Insufficient Disclosure** mentioned, for example, that data was *transcribed verbatim*. Some mentioned that a professional or automatic transcription service was used, which provided a certain transparency of the overall process. However, they did not provide sufficient details to support readers in understanding which standards were applied and which information was preserved, thereby hindering comprehension of interpretations and results. An example of this category is: “[t]he interviews were transcribed using automatic transcription software and were afterward manually corrected. The transcripts were anonymized and used for subsequent processing” (Aivaloglou and Meulen, 2021).

⁶ The full list of papers is available at https://osf.io/8ejcu/?view_only=15343ef09ff24e3e98571994294dd484

Out of 107 articles, 87 (81.3%) quoted excerpts from their transcripts. This can generally be considered good practice in qualitative research. However, as the articles did not share sufficient information about the standards to which the transcripts were created, it was unclear whether the quoted excerpts were altered for the publication. Consequently, even though the excerpts were quoted, it cannot be determined whether these excerpts were formatted or cleaned for readability, or if the format matched that which was coded in their analysis. Identification of standards would resolve this issue as well.

4.3. Results

In this section, we address the research question RQ1: To which extent are transcription standards reported in qualitative research articles in the CER field? The analysis of articles revealed a significant lack of transparency in the transcription process within qualitative CER articles that base their results on audio or video data (see Table 3 for an overview of the degree of disclosure of the transcription process within the articles).

In summary, out of the 107 articles reviewed:

- A substantial 76 articles (71.0%) disclosed almost no details about their transcription methods.
- 31 articles (28.9%) provided some but insufficient disclosure.
- No article (0%) provided full disclosure of their transcription processes.

These findings indicate a prevalent culture of insufficiently reporting transcription practices in qualitative CER. Consequently, the critical evaluation and reflection of the results within our community is severely impacted. In contrast to practices within EMCA, in our community neither the status of transcriptions in the research process nor their underlying principles of production are reflected, discussed, or made transparent to the reader. While a small percentage of articles share some details about the type of transcript that were used, no articles shared arguments for the choice of transcription system or standards. We therefore claim that the choice is *not actively reflected*, or at least these reflections are not transparently communicated, in articles in our field. Even though we would like to provide some reasons, the review results do not provide any insights into the underlying reasons for this trend.

5. Discussion

In this section, the issues arising from transcription strategies not being communicated in method sections are discussed. By doing so, this paper follows the path of publications in other fields that have also raised the need for better transcription standards (Point and Baruch, 2023). Our analysis highlights the potential to further develop transcription practices within our research community. In particular, fostering greater reflexivity in dealing with transcripts and establishing more standardized ways of reporting transcription strategies could enhance transparency and comparability in research findings. Before diving into specific discussion points, we want to address two issues: The fact that

all the articles reviewed went through a rigorous peer-review process yet still do not provide sufficient details regarding transcription in their method sections indicates room for improvement. Second, the absence of clear communication regarding the standards applied in transcript creation makes it difficult to conduct meta-analyses of our community's research practices. For instance, systematically assessing whether the interpretations and results of qualitative studies in our field are built on solid foundations becomes a challenge. With our pragmatic transcription framework, we aim to provide a useful tool that facilitates a more structured evaluation of spoken data and enhances the overall transparency and reliability of research outcomes.

In CER, verbatim transcripts are the predominant choice. As long as the choice is

(a) reflected upon and (b) transparently communicated in papers, this is not a problem per se. A suitable way to reflect is to ask: How much information does a transcript *need*? Firstly, simply adding details in transcripts for the sake of having detailed transcripts is not recommendable. "One of the important features of a transcript is that it should not have too much information. A transcript that is too detailed is difficult to follow and assess. A more useful transcript is a more selective one" (Ochs, 1979). The amount of detail a transcript needs, therefore, depends on the study's objective and thus requires researchers to reflect on which features of the transcript should be excluded and included. More precisely, aligning the transcript system with the study's objectives requires researchers to think about how the potential exclusion of information may "blind us to other features of language which are equally important to human communication" (Olson, 1993).

In educational contexts, these other features can be relevant for analysis. Collaborative learning is common, and understanding the dynamics of group interactions is, for some research interests, important. In retrospective techniques, retracing an inner debate of someone may not be possible if only verbatim transcripts that contain information about *what* was said are used. While there is definitely a place for research in our field concerned with *what* was said (e.g., identified themes in Fowler *et al.* (2021)), there is also considerable research in our community that is more interested in or additionally focused on *what* was meant (e.g., Alshahrani *et al.* (2018)). Two examples relevant to CER help illustrate this point: First, in the context of block-based programming, a transcription system that includes gestures such as pointing captures the interplay between verbal interactions and visual coding blocks in pair-programming settings. This is potentially relevant, for example, for understanding how students jointly navigate and interpret programming tasks. Second, incorporating prosodic and multimodal data can provide deeper insights when researching how students struggle during pair programming. Capturing pauses or hesitations in speech might reveal moments of cognitive struggle, which can point researchers to significant instances in the data. In these impromptu examples, smoothed, verbatim transcripts would exclude phenomena relevant for analysis; details relevant for interpretation are missing (cf. the seven interpretations of the phrase "I never said she stole my money" (Rudzicz, 2016)). Sometimes capturing such nuances might be directly related to research interests and can serve as a guide. Thus, if not reflected upon, verbatim transcripts may lead to unintentionally ignoring nuances, resulting in inaccurate or incomplete analyses and interpretations.

Concerning the understandability and transparency of qualitative research results, one important aspect needs reflection: In qualitative research, trustworthiness is increased through transcription. Transcripts provide evidence for the analysis (Duranti, 2006). Therefore, if the process of creating the evidence is not reflected upon and made transparent, analytical claims (Ashmore and Reed, 2005) may be based on a weak foundation. As this would generally require more text in publications, Lapadat (2000, p. 217) states, “[w]hen standardized procedures are used, a few words will suffice, but when researchers contextualize and negotiate a method as a means of interpretive seeing, there is no shortcut to explicit description.”

Lastly, we want to discuss an important issue concerning insufficient reflection on the choice of transcription system: Our findings in the case example suggest that unreflected adherence to verbatim transcription could risk prematurely dismissing promising research ideas. This would have led us to overlook valuable insights central to our research proposal in our case example. Therefore, we want to highlight the risks of selecting verbatim transcripts without critical evaluation, particularly the inefficiencies and biases this practice introduces. An unreflected choice of transcription system can potentially cause researchers to dismiss seemingly unreliable coding manuals and doubt their research approach. This risk should motivate us as a community to reflect and adapt current transcription practices.

6. A Pragmatic Framework for Transcribing

In this section, we provide practical recommendations to support researchers in our community in improving their transcription practices. The SIGSOFT standards for qualitative research require authors to “identif[y] data recording methods (audio/visual), field notes, or transcription processes used” (Ralph *et al.*, 2021). For this, a practical yet theoretically sound guideline can be helpful. We condensed various advice found in the literature (e.g., (Point and Baruch, 2023; O’Connell and Kowal, 1995b; Davidson, 2009; O’Connell and Kowal, 1995a)) into a three-step process (see Fig. 5). In the first step, after data acquisition, researchers should – keeping their research questions in mind – choose what information needs to be included in their transcripts. During the development of the transcripts in step two, the rules of the selected system should be followed rigorously to create standardized and comparable transcripts. In the final step, when writing the publication, the choice of transcription system needs to be made transparent and discussed in the context of the research question. Only if the article contains sufficient information will readers be in a position to fully reproduce the research and create appropriate transcripts themselves. Furthermore, the argumentation provides insights for readers to evaluate how the research questions are planned to be answered. Although not part of the corpus of articles we reviewed, Tenenberg and Chinn (2019); Kong *et al.* (2022) provide good examples of how to address transcription practices and can serve as references. In the following subsections, we will discuss further considerations for each of the three steps.

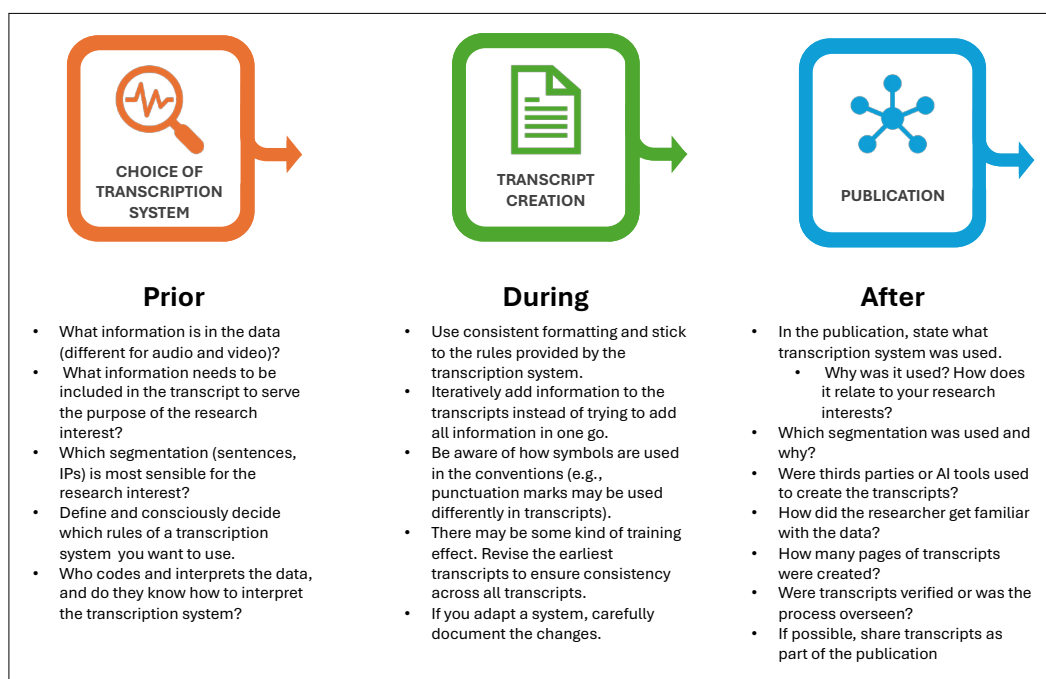


Fig. 5. Flowchart illustrating the steps in the process, with a synthesis of important guiding questions for each step. Depending on the complexity of the qualitative study as well as individual interests, not all aspects need to be strictly adhered to and serve as a guideline supporting the reflection and decision-process.

6.1. The Choice of System – Before Transcribing

The process of transcribing transfers the volatile nature of spoken language in interaction into the readable and reproducible format of the transcript. To pick the right notation for the transcription, it is important to reflect on the research interest and research question to decide which details need to be preserved (Ochs, 1979). To aid in this decision, a few guiding questions can help select an adequate notation system (Spiegel, 2009, p. 7f):

- What is the transcript for? In which context and for what kind of phenomenon do I transcribe?
- What is the research interest?
- Which information must the transcript preserve so that it can serve this purpose?
- What do the readers of the transcripts (e.g., researchers, student assistants) know about transcripts and transcribing? Will they be able to work with them?

For the analysis, it is essential to remember that research assistants may not be familiar with the transcription system. If multiple people will work with the transcript for interpretation and analysis, they should be familiar with the rules and conventions of the chosen system. Especially with more detailed systems, training may be required.

6.2. Creating the Transcripts – While Transcribing

It can be helpful, especially when using more complex transcription systems for the first time, to iteratively improve the level of detail in the transcripts. Rather than aiming for perfect transcripts in one go, it is sensible to start with a simpler initial transcript and refine it through successive iterations. As more complex transcription systems require adherence to a larger set of rules and conventions, keeping all of them in mind simultaneously is often overwhelming and can lead to errors. Instead, one may begin with a simple verbatim transcription. In a subsequent step, pauses and stress are added according to the conventions. Then, intonation phrases may be identified, and the segmentation is changed accordingly, until a refined transcript is ultimately achieved. One final step should always be to revise the transcripts and listen to the source file while reading along with the transcript to spot errors.

Transcribing is a time-consuming task. Nowadays, however, there are helpful ways to increase transcription speed, even when the system is more complex than verbatim. Openly available, locally executable transcription models such as Whisper (Radford *et al.*, 2022) are becoming increasingly popular and effective. Many commercial products for qualitative analysis also provide automatic transcription services (e.g., MaxQDA, Atlas.ti, Transana). In the past, the quality of auto-transcripts was insufficient, and correcting errors often proved to be as time-consuming as transcribing manually. This, however, has changed. Today, models perform well and run locally, so they can be used with sensitive data (if the recorded audio data are of high quality). Even automatic speaker diarization⁷ is supported. However, when using such tools, the question of how one familiarizes oneself with the data appropriately becomes increasingly important. Additionally, these tools tend to smooth the language, often more than standard verbatim transcription does. In other words, there is a trade-off between obtaining transcripts quickly, knowing the data well, and having accurate transcripts, all of which are helpful for improving the analysis. While these tools certainly make the creation of a first, rough transcript almost effortless, there are two aspects that we want to emphasize. First, one should use the time saved by auto-transcripts to better familiarize oneself with the data-perhaps even by extending the transcripts beyond their verbatim nature and including certain paralinguistic elements. Second, these tools can be problematic if researchers are unaware of the issues and potential consequences of choosing an inappropriate transcription system (as discussed in this paper). Critically speaking, research communities could increase their distance from the practice of transcribing, thereby missing out on analyzing rich transcripts and yielding fruitful interpretations.

Regarding segmentation in transcripts, it is sensible for each speaker's contribution to be placed in a single paragraph. If one decides on sentence segmentation, each sentence should contain one coherent contribution. However, it is important to remember that this decision process introduces subjectivity. To address this issue, it

⁷ <https://github.com/huggingface/speechbox/tree/main?tab=readme-ov-file#asr-with-speaker-diarization>

can be sensible to adhere to rules for segmentation (e.g., a sentence ends after every pause of at least 0.5 seconds) and to repeatedly listen to the source data closely while reading along to identify instances that may need correction or discussion with colleagues.

If, however, the research interests require a more sophisticated approach to segmentation, such as intonation phrases, sticking to the GAT2 ruleset is viable but can be rather overwhelming for those outside linguistics. To enable broader use of the benefits of segmenting transcripts into prosodic units such as intonation phrases, alternative methods such as Rapid Prosody Transcription (RPT) – though perhaps not as linguistically rigorous – exist (Cole and Shattuck-Hufnagel, 2016). Although these methods are not based on the GAT2 ruleset, they ensure accessibility for a broader audience to make use of different forms of segmentation. In the first step, all words present in the data are transcribed without punctuation as perceived by one person. Afterward, in two passes, other researchers or research assistants identify (1) prominent words and (2) boundaries of information chunks by listening to the audio while annotating the transcripts from step 1. The method stipulates that they cannot pause or stop the playback; however, they should listen twice per pass. No feedback or sample solution of another transcript should be given to the annotators beforehand, as they should rely solely on their intuition without concern for correctness. For these two passes, they are instructed uniformly (“Mark as prominent words those that the speaker has highlighted for the listener, to make them stand out,” and “Mark boundaries between words that belong to different chunks that serve to group words in a way that helps listeners interpret the utterance” (Cole and Shattuck-Hufnagel, 2016)). Afterward, the annotations need to be tested for their reliability. In the original method, Cole and Shattuck-Hufnagel (2016) provide a sophisticated approach that we deem useful only if, for example, at least five people annotate the transcripts. As that is often not the case, we would argue that intercoder reliability tests and intercoder sessions are also sensible and quick alternatives.

6.3. Transparency in Publications – After Transcribing

When writing the publication, we would strongly recommend reporting sufficient details to increase transparency in such a way that other researchers can produce transcripts of comparable form. It is advisable to share (excerpts of) the transcripts in open-science repositories in anonymized form, not only for transparency but also in the interest of open science.

To quickly share necessary details about the transcription system that was used, it is advisable to simply refer to a pre-existing system by citing it (Kowal & O’Connell provide a good overview in (Flick, 2013, p. 64ff)). If rules of existing systems were combined or adapted, it is helpful to state whether the (1) verbal component (e.g., verbatim, standard orthography), (2) prosodic component (e.g., rhythm, pitch contour), (3) paralinguistic component (e.g., breathing, laughing), or (4) additional information such as gestures or gaze are included, omitted, or adapted.

7. Conclusion

This paper explores the role and standards of transcription in qualitative computing education research (CER). Besides highlighting the need for better standards, it examines how different transcription choices may influence research outcomes. Drawing on Ethnomethodological Conversation Analysis (EMCA), the paper elaborates on transcription theory. It emphasizes that transcription is more than just a preparatory step; it actively shapes analysis and should be critically considered. A systematic review of 107 CER articles was conducted to assess how transcription standards are reported and to identify ways to improve transparency and rigor. Based on these findings, a three-step framework is proposed for selecting, applying, and communicating transcription strategies. A case study illustrates how a structured approach to transcription enabled the pursuit of a research interest that would have been discarded using a different transcription system. Rather than advocating for a one-size-fits-all solution, this paper emphasizes the importance of selecting a transcription method that aligns with research goals and ensuring clear documentation. The aim is to contribute to a more reflective and consistent approach to transcription in CER.

To improve methodological clarity and the qualitative analysis of spoken data in CER: (1) A conscious reflection on the choice of transcription system in light of the research question is required. For example, if one wants to retrace the thought process of someone who was interviewed about a certain topic, verbatim transcripts may already provide too little information to *closely* retrace the thought process. If one decides that a verbatim transcript contains sufficient information, it should be explained why that is the case. (2) All publications should identify how the transcripts were developed and what information they include. It is good practice to use well-defined transcription systems whenever possible (see, e.g., Flick (2013, p. 64ff) for an overview).

As discussed in the previous sections, there is a plethora of empirical and theoretical ideas that underline the importance of rigor in the context of transcription. Ultimately, an unreflective choice or poor strategy may even lead to the discarding of promising qualitative research ideas. We have provided theoretical foundations, shared practical experiences, argued for better transcription practices, and presented a practical three-step scheme (see Fig. 5) that researchers can follow. More rigorous standards can improve the findings of our field's research in the long term, especially regarding:

- **Contextual Richness:** Transcribing layers of conversation that preserve details in addition to semantic content can improve the accuracy and reliability of coding.
- **Pedagogical Insights:** Thorough transcription offers additional, valuable resources for researchers in education, potentially aiding in better understanding, for example, teaching methodologies and student interactions.
- **Inclusivity and Equity:** Accurate transcription acknowledges diverse linguistic expressions, ensuring fair representation and potentially avoiding bias in educational research interpretations.

- **Validity of Results:** Detailed transcription may sometimes be needed to pursue certain research interests. Therefore, better transcription practices can enable researchers to access interpretations that are inaccessible with other transcription systems. This can improve reliability and, ultimately, the validity of results.

Admittedly, opting for the most complex transcription system and strategy is not always practical, as transcription is “[t]ime consuming and can therefore also be a cost factor, [...] it is important to consider what level of accuracy is truly necessary to answer your research questions” (Kuckartz and Rädiker, 2019). Thus, again, our appeal is not an endorsement of always selecting the most complex system. Rather, we call for more rigor and consciousness of the topic, as the identified potentials can only be realized if the community follows clearer standards. By no means do we argue that the method sections of all qualitative research papers should grow significantly in the future. Instead, this paper encourages researchers to articulate, reflect on, and justify their choice of transcription system and to make their guiding rules for transcript creation transparent.

Higher standards can serve as a catalyst for meaningful progress, deter incorrect or erroneous interpretations, and open up fresh avenues in our field. Our community, despite being relatively young compared to others, has established and matured significantly over time, and the inquiries being addressed are becoming increasingly complex, intricate, and often interdisciplinary. To be adequately prepared and to foster openness for interdisciplinary collaborations, maintaining a rigorous standard in the context of transcription is important to remain both relevant and progressive as a field.

Acknowledgments

We thank the anonymous reviewers of a previous version of this article for their valuable feedback and tips that improved the paper’s quality. We also thank our colleagues for their feedback on different drafts of the paper and during discussions. We also would like to thank all research assistants for their support during transcription, data acquisition, and data analysis.

Funding

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation): TRR 318/1 2021 - 438445824.

References

- Aivaloglou, E., Meulen, A.v.d. (2021). An Empirical Study of Students' Perceptions on the Setup and Grading of Group Programming Assignments. *ACM Transactions on Computing Education*, 21(3), 17–11722. <https://doi.org/10.1145/3440994>
- Alshahrani, A., Ross, I., Wood, M.I. (2018). Using Social Cognitive Career Theory to Understand Why Students Choose to Study Computer Science. In: *Proceedings of the 2018 ACM Conference on International Computing Education Research*. ACM, Espoo Finland, pp. 205–214. 978-1-4503-5628-2. <https://doi.org/10.1145/3230977.3230994>
- Ashmore, M., Reed, D. (2005). Innocence and Nostalgia in Conversation Analysis: The Dynamic Relations of Tape and Transcript.
- Auer, P. (1992). The neverending sentence: Rightward expansion in spoken language. In: Kontra, M., Tamás, V. (Eds.), *Studies in Spoken Language: Englisch, German, Finno-Ugric*. Hungarian Academy of Sciences, Linguistics Institute, Budapest.
- Ayaß, R. (2015). Doing data: The status of transcripts in Conversation Analysis. *Discourse Studies*, 17(5), 505–528. <https://doi.org/10.1177/1461445615590717>
- Bender, E., S. Niclas., C. Michael E., M. Melanie, and Hubwieser, P. (2016). Identifying and Formulating Teachers' Beliefs and Motivational Orientations for Computer Science Teacher Education. *Studies in Higher Education*, 41(11), 1958–1973. <https://doi.org/10.1080/03075079.2015.1004233>
- Bergmann, P., Mertzluft, C. (2009). Segmentierung spontansprachlicher Daten in Intonationsphrasen – Ein Leitfaden für die Transkription. In: *Die Arbeit mit Transkripten in Fortbildung, Lehre und Forschung* (Gedruckte Ausgabe ed.). Verlag für Gesprächsforschung, Mannheim, pp. 83–95. 978-3-936656-34-3.
- Bloch, C. (1996). Emotions and Discourse. *Text & Talk*, 16(3), 323–342. <https://doi.org/10.1515/text.1.1996.16.3.323>
- Carlson, K. (2009). How Prosody Influences Sentence Comprehension. *Language and Linguistics Compass*, 3(5), 1188–1200. <https://doi.org/10.1111/j.1749-818X.2009.00150.x>
- Chafe, W. (1994). *DISCOURSE, CONSCIOUSNESS, AND TIME*. Discourse, Consciousness, and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing. University of Chicago Press, Chicago, IL. 978-0-226-10054-8.
- Cole, J., Shattuck-Hufnagel, S. (2016). New Methods for Prosodic Transcription: Capturing Variability as a Source of Information. *Laboratory Phonology*, 7(1). <https://doi.org/10.5334/labphon.29>
- Couper-Kuhlen, E., Selting, M. (2018). *Interactional linguistics: studying language in social interaction*. Cambridge University Press, Cambridge, United Kingdom ; New York, NY. 978-1-107-03280-4 978-1-10761603-5.
- Davidson, C. (2009). Transcription: Imperatives for Qualitative Research. *International Journal of Qualitative Methods*, 8(2), 35–52. <https://doi.org/10.1177/160940690900800206>
- Du Bois, J.W., Schuetze-Coburn, S., Cumming, S., Paolino, D. (1993). Outline of Discourse Transcription. In: *Talking Data*. Psychology Press, New York. 978-1-315-80792-8.
- Duranti, A. (2006). Transcripts, Like Shadows on a Wall. *Mind, Culture, and Activity*, 13(4), 301–310. https://doi.org/10.1207/s15327884mca1304_3
- Fitzgerald, S., McCauley, R., Plano Clark, V.L. (2011). Report on qualitative research methods workshop. In: *Proceedings of the 42nd ACM technical symposium on Computer science education -SIGCSE '11*. ACM Press, Dallas, TX, USA, p. 241. 978-1-4503-0500-6. <https://doi.org/10.1145/1953163.1953237>
- Flick, U. (2013). *The SAGE Handbook of Qualitative Data Analysis*. SAGE. 978-1-4462-9669-1.
- Fowler, M., Chen, B., Zilles, C. (2021). How should we 'Explain in plain English'? Voices from the Community. In: *Proceedings of the 17th ACM Conference on International Computing Education Research*. ICER 2021. Association for Computing Machinery, New York, NY, USA, pp. 69–80. 978-1-4503-8326-4. <https://doi.org/10.1145/3446871.3469738>.
- Frank, M.G., Griffin, D.J., Svetieva, E., Maroulis, A. (2015). Nonverbal Elements of the Voice. In: Kostić, A., Chadee, D. (Eds.), *The Social Psychology of Nonverbal Communication*. Palgrave Macmillan UK, London, pp. 92–113. 978-1-137-34586-8. https://doi.org/10.1057/9781137345868_5
- Goodwin, C., Heritage, J. (1990). Conversation Analysis. *Annual Review of Anthropology*, 19, 283–307. <https://doi.org/10.1146/annurev.an.19.100190.001435>
- Jefferson, G. (2004). Glossary of transcript symbols with an introduction. *Conversation analysis*, 13–31. <https://doi.org/10.1075/pbns.125.02jef>

- Kong, M., Mauriello, M.L., Pollock, L. (2022). Exploring K-8 Teachers' Preferences in a Teaching Augmentation System for Block-Based Programming Environments. In: *Proceedings of the 22nd Koli Calling International Conference on Computing Education Research*. Koli Calling '22. Association for Computing Machinery, New York, NY, USA, pp. 1–12. 978-1-4503-9616-5.
<https://doi.org/10.1145/3564721.3564725>
- Kroes, P. (1998). Technological Explanations. *Phil & Tech*, 3(3), 124–134.
<https://doi.org/10.5840/techn19983325>
- Kuckartz, U. (2014). *Qualitative Text Analysis: A Guide to Methods, Practice and Using Software*. SAGE. 978-1-4462-9776-6.
- Kuckartz, U., Rädiker, S. (2019). Transcribing Audio and Video Recordings. In: Kuckartz, U., Rädiker, S. (Eds.), *Analyzing Qualitative Data with MAXQDA: Text, Audio, and Video*. Springer International Publishing, Cham, pp. 41–49. 978-3-030-15671-8. https://doi.org/10.1007/978-3-030-15671-8_4
- Lapadat, J.C. (2000). Problematizing transcription: Purpose, paradigm and quality. *International Journal of Social Research Methodology*, 3(3), 203–219. <https://doi.org/10.1080/13645570050083698>
- Mandal, F. (2014). Nonverbal Communication in Humans. *Journal of Human Behaviour in the Social Environment*, 24, 417–421. <https://doi.org/10.1080/10911359.2013.831288>
- Mondada, L. (2011). Understanding as an embodied, situated and sequential achievement in interaction. *Journal of Pragmatics*, 43(2), 542–552. <https://doi.org/10.1016/j.pragma.2010.08.019>
- Mondada, L. (2018). Multiple Temporalities of Language and Body in Interaction: Challenges for Transcribing Multimodality. *Research on Language and Social Interaction*, 51(1), 85–106.
<https://doi.org/10.1080/08351813.2018.1413878>
- Mondada, L. (2019). Transcribing silent actions: a multimodal approach of sequence organization. *Social Interaction. Video-Based Studies of Human Sociality*, 2(1).
<https://doi.org/10.7146/si.v2i1.113150>
- Murphy, L., Fitzgerald, S., Hanks, B., McCauley, R. (2010). Pair Debugging: A Transactive Discourse Analysis. In: *Proceedings of the Sixth International Workshop on Computing Education Research*. ICER '10. Association for Computing Machinery, New York, NY, USA, pp. 51–58. 978-1-4503-0257-9. <https://doi.org/10.1145/1839594.1839604>
- Ochs, E. (1979). *Transcription as Theory*. (2nd ed.), New York.
- O'Connell, D.C., Kowal, S. (1995a). Basic Principles of Transcription. In: *Rethinking Methods in Psychology*. SAGE Publications Ltd, 1 Oliver's Yard, 55 City Road, London EC1Y 1SP United Kingdom, p. 93. 978-08039-7733-4 978-1-4462-2179-2. <https://doi.org/10.4135/9781446221792>
- O'Connell, D.C., Kowal, S. (1995b). *Rethinking Methods in Psychology*. SAGE Publications Ltd, London. <https://doi.org/10.4135/9781446221792>
- Oleson, A., Xie, B., Salac, J., Everson, J., Kivuva, F.M., Ko, A.J. (2022). A Decade of Demographics in Computing Education Research: A Critical Review of Trends in Collection, Reporting, and Use. In: *Proceedings of the 2022 ACM Conference on International Computing Education Research -Volume 1*. ACM, Lugano and Virtual Event Switzerland, pp. 323–343. 978-1-4503-9194-8.
<https://doi.org/10.1145/3501385.3543967>
- Oliveira, M. (2002). The Role of Pause Occurrence and Pause Duration in the Signaling of Narrative Structure. In: Ranchhod, E., Mamede, N.J. (Eds.), *Advances in Natural Language Processing*. Springer, Berlin, Heidelberg, pp. 43–51. 978-3-540-45433-5. https://doi.org/10.1007/3-540-45433-0_7
- Olsen, R.M. (2019). The Acoustics of Feeling: Emotional Prosody in the StoryCorps Corpus. *The Journal of the Acoustical Society of America*, 145(3_Supplement), 1930. <https://doi.org/10.1121/1.5102025>
- Olson, D.R. (1993). How writing represents speech. *Language & Communication*, 13(1), 1–17.
[https://doi.org/10.1016/0271-5309\(93\)90017-H](https://doi.org/10.1016/0271-5309(93)90017-H)
- O'Connell, D.C., Kowal, S. (2009). Transcription systems for spoken discourse. *John Benjamins Publishing*, 4, 240.
- Park, J.S.-Y. (2002). Cognitive and interactional motivations for the intonation unit. *Studies in Language. International Journal sponsored by the Foundation "Foundations of Language"*, 26(3), 637–680.
<https://doi.org/10.1075/sl.26.3.07par>
- Point, S., Baruch, Y. (2023). (Re)thinking transcription strategies: Current challenges and future research directions. *Scandinavian Journal of Management*, 39(2), 101272.
<https://doi.org/10.1016/j.scaman.2023.101272>
- Radford, A., Kim, J.W., Xu, T., Brockman, G., McLeavey, C., Sutskever, I. (2022). Robust Speech Recognition via Large-Scale Weak Supervision. <https://doi.org/10.48550/ARXIV.2212.04356>
- Ralph, P., Ali, N.b., Baltes, S., Bianculli, D., Diaz, J., Dittrich, Y., Ernst, N., Felderer, M., Feldt, R., Filieri,

- A., de França, B.B.N., Furia, C.A., Gay, G., Gold, N., Graziotin, D., He, P., Hoda, R., Juristo, N., Kitchenham, B., Lenarduzzi, V., Martinez, J., Melegati, J., Mendez, D., Menzies, T., Molleri, J., Pfahl, D., Robbes, R., Russo, D., Saarimäki, N., Sarro, F., Taibi, D., Siegmund, J., Spinellis, D., Staron, M., Stol, K., Storey, M.-A., Taibi, D., Tamburri, D., Torchiano, M., Treude, C., Turhan, B., Wang, X., Vegas, S. (2021). Empirical Standards for Software Engineering Research. arXiv. <https://doi.org/10.48550/arXiv.2010.03525>
- Roberts, C., Street, B. (2017). Spoken and Written Language. In: *The Handbook of Sociolinguistics*. John Wiley & Sons, Ltd, pp. 168–186. 978-1-4051-6625-6. <https://doi.org/10.1002/9781405166256.ch10>
- Romero, C., O’Connell, D.C., Kowal, S. (2002). Notation Systems for Transcription: An Empirical Investigation. *Journal of Psycholinguistic Research*, 31(6), 619–631. <https://doi.org/10.1023/A:1021217105211>
- Rudzicz, F. (2016). *Clear Speech: Technologies that Enable the Expression and Reception of Language*. Morgan & Claypool Publishers. 978-1-62705-827-8.
- Sanders, K., Vahrenhold, J., McCartney, R. (2023). How Do Computing Education Researchers Talk About Threats and Limitations? In: *Proceedings of the 2023 ACM Conference on International Computing Education Research V.1*. ACM, Chicago IL USA, pp. 381–396. 978-1-4503-9976-0. <https://doi.org/10.1145/3568813.3600114>
- Schulte, C. (2008). Duality Reconstruction -Teaching digital artifacts from a socio-technical perspective. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5090 LNCS, 110–121. https://doi.org/10.1007/978-3-540-69924-8_10
- Selting, M. (2000). The construction of units in conversational talk. *Language in Society*, 29(4), 477–517. <https://doi.org/10.1017/S0047404500004012>
- Selting, M., Auer, P., Barth-Weingarten, D. (2011). A system for transcribing talk-in-interaction : GAT 2. *Gesprächsforschung : Online-Zeitschrift zur verbalen Interaktion*, 12, 1–51.
- Sidnell, J. (2012). Basic Conversation Analytic Methods. In: Sidnell, J., Stivers, T. (Eds.), *The Handbook of Conversation Analysis* (1st ed.). Wiley, pp. 77–99. 978-1-4443-3208-7 978-1-118-32500-1. <https://doi.org/10.1002/9781118325001.ch5>
- Simpson, H.E. (2016). *The role of intonation units in memory for spoken English*. Dissertation, University of California, Santa Barbara, Santa Barbara, Calif.
- Speer, S.R., Ito, K. (2009). Prosody in First Language Acquisition – Acquiring Intonation as a Tool to Organize Information in Conversation. *Language and Linguistics Compass*, 3(1), 90–110. <https://doi.org/10.1111/j.1749-818X.2008.00103.x>
- Spiegel, C. (2009). Transkripte als Arbeitsinstrument: Von der Arbeitsgrundlage zur Anschauungshilfe. In: *Die Arbeit mit Transkripten in Fortbildung, Lehre und Forschung* (Gedruckte Ausgabe ed.). Verlag für Gesprächsforschung, Mannheim, pp. 7–15. 978-3-936656-34-3.
- Stein, S. (2018). 1. Oralität und Literalität. In: *Handbuch Text und Gespräch*. De Gruyter, Berlin, Boston, pp. 3–25. 978-3-11-029605-1. <https://doi.org/10.1515/9783110296051-001>
- Tenenberg, J., Chinn, D. (2019). Social Genesis in Computing Education. *ACM Transactions on Computing Education*, 19(4), 34–13430. <https://doi.org/10.1145/3322211>
- Terfloth, L., Schaffer, M., Buhl, H.M., Schulte, C. (2023). Adding Why to What? Analyses of an Everyday Explanation. In: Longo, L. (Ed.), *Explainable Artificial Intelligence*. Communications in Computer and Information Science. Springer Nature Switzerland, Cham, pp. 256–279. 978-3-031-44070-0. https://doi.org/10.1007/978-3-031-44070-0_13

L. Terfloth is a PhD student at the Computing Education Research Group at the University of Paderborn, Germany. His primary research focuses on the empirical study of explanations of technological artifacts. In his dissertation, he develops a novel analytical approach grounded in techno-philosophical theory to systematically investigate how explanations about technology are constructed and understood. By integrating philosophical perspectives with empirical research methods, his work aims to deepen our understanding of the processes and challenges involved in making complex technological systems explainable in educational and practical contexts.

V. Lohmer is a PhD student at the Transregional Collaborative Research Centre 318 (Sonderforschungsbereich Transregio 318) at Bielefeld University, where she works as a research associate and doctoral researcher in project A04. Her research focuses on multi-modality and gestural behavior in everyday explanations.

F. Kern teaches German linguistics and their didactics at Bielefeld University. After studying German literature, linguistics and philosophy in Berlin and London, she was awarded the Dr. phil., from the University of Hamburg on the communicative differences between East and West Germans in Job Interviews. Her research interests are in the area of Ethnomethodological Conversational Analysis, multimodal interaction analysis and ethnography, discourse acquisition and multimodal language development, classroom interaction, and learning in multimodal environments. Her publications include work on rhythm in Turkish German, on the development of children's multimodal storytelling and explanations, and on interactions in learning situations.

C. Schulte is a professor for Computing Education Research at Paderborn University, Germany. His work and research interests are the philosophy of computing education, artificial intelligence in education, and empirical research on teaching and learning processes (including eye movement research). Since 2017, he has been working together with Didactics of Mathematics (Paderborn University) on the ProDaBi project, in which data science and artificial intelligence are prepared as teaching topics. He is also a PI in the collaborative research centre TRR318 'Constructing Explainability' on explainable AI.

3.2 Building Block 2: Dual Nature in Naturalistic Explanations

Building Block 2

L. Terfloth et al. (2023). “Adding Why to What? Analyses of an Everyday Explanation”. en. In: *Explainable Artificial Intelligence*. Ed. by L. Longo. Vol. 1903. Cham: Springer Nature, pp. 256–279. ISBN: 978-3-031-44070-0. DOI: [10.1007/978-3-031-44070-0_13](https://doi.org/10.1007/978-3-031-44070-0_13)

3.2.1 Summary

This paper presents the first empirical investigation of how the dual nature of a technological artifact is addressed in naturalistic, everyday explanations. Using the board game Quarto as the explanandum, 20 dyadic explanatory interactions were recorded and analyzed using qualitative content analysis. The coding manual was grounded in the dual nature theory (Kroes, 1998). GAT2-transcribed speech was coded into utterances addressing Architecture and utterances addressing Relevance, achieving satisfactory intercoder reliability ($k=0.80$). The analysis was triangulated with a video recall in which explainers retrospectively justified their content choices at selected moments. Three research questions were addressed: (1) how is the dual nature expressed in explanatory utterances, (2) what patterns emerge across explanations, and (3) how do explainers justify their content choices. Most importantly, both sides of the dual nature were identifiable in almost all explanations. Architecture was addressed substantially more frequently than Relevance overall, and was dominant particularly in the first third of explanations — a pattern the paper terms *laying out the tools*. A subsequent video recall, in which explainers watched segments of their own explanation and commented on their reasoning for specific utterances, revealed that explainers regarded the Architecture as prerequisite knowledge required before more complex aspects of Relevance could be meaningfully introduced. Shifts toward Relevance in later stages were driven by the explainer’s intention, the explainee’s expressed knowledge needs, and dynamics emerging from the interaction itself. The papers’ finding imply that both sides of the dual nature are necessary for genuine understanding. The identified patterns offer a first empirical basis for recommendations regarding the construction of synthetic explanations in XAI.

3.2.2 Authors' Contributions

Lutz Terfloth conceived and designed the paper, developed the contextual framing for the XAI community, created and refined the coding manual through iterative pilot studies, led the qualitative content analysis of the explanations, co-conducted the data collection, wrote the majority of the manuscript including finalizing all sections, revised the paper in response to peer review, and presented the paper at the conference. Michael Schaffer designed and administered the sociodemographic questionnaires, co-conducted the data collection, designed and conducted the analyses of the video recall data, wrote first drafts of the video recall sections, and provided feedback on the manuscript. Heike Buhl contributed expertise in educational psychology, supervised Michael Schaffer's contributions, and provided feedback on the manuscript. Carsten Schulte contributed to the theoretical framing within the XAI and computing education context and provided feedback on successive drafts.

Lutz Terfloth did greater than or equal to 70% of the work.

Acknowledgements

The experimental setting was designed as part of a research proposal by Heike Buhl, Friederike Kern, and Carsten Schulte within the TRR 318 *Constructing Explainability*. Vivien Lohmer and Friederike Kern contributed to the study design, data collection, and transcription procedure. I thank all student assistants for their support during data acquisition.

3.2.3 Author's Commentary

This paper was presented at the XAI World Conference 2023 in Lisbon, which I attended together with Carsten Schulte and Michael Schaffer. It is the empirical core of the dissertation; the study in which the dual nature of technical artifacts as an analytical lens for explanations, was first put to an empirical test in a naturalistic setting. A central question, besides to what extent explainers address both Architecture and Relevance, was whether using these categories for content coding is sensitive enough to reveal structured patterns in how they do so across a corpus of 20 explanations. The answer was affirmative on both counts.

The finding that most resonates throughout the remainder of this dissertation is *laying out the tools*: the systematic tendency of explainers to concentrate Architecture-oriented content in the early stages of an explanation, treating and rationalizing it as a prerequisite scaffold before introducing Relevance. This was not an arbitrary observation but a pattern that held across almost all dyads and was explicitly confirmed and verified through, for example, confirmations by explainers in the

video recalls. It suggests that the Architecture-first approach is not incidental but reflects a genuine explanatory intuition: the belief that some structural knowledge must precede understanding of the Relevance as some sort of a baseline.

This intuition, however, deserves scrutiny. The pattern is prevalent, but prevalence does not imply optimality. Whether an Architecture-first approach actually produces *better* understanding in the explainee, or whether a more integrated presentation of Architecture and Relevance would be superior, is a question this paper does not answer. That question is taken up in Building Block 4. What this paper does indeed establish is the descriptive baseline. Given the discussion surrounding a default mode of construal, the laying-out-the-tools pattern is particularly interesting.

A further finding that gains significance in light of the later Building Blocks is that Relevance was not always absent from the early stages of an explanation even when it was not verbally expressed. In several cases, the video recall revealed that explainers had structured their Architecture-focused content with Relevance in mind, selecting which components to introduce based on their later functional significance. This suggests that the relationship between Architecture and Relevance in the explainer's reasoning may be more integrated than the sequential surface pattern implies.



Adding Why to What? Analyses of an Everyday Explanation

Lutz Terfloth^(✉), Michael Schaffer, Heike M. Buhl,
and Carsten Schulte

Paderborn University, Paderborn, Germany
{lutz.terfloth,michael.schaffer,heike.buhl,
carsten.schulte}@uni-paderborn.de

Abstract. In XAI it is important to consider that, in contrast to explanations for professional audiences, one cannot assume common expertise when explaining for laypeople. But such explanations between humans vary greatly, making it difficult to research commonalities across explanations. We used the dual nature theory, a techno-philosophical approach, to cope with these challenges. According to it, one can explain, for example, an XAI’s decision by addressing its dual nature: by focusing on the Architecture (e.g., the logic of its algorithms) or the Relevance (e.g., the severity of a decision, the implications of a recommendation). We investigated 20 explanations of games using the theory as an analytical framework. We elaborate how we used the theory to quickly structure and compare explanations of technological artifacts. We supplement results from analyzing the explanation contents with results from a video recall to explore how Explainers (EX) justified their explanation. We found that EX were focusing on the physical aspects of the game first (Architecture) and only later on aspects of the Relevance. Reasoning in the video recalls indicated that EX regarded the focus on the Architecture as important for structuring the explanation initially by explaining the basic components before focusing on more complex, intangible aspects. EX justified shifting between addressing the two sides by explanation goals, emerging misunderstandings, and the knowledge needs of the explainee. We discovered several commonalities that inspire future research questions which, if further generalizable, provide first ideas for the construction of synthetic explanations.

Keywords: Analysis of Human Explanations · Naturalistic Explanations · Qualitative Analysis · Technological Artifacts

1 Introduction

Enabling laypeople to “effectively understand, trust, and manage” [16] AI applications requires XAI systems that provide *understandable explanations*. In a similar sense, the General Data Protection Regulation (GDPR) calls for “meaningful explanations of the logic involved” in the context of automated decision-making,

© The Author(s) 2023
L. Longo (Ed.): xAI 2023, CCIS 1903, pp. 256–279, 2023.
https://doi.org/10.1007/978-3-031-44070-0_13

too. However, in many instances, experts rely on a notion of what constitutes an explanation for the implementation [48]. We see potential in adhering to an empirically grounded conceptualization of explanations instead. If an XAI system could mimic how explanations evolve in naturalistic, everyday settings, it could be used to make an AI's output more understandable, especially for laypeople unfamiliar with the technology.

Over the years, plenty of research has been done on explanations [13, 34]. For XAI, researching everyday explanations offers large potential for understanding how both parties of an explanation engage interactively; for a "Social Design of AI Systems" in which the explainer and explainee are put in focus [38]. But little research focused on explanations of technological artifacts in particular, even though important in the context of explaining and understanding such artifacts [37]. One central idea can be found across different sources: Technological artifacts are human-made objects engineered to serve as means to certain ends. Therefore, explaining technological artifacts can be done addressing the two sides of their dual nature [22, 47]. Whenever, for example, engineers explain their inventions in texts, they are explaining aspects about the physical properties (e.g., the shape, the algorithms, or the logic – its Architecture) of the artifact in alignment with its functional capabilities (e.g., how it serves as a means to an end; its Relevance). In the context of XAI, in which synthetic explanations are constructed by technological artifacts, and based on decisions by technological artifacts, the peculiarities of their dual nature cannot be neglected. We believe, the theory provides a rich background for researching everyday explanations of technological artifacts.

Many scholars agree that knowledge about the dual nature of a technological artifact is relevant for understanding it, yet are undecided whether both sides are equally important or whether one side may be the precursor for the other [42, 45]. However, to date, no study investigates specifically if and how the dual nature is addressed in naturalistic everyday explanations of technological artifacts. Thus, how the dual nature is addressed in explanations, which of the two sides are explained first, more frequently, or whether one or the other should be addressed in more detail or omitted completely, is not empirically investigated yet. An empirically grounded conceptualization of how technological artifacts are naturally explained in a verbal, interactive settings would be a useful prerequisite for many research areas, but especially for XAI and education in general. It could provide a foundation for formulating recommendations on whether and which of the artifact's properties should be referred to within a (synthetic) explanation, and thus improve the understandability and interpretability of XAI systems.

This paper is a step towards addressing these shortcomings and investigating how the content of naturalistic explanations evolves and is justified by the explainers (EX). It presents results from a study of dyadic explanatory dialogues in which a technological artifact, the board game *Quarto!*, is explained in a naturalistic setting. The contribution of this paper is twofold. First, we elaborate on the theoretical foundation used for an empirical study, after which we present results from that study. The paper consists of two parts: In the Sections Background and Study, we provide a brief introduction to the theoretical

foundation of the research and connect it to the research questions of the paper, and elaborate how and why we use the dual nature theory as a framework. In the Method, Results and Discussion sections, the study, data acquisition and analysis are elaborated after which the results are presented and discussed.

2 Background

Considerable research attention has been directed towards the conceptualization of explanations from different perspectives and disciplines. A philosophical standpoint sheds light on quality characteristics of explanations and tries to identify factors that contribute to sound and satisfying explanations [11, 40]. Social psychology examines why and when individuals explain [11, 15, 18, 30]. Explanations can cover a vast array of topics and ideas, but regardless of content diversity, explanations in their composition are shaped by causal and logical factors [40], too. Cognitive sciences are investigating important elements of explanations, the process of explanation generation and which cognitive processes succumb to explanatory processes [11, 27].

During extemporaneous explanations, EX are facing various challenges to give meaningful, and complete and precise explanations that are well-structured [39]. Structuring explanations requires organization of knowledge [39]. Therefore, explaining is a cognitive process, that is highly constructive [9, 17] and utilizes existing domain knowledge [27], that is part of mental representations [9, 20]. Thus, EX need to decide, over the course of explanations, which aspects of a domain they want to explain and especially at what point of time, for example by monitoring the explainee (EE) [26].

These challenges an EX faces makes researching such naturalistic explanations challenging as well as necessary, if the goal is to provide recommendations for XAI systems. To circumvent some of these challenges, shared properties of explananda are interesting. The content of an explanation is influenced by the explanandum (the subject of explanation), more precisely the EX's mental representation of that explanandum, and the knowledge needs of the Explainee (EE) regarding that explanandum [3, 10]. Whenever the explanandum is a technological artifact, one can refer to its dual nature, following a techno-philosophical theory [22, 46]. According to the dual nature theory, an explanation of a technological artifact – an object engineered to be a means to an end – can be formulated by addressing two properties [21]. On the one hand, a structural mode of description, “makes use of concepts from physical laws and theories and is free of any reference to the function of the object” and addresses the Architecture of the artifact [22, 43]. On the other hand, the functional/teleological mode of description is a way in which “[w]ith regard to its function, a technological object is described in an intentional (teleological) way” [22] and addresses the Relevance of the artifact.

Interestingly, in the context of this dual nature theory, the idea of an engineer's ability to “bridge the gap” [23] between the two sides of the dual nature resulting from their deep understanding of the artifact is discussed. Not only, but especially because of this, the dual nature theory found application within the computing

education community [43], as well as in the technology education community in the context of understanding technological artifacts [8]. It guided a development of an analytical framework to investigate pupil's understanding of programmed technical solutions (PTS), especially concerning their ability to open the black boxes [7]. The results indicated that both sides of their dual nature need to be understood to have a more profound understanding of a PTS: "[t]hus, these key elements are important to consider in pedagogical practice to promote learning regarding PTS." [7]. The theory served as a theoretical foundation for a program comprehension model, too [42]. On a more abstract level, comparable ideas were referred to using different foci and terms, but similar ideas in the context of program comprehension: "mechanisms" vs. "explanations" [45], and "text base" vs. "situation model" [36]. [1] investigated how children discuss a machine they got to know while visiting a museum, concerning which prompt resulted in the children addressing either the mechanisms or the components when asked questions about the machine. In summary, the main ideas of the dual nature theory have a rich history not only within and outside the computing community, but in the context of making sense of technological artifacts in general.

Whether the dual nature theory is useful for the analysis of explanation content whenever a technological artifact is explained, still needs investigation. We call the sides of the dual nature, the *Architecture* (roughly: what it is, what components are) and *Relevance* (roughly: why and what it and its components are for) of the artifact. Our conceptualization of the two sides is as follows: one can, for example, (1) explain how a navigation system works on the level of data, or the logic of the algorithms (i.e., addressing the *Architecture*), or by explaining (2) how one may use it for relaxed journey across foreign countries (i.e., addressing the *Relevance*). But there is uncertainty how the dual nature of the technological artifact influences the structure and evolution of an explanation, and whether these sides can be clearly identified within naturalistic explanations. This paper aims at answering three questions:

- (RQ1) How is the dual nature of the explanandum expressed in the utterances in the explanations?
- (RQ2) What are common patterns regarding sequences of utterances about the dual nature across the explanations?
- (RQ3) How do EX justify their choice of explanation content regarding which side of the dual nature was addressed?

3 Method

Given our overarching objective to analyze how EX address the sides of the dual nature in their explanations, we conducted a study in which a dyad of people engaged in a naturalistic explanation scenario. To acquire comparable explanations, a study design that allowed to control certain variables that would not be controllable with a more complex explanandum and in the field, was developed. Even though the naturalness of the explanations in the field would be higher, at this early stage of researching the phenomenon a certain uniformity

is preferred. Researchers and research assistants adhered to a predefined study guideline. The study was conducted at Paderborn University as well as Bielefeld University in Germany. This study is part of a larger, more complex study consisting of five different phases: (1) pretesting, (2) explaining, (3) video recall, (4) playing and (5) post testing. This paper focuses on the analysis of phase 2 and 3 (explanation phase and video recall). We recorded the dyads during the explanation. To observe the explanations from another room, we placed a webcam for live-streaming, too. Prior to the explanation, every participant filled out a questionnaire for sociodemographic details. After the explanation (phase 2), we conducted a video recall to assess the EX underlying thinking and reasoning during the explanations based on questions of a structured interview. The analyses of the content of the explanations, and the justifications mentioned in the video recalls, is guided by the dual nature theory as an analytical framework.

3.1 Participants and Recruitment

For the recruitment, we relied primarily on onsite recruiting via handing out flyers on campus, hanging posters, and hearsay. We compensated all participants monetarily (10 per hour of participation). Participants signed up voluntarily for the study. All studies included COVID-19 safety measures and obtaining written informed consent before the studies. Communication with participants was based on boilerplate to ensure that all participants had the same information and instructions. We divided a total of 48 participants into two groups: Explainers and Explainees (24 dyads). Interaction between EE and EX before the study was prevented as much as possible by meeting and picking them up independently of each other on the day of the study. Through mail communication, we instructed the group of EX to make themselves familiar with the explanandum. They had the option to either pick up a physical copy of the game from our lab or use an online version of the game. EE received only directions to, and the location of the study. Exclusion criteria were (1) prior knowledge of the game *Quarto!* (for EE), (2) non-C2 level language skills (for both EE and EX), or (3) prior participation and knowledge about the overarching research endeavor. Four dyads were excluded from the corpus due to a language barrier, two instances in which the EE had prior knowledge of the game, and one instance in which the game was mistakenly present during explanation. The final sample of video data consists of 20 lab studies in which 20 EX explained the game to 20 EE (18 female, 19 male, 1 non-binary)¹. Age ranged from 18 to 39 ($M=24.92$, $SD=4.42$). 36 participants had an academic background. 35 reported to be students (e.g., engineering, education, economics, law, computer science, media sciences, linguistics), 2 were full-time employees. Out of 19 EX, 7 reported to have explained the game to someone else before the study. EX reported to have between 0 and 18 rounds of gameplay experience ($M=5.46$, $SD=5.18$). 10 EX reported to have experience in explanations (e.g., tutoring). The studies were planned to take 2 to 3 h for

¹ 3 participants provided no information and are therefore not included in the descriptive statistics.

all assessments. Phase 2 (explanation) varied from 02:23 mm:ss to 16:17 mm:ss ($M=07:24$, $SD=03:22$).

3.2 The Explanandum: *Quarto!*

The explanandum of the study is the strategical, two-player board game *Quarto!*. *Quarto!* is a game developed by Swiss mathematician Blaise Müller. It demands deductive reasoning and thinking, similar to chess. At any moments during the game, all information is openly available to every player. The game is made up of 16 pieces and a board consisting of 4×4 squares. Each of the 16 pieces one places on the board, is unique and differs in at least one of the following traits: size (tall or short), shade (light or dark), shape (round or square), and solidity (solid or hollow). The goal of the game is similar to Connect Four: connect a row, column, or diagonal of four pieces that share *at least one trait*. Contrary to the way Connect Four is played, players do not choose which piece they place themselves, but instead choose the piece the opponents have to place in their next move. After the opponent places the piece handed to them, they choose a piece, that the other person has to play. Turn-by-turn, the game continues until either a *Quarto!* is called, and a person wins, or all squares are occupied and the game ends in a draw.

We chose a game and specifically *Quarto!* as the explanandum for several reasons. As our interdisciplinary team strives to triangulate the results from different approaches to get a rich perspective of the phenomenon of natural explanations of technological artifacts, certain tradeoffs were inevitable. Explaining games is a common, social practice, making it an accessible domain for a large variety of participants. The occasion to explain a game creates a setting for natural explanations and lessens any perceived pressure or exam-like atmosphere. Regarding participant acquisition, one can learn the game quickly, whereas mastering the game can be considered challenging. As *Quarto!* provides strategical depth, it allows for some sort of variance in the explanations. Yet, the scope of what one may explain is narrow enough to make the explanations comparable.

Games, being invented by humans, share the main characteristics of technological artifacts. We believe, games exhibit their dual nature quite well in the context of explanations: One has to explain the pieces, board and rules, hence the Architecture but one also needs to explain how these things go together and create an interesting game in which two players struggle to win. Comparable to chess, for example, one could explain how certain moves can be interpreted as offensive or defensive. In theoretical terms, the explanation also needs to elaborate how one follows their intentions within the game and thus address the game's Relevance. In summary, providing an explanation that results in a deep understanding of a game does not only require explaining its components and rules (knowing what, Architecture), but also explaining why certain aspects of the game are important to consider during gameplay (knowing why; Relevance). This is comparable to understanding XAI explanations.

The decision for the absence of the game had different reasons. Regarding the theoretical framework for the analysis, we expected having the game on

hand would limit utterances that address the Architecture of the game (e.g., about the different shapes of the pieces). Thus, the EX needs to decide which detail requires explanation, allowing us to perceive strategies that EX use to, for example, generalize such details. As gestures are linked to or results of cognitive processes, a gesture analysis offers additional insights to get an even more detailed understanding of what happens during explanations at later stages of the project [31,32]. As a consequence, the table between the participants needs to be free of any objects to not influence their gestures (e.g., participants could hold on to pieces of the game).

3.3 Procedure

The study was conducted by researchers and research assistants from an interdisciplinary background (linguistics, psychology, and computing education). After participants arrived, they were welcomed, and, for the sociodemographic questionnaire, led into two different rooms independently of each other. During the study, all communication between researchers and participants was based on boilerplate, too. In phase 2, the explanations were recorded from three different camera angles: a view of the table, a medium-shot of the EE upper body, and a medium-shot of the EX upper body.

After EE and EX filled out the questionnaires for sociodemographic details, they were instructed about the next phase. The EX's instruction was: "In the next room, please explain the game to the person opposite you in such a way, that the person would have a chance to win the game"². The EE's instruction was, "[i]n a moment, a person will enter the room and explain a board game to you. Please participate actively in the explanation" before the EX entered the room³.

In phase 2, the EX explains the board game *Quarto!* to the EE sitting vis-à-vis at a table. During the explanation, the game is *not* present due to reasons given in Sect. 3.2. After the instruction, the EX entered the room of the EE, sat down opposite of the EE and a researcher gave a starting signal, left the room, and the explanation phase started.

For the video recall, we observed the explanation via livestream and took notes of important scenes during the explanation. We used an identification scheme for the ad hoc selection of scenes to be used in the video recall. Criteria for selection were: substantial contributions, self- and other initiated repairs, turn taking, misunderstandings and questions that assessed understanding and knowledge of the EE. After the explanation ended (phase 2), we quickly started the video recall with the EX to investigate the EX's thoughts regarding specific points in the explanation [14], especially to discover reasons for addressing either of the two sides of the dual nature in their explanations.

² Original: "In dem nächsten Raum erklären Sie bitte dem Gegenüber das Spiel so gut, dass Ihr Gegenüber eine Chance hätte, das Spiel zu gewinnen".

³ Original: "Im nächsten Raum ist eine Person, die Ihnen ein Spiel erklärt, bitte nehmen Sie aktiv an der Erklärung teil".

For the video recall with the EX, we used a semi-structured interview which included questions about explanation content, explaining intention, explanation quality, the assumed EE's understanding of the game, and perceived knowledge needs of the EE. During the video recall, a researcher watched the videotaped explanation with the EX from start to end, and stopped at the previously selected scenes. This served as a stimulus to enable the EX to give detailed insights into certain sequences of interaction, their interpretation of a particular moment [14] by allowing participants to not solely rely on memory when trying to answer the questions [28]. Questions were answered before information has been transferred from short- to long-term memory to avoid conflation of experiences and conjoined memories [29] by starting phase 3 as quickly as possible. The whole video call procedure was standardized.

3.4 Analyses

This paper puts the focus on the analysis of the explanation of the game, especially on what was said and why regarding the dual nature of the explanandum. An analysis of the explanations content provides the “what was said at a certain time”. That analysis is, in a second step, supported by an analysis of the video recall to support the interpretation by investigating the underlying reasoning. The video recall data provides the EX justifications, and helps to understand the development of the explanation regarding the reasons for utterances addressing Architecture and Relevance at certain times.

Explanation Content Analysis. The explanations recorded in phase 2 served as the foundation for this analysis. Student assistants and colleagues from linguistics transcribed all recordings of the explanations according to the second complexity level (basic transcript) of the GAT2 transcription system [44]. The GAT2 transcription system level 2 segments the spoken text into intonational units. The length of the transcripts was between 118 and 653 intonational units ($M=332$, $SD=146$). In the final transcript, an intonational unit represented one line in the transcript.

We used qualitative content analysis [24, p. 70] to analyze the transcripts. The code system consisted of two deductive code categories: (1) utterances addressing the Architecture, and (2) utterances addressing the Relevance of the explanandum. We developed a coding manual over the course of multiple pilot studies during which the first and second coder (first author and a student assistant) discussed their coded segments in data sessions, and refined the coding manual. All transcripts were double coded using the final version of the manual. All coded segments were additionally labeled to represent whether the utterance was by the EX or EE. Inter-rater and intra-rater reliability were calculated using MaxQDA, which reports the Brennan and Prediger Kappa [2]. Across all transcripts, the inter-rater reliability between both coders was “Almost Perfect” ($k=0.80$) [25, p. 165]. The intra-rater reliability was calculated for both coders who coded four transcripts twice three weeks apart ($k=0.91$). The performance

of the coding manual is satisfactory for the aim of the study. Overall, the content of the segments coded with the two code categories align with the theoretical ideas, and thus we were able to identify which sequences of the explanations contained utterances addressing the Architecture or Relevance reliably.

Table 1. Overview of the Code Categories used for the Analysis of the Explanations' Content

Code	Examples
Utterances addressing Architecture (UA)	EX h and they will then be ALternately-put onto this BOARD, (P01, Pos. 55–56) if I would STARt now i would have to hand YOU one [a piece] (VP22, Pos. 213–214) and these PIECES have FOUR properties they are either uhm TALL or SMALL (VP06, Pos. 27–30)
Utterances addressing Relevance (UR)	EX SO what makes this game interestING is that uhm if YOU plAced one [piece] then YOU decide which piece your opPOnent places next (VP02, Pos. 204–208)

different plausible interpretations. Examples for that were transcripts starting with an example (“the GA:ME / is a variANT / of coNNECT four” (VP11, Pos. 5–7)⁴), or areas in which a shift from Architecture to Relevance occurred within one intonational phrase and thus both codes needed to be applied. In general, the content of each segment corresponded well (98%) to one of the code categories, thus allowing us to consistently identify whenever the Architecture or Relevance were addressed during an explanation. For an overview of the contents of each code category, see Table 1.

For the analyses, we used MaxQDA for coding, reliability calculations and the majority of qualitative work, and Python in subsequent steps, especially for visualizations and calculations, to investigate patterns regarding combination of sequences of the two code categories. For these visualizations, all transcripts were standardized in length. The starting point was always the first line in the transcript, that was coded. The end was the last line which received a code. For the standardization, all transcripts are divided into 100 sections with equal length and labeled regarding the codes used in each of the 100 sections, similar

A total of 878 segments were coded across all transcripts (605 utterances addressing the Architecture, and 273 utterances addressing the Relevance of *Quarto!*). On average, 56% ($SD=10,37\%$) of the transcripts were coded, while 44% ($SD=10,12\%$) received no code. This is mainly due to the transcription system which makes use of spaces and special characters to include additional information (e.g., pauses, breathing, overlaps). Another reason is that most explanations contained some small talk which was not part of the explanation. Out of 605 coded segments, 2,62% received both codes, as either one or the other category was possible due to

⁴ All examples we use from the transcripts of this paper were carefully translated from the German original to English by the first and second author of this paper.

to how MaxQDA realizes this feature⁵. We used these visualizations to identify specific points of interests in the corpora of explanations (e.g., switches from Architecture to Relevance, EE remarks regarding Relevance after a long phase of utterances about Architecture by the EX and vice versa). These points of interest served as a guide for the selection of video recalls we analyzed in the next section, which are used to underpin the analysis of the content of explanations.

Video Recall. To investigate the justification for the content of the explanations as perceived by the EX we analyzed the interview questions from the video recall. The video recall consisted of a larger variety of questions. Only questions regarding the EX's explaining intention, as well as the EX's perceived knowledge needs of the EE are analyzed. To achieve this, a student assistant transcribed the data from ten video recalls using standard orthography [35].

Material was coded using a deductive category system to categorize reasons for explanation content in consideration of the dual nature theory, and to categorize the reasons into two subordinated code categories: Knowledge Needs of the EE, and Explaining Intention of EX (see Table 2). Depending on whether a statement of the EX provided reasons for addressing "Architecture" or "Relevance" in their explanations, either the perceived knowledge needs of the EE or the explaining intention of the EX were coded with the corresponding sub-category. Statements which could not clearly be assigned to Architecture or Relevance were collected under a main category (Reasons for Explanation Content). The questions for the semi-structured interview guideline were derived from literature. We formulated the questions so that the explaining intention of the EX (why the EX wanted to explain specific aspects in certain situations irrespective of the situation, or the EE), as well as the knowledge needs of the EE as perceived by the EX, were focused on in the answers. We considered both of these aspects as important factors for the contents of explanations [10, 18, 26, 39].

We coded the material using a coding manual that we developed beforehand. The length of units of coding was flexible, ranging from at least one phrase to sometimes several sentences. Most important was that the unit encapsulated one specific reason stated by the EX (EX's explaining intention, or EE's knowledge needs). If multiple reasons were mentioned within a one answer to a question, each reason was assigned to a separate unit of coding [41]. All video recall transcripts were coded twice, and intra-rater reliability was "substantial" ($k = 0.72$) [25, p. 165], which is satisfactory as the data were complex. Only ten video recalls are part of this analysis, as the final solution for the video recall was not fully developed until the second half of data collection, especially due to finding a reliable technical solution.

⁵ If, for example, one section contained three lines of transcripts, one addressing the Architecture and two addressing the Relevance, the visualization represented this by drawing a rectangle which is to 1/3 colored according to the code Architecture and 2/3 colored according to the code Relevance in order of their occurrence in the transcripts. In other words, there is no large loss of precision in the visualizations as a result of the standardization of length.

4 Results

Table 2. Category System to code the reasoning for explanation content stated by EX during the video recall

Code Category	Subcategory
Architecture	Knowledge Needs of EE
	Explaining Intention of EX
Relevance	Knowledge Needs of EE
	Explaining Intention of EX

We start this section by presenting results from the analysis of the content of the explanations. Afterward, to underpin the commonalities we found across the explanations, we support the findings with the results from the analysis of the video recalls.

Explanation Content. The investigation of RQ1 (How is the dual nature of the explanandum expressed in the utter-

ances of the explanation?) is based on a qualitative content analysis. As an analytical framework, the techno philosophical theory of the dual nature provided two deductive code categories, which we used for coding the corpus.

Except for the EX in VP23, all EE and EX addressed both sides of the dual nature in the corpus (see Fig. 1). In general, the content of the coded segments within each category aligned well with the ideas of the dual nature theory. When the Architecture of *Quarto!* was addressed, EX and EE alike, used, for example, causal reasoning to address certain rules, described how the rules of the game result in a certain sequence in which each of the two players has to act, or described the physical makeup of different parts of the game. Whenever they addressed the aspects of Relevance of *Quarto!*, they shifted to teleological aspects, for example, that the complexity rises continuously throughout the game (“that means the MORE pieces are put onto the board (0.7) / the MORE you need to thInk / because there are MULTiple rows containing three pieces”, VP25, Pos. 174–176), strategical recommendations (“a:nd it is a: stra_strategical GAME / because if a few pieces are PLACed / then YOU are able to see / which PIECE / would be GOOD / to hand to me / such that i would not WIN / that means you also have to thiNK a little” VP16, Pos. 75–82), and emotional aspects for example how it is especially annoying to personally hand the piece to the opponent who may win using that piece in their next move (“EX (0.3) the frus-TRATIng bit about the game is that / that ONESELF is always the reason / for loSING / because one HANDS OVER (.) a piece” VP18, Pos. 372–375). A summary of the contents of the two categories, as well as examples from the transcripts, is provided in Table 1.

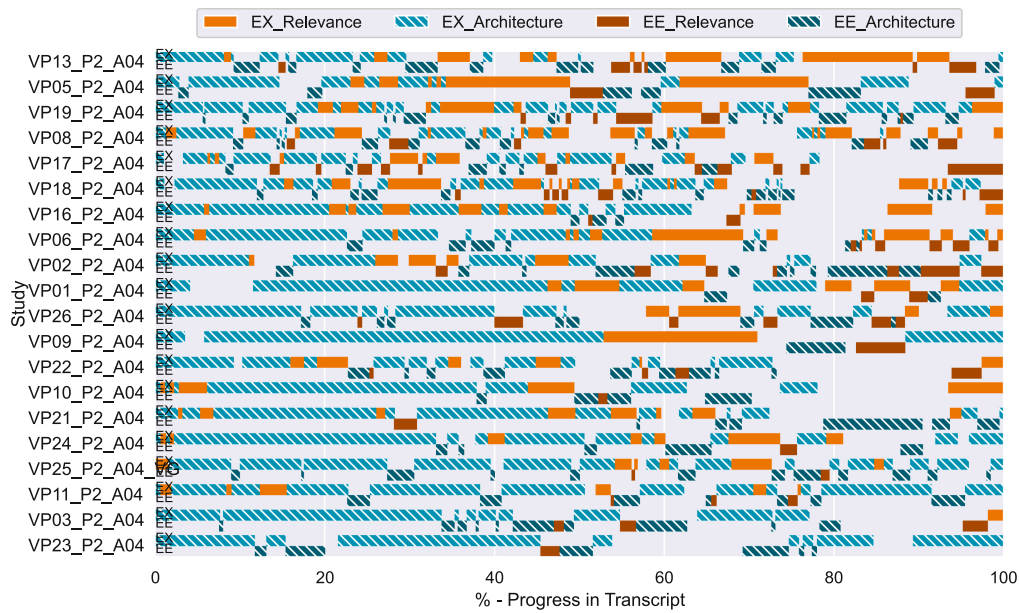


Fig. 1. Visualization of explanation content across studies. All transcripts standardized in length. Start- and endpoint are the first and last coded segments in the transcripts. Sorted by utterances addressing Architecture (least to most; EX = Explainer, and EE = Examinee)

Architecture Addressed More Frequently: Both EX and EE addressed each side of the dual nature in their explanations, but the balance of which side was focused on shifted throughout the explanations. While some EX (e.g., VP23, VP03) were focusing almost exclusively on aspects of the Architecture, other EX addressed aspects of the Relevance of *Quarto!* more often (e.g., VP13, VP05). The amount of talk EX and EE spent addressing each of the sides varied quite a lot across dyads, but on average, the Architecture was addressed more frequently than the Relevance for EX and EE (see Fig. 2).

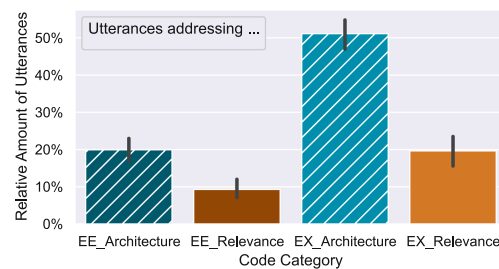


Fig. 2. Average code distribution across studies

Architecture First – Laying out the Tools? In the context of RQ2, we were especially interested in whether we would find any patterns across explanations regarding *when* and *in which configurations* the sides of the dual nature were addressed. A comparative visualization of all transcripts of the corpus grouped by utterances by EE and EX and colored according to which side of the dual nature was focused on throughout the transcripts, see Fig. 1. As indicated by the visualization, utterances addressing the Architecture were uttered more frequently overall. Especially in the early stages of the explanations, the aspects

of the Architecture were addressed more frequently than in the middle and last third (cf. Figure 3).

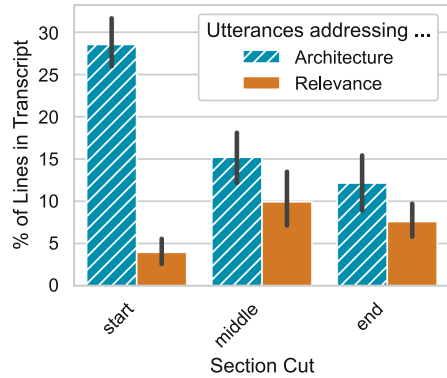


Fig. 3. Comparison of the percentage of lines in the transcripts addressing the two code categories by speaker sectioned in thirds.

third. But utterances about the Architecture were dominating the explanations across all dyads. Interestingly, when the dyads decided that the EE should share their understanding – that happened rather frequently at the end of the transcripts – the EE’s reiteration addressed mainly aspects of the Architecture, too.

Video Recall. For answering RQ3 (How do EX justify their choice of explanation content regarding which side of the dual nature was addressed?) we used qualitative content analysis, too. Based on the results from Sect. 4, we selected 20 scenes from the video recalls: ten scenes focusing on the beginning of the explanation, and ten scenes after the content shifted from Architecture to Relevance.

Reasons for Architecture First. Reasons for addressing Architecture at the beginning of an explanation were found in the EX’s explaining intention, i.e., what are the needs of the EX regarding the explanation content, and to a lesser degree in the EX reaction to EE’s knowledge needs. From a total of 28 reasons stated by EX for explaining content addressed to Architecture, 26 were rooted in the explaining intention of the EX irrespective of the EE. No EX mentioned the need to address aspects of the Relevance at the beginning. In two instances (VP10, VP13), the EE expressed knowledge needs at the beginning of the explanation. Yet interestingly, the EX did not justify the explanation content by a desire to address these knowledge needs (Fig. 4).

Figure 3 provides additional evidence that EX referred to aspects of the Architecture in the first third of their explanation more often than in other parts of the explanation. In the explanations, the content of the first third, especially EX, seemed to address important components and pieces of the game, which they often referred to later in their explanations. We describe this phenomenon as *Laying out the Tools*. In the later stages of the explanation, the utterances are more balanced regarding the side they address. In the second third of the explanations, utterances addressing Relevance were more prominent than in the first or last

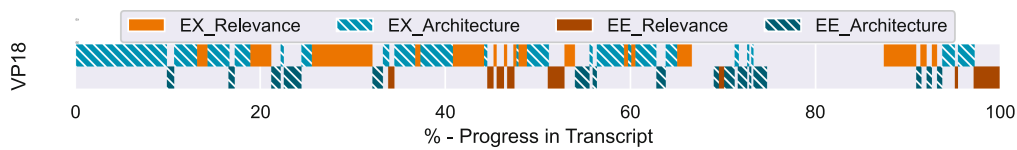


Fig. 4. Combination of analyses of explanation content, and EX’ justification for the example of VP18. The EX addressed the Architecture in roughly the first 15% and justified it using the following reasoning: “starting/ at first I thought it was important to know the components of the game before saying how (...) the game progresses” (VP18, VR1, Pos. 5)

The reasons for focusing on the Architecture at the beginning (cf. *Laying out the tools*) as stated by the EX were manifold. Most EX regarded it as important that the explanation has a certain structure, meaning that some aspects were believed to be prerequisites that need to be explained, before explaining further details. Two EX stated, for example, that explaining game components and their physical characteristics was a requirement to be able to explain the goal of the game (VP10, VP18). In VP06, the EX reported: “In the beginning I wanted to explain the course of the game swiftly before explaining in-depth” (VP06, VR1, Pos. 3). Another reason for explaining Architecture first was found in the EX’s wish to create an image of the game that the EE can imagine, and therefore explained game components and their appearance first. With a similar intention, one EX made a comparison to chess and described the game board as a smaller chessboard (VP21). In VP 24 the EX stated: “I just wanted to create a picture of the game so that he knows [...] what kind of game it is.” (VP24, VR1, Pos. 4).

Even though the EX primarily addressed Architecture in the beginning and thus Relevance is not addressed in the explanations, we found indicators that Relevance might already be part of an explanation. We found cases, in which Relevance is not expressed verbally, yet considerations of the Relevance had an impact on the structure of the explanation (e.g., VP13, VP17). One EX explained the different pieces and their similarities in detail and justified it by addressing the Relevance in the interview stating: “Because that’s what the goal of the game actually means to me. To find these similarities.” (VP17, VR1, Pos. 5). This addresses the Relevance by implying that getting four in a row is *not* important, but instead finding the similarities on the board is.

In summary, reasons for addressing the Architecture in the beginning were rooted predominantly in the EX’s explaining intention and covered, for example, structuring the explanation, as well as pictographic descriptions of main aspects.

Reasons for shifting to Relevance after addressing the Architecture: Reasons for shifting in later stages of the explanation were rooted in both the EX’s explaining intention and the perceived knowledge needs of the EE. In total, 17 reasons in the 10 scenes included reasons for addressing the Relevance, whereof nine were related to knowledge needs of the EE and 8 were related to the explaining intention of the EX.

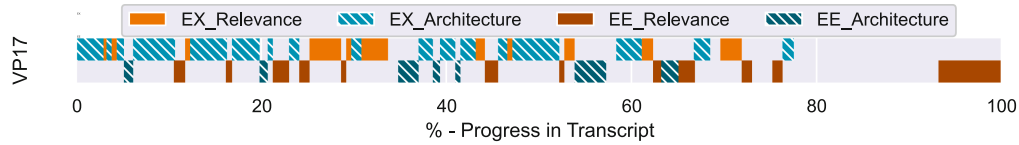


Fig. 5. Combination of analyses of explanation content, and EX’s justification in VP17. The EX shifted from Architecture to Relevance between 20% and 40% of the explanations progress, and justified it using the following reasoning: “My counterpart wanted to know why or what the rationale is for the fact that one decides for the opponent [which piece to play]. Who then could have won the game using that piece. Which he would have handed to me [and thus I can win instead].” (VP17, VR3, Pos. 7)

There were three main reasons for addressing the Relevance rooted in the knowledge needs of EE, ranging from the desire to be able to assess the game’s degree of complexity (VP6, VP24), to certain aspects that are important to consider when trying to build a row of four (VP25), and the expressed wish to learn more about the game’s strategies (VP13). The EX’s explaining intention, resulted in explanation content that was addressed towards Relevance, too (e.g., VP10, VP18, VP26). The EX of VP26 referred to gameplay experience: “It was about the exchange of my experience, it is not about knowledge, but about experience and about situations experienced during a round or the game.” (VP26, VR 5, Pos. 9). There were instances within the explanations for which the reasons could not be categorized to EE knowledge needs or EX explaining intention. In these cases, the reasons for addressing the Relevance in the explanation emerged from the interaction between both EX and EE. Sometimes, for example, questions were posed by the EE that addressed the Architecture and the EX replied by addressing the Architecture while integrating information regarding the Relevance as well (VP21, VP22). In VP21, the EE asked what the end of the game looks like and the EX answered the question by referring to the Architecture, but also addressed to Relevance by mentioning the difficulty of keeping an eye on every game detail as the game progresses (VP21). Another example of how shifts to Relevance occur within an explanation is found in VP17, where the EX was explaining content addressed towards Architecture, specifically that players have to choose pieces for the opponent. The EE struggled to find the rationale for this, especially as it would mean that one would have to hand the winning piece to the opponent, too, and by doing that forced a shift to Relevance in the explanation (see Fig. 5).

The reasons for explanation content after shifting from Architecture to Relevance can be highlighted as highly diverse and for now, it can be said that reasons are rooted in EE’s knowledge needs, the EX’ explaining intention, and in questions emerging from the interaction between both EX and EE during the explanation.

5 Discussion

This paper investigated how the content of a naturalistic explanation of the board game *Quarto!* evolves, and how this evolution of content is justified by the EX. Regarding potentials for XAI research, we believe that researching how explanations of technological artifacts evolve naturally is an important prerequisite for formulating recommendations on how to construct synthetic explanations. One of the aims of this study was to investigate how the dual nature of the explanandum is addressed in the utterances of the explanations (RQ1). Existing literature was very sparse regarding this question. The objective of the content analysis was to categorize utterances reliably into the two code categories: Architecture and Relevance. Throughout all explanations, both sides of the dual nature were addressed to a varying degree. Answering RQ2 (what are common patterns regarding sequences of utterances about the dual nature across the explanations?) resulted in a deeper understanding of when which side was addressed more frequently. We found that in most explanations, aspects of the game's Architecture were addressed in the first third of the explanation, and only in the middle and last third, aspects of the Relevance were addressed more frequently. Based on these findings, RQ3 (How do EX justify their choice of explanation content regarding which side of the dual nature was addressed?) helped to gain better insights to the reasons why EX explained predominantly the Architecture at the beginning of explanations, and how shifts from Architecture to Relevance were justified by the EX. We found that at the beginning of an explanation, especially the explaining intention of the EX, is the reason for explaining content addressed towards Architecture. In later stages, shifts from addressing Architecture to addressing Relevance were justified by the EX's explaining intention, the perceived knowledge needs of the EE, and their interactive sequences of EX and EE throughout the explanation.

The study design worked as anticipated. All dyads engaged quickly into the explanation after getting the start signal. The EX explained intuitively, and EE frequently asked questions or shared their confusion. Our observation is that all dyads were interested in a successful explanation, and motivated to continue the explanation, until they thought they had finished. The comparability across explanations was satisfactory, as the time and content did not deviate too much and dyads kept their focus on explaining in such a way, that the EE would be able to win the game, too. Thus, instruction worked as well. Our overarching motivation is that finding certain patterns would provide insights useful for future research on other explananda to, eventually decide, for example, whether certain patterns are generalizable and maybe even a useful scheme for synthetic explanations, too.

As stated earlier, we believe, the theory provides a rich background for research on everyday explanations of technological artifacts and offers potentials for XAI research, too. Based on our findings, we still believe that. It could provide a foundation for formulating recommendations on whether and which of the artifacts properties should be referred to within a (synthetic) explanation. In our study, EX tended to put more effort into explaining aspects that address the

Architecture than the Relevance. In the case of *Quarto!*, the questions of which side of the dual nature is more important or whether one is a precursor for the other (cf. Sect. 2) can thus be answered: Explaining the Architecture first before diving into what EX considered to be more complex details of the Relevance of the game, seemed to be a natural way to approach an explanation in the context of *Quarto!*. Regarding what we saw in the explanations, we would interpret it as a verbal realization of *laying out the tools*, by addressing the bits of Architecture which the EX can refer to in later stages of the explanation to connect aspects of the Architecture of *Quarto!* to aspects of its Relevance. Thus, we believe that, especially looking at the fairly large percentage of utterances addressing the Relevance with the justifications that the EX stated in the video recalls, provide a solid foundation for the claim that **both** sides can be considered important in the case of *Quarto!*. Whether and how these findings are generalizable, especially for more complex artifacts, is still an open research question.

The question of (RQ3) how do EX justify their choice of explanation content regarding which side of the dual nature was addressed, brought further insights into the reasons why the EX addressed either the Architecture or Relevance. The EX provided a broad range of reasons for explaining aspects addressed to Architecture and Relevance at different stages of the explanation. The variety of different reasons indicate, that deciding on which aspects to explain is a process in which the EX has to consider a large variety of factors, not only concerning their intention but also guided by what they intend to explain. At the beginning of the explanations, EX dominated the explanation and the reasons for addressing the Architecture are usually rooted in the EX's explaining intention. Therefore, the first part of the explanation follows the explanation goal and plan of the EX without further consideration of the specific needs of the EE, albeit important for understanding [5, 26]. EX mentioned reasons like explaining physical components and their characteristics and appearance first as a preparation for providing more details later in the explanation, especially in regard to shifting to Relevance. At times, the shift from Architecture to Relevance did not occur suddenly, but emerged gradually in the context of the interaction of both participants. In these instances, even though the explaining intention of the EX was still apparent, the knowledge needs of EE became more evident. Reasons for that were not mentioned, but could potentially be out of general interest or due to identifying certain misunderstandings [10]. Aspects that were addressed to Relevance were seldom part of the first third of the explanations. Yet, our analyses of the video recall showed that in some instances, EX structured their explanations with aspects of the Relevance in mind. They interpreted these bits of information as highly important for their explanation. This might be an indicator that Relevance is (1) an important part of an explanation and (2) even though verbally the Relevance is not yet addressed in the explanation, it still guides the choice of which aspects of the Architecture are addressed at the start.

The theoretical background puts focus on the characteristics of technological artifacts, but empirical investigations only addressed more technical, text-based explanations (e.g., patents) [37]. As texts are very different from extemporaneous

explaining, we were surprised how clearly, and easily distinguishable Architecture and Relevance were addressed within the explanations of *Quarto!*. In the cognitive sciences, the theory of people taking different stances to explain things may be useful for explaining this phenomenon [12, 19]. Two stances seem to be connected to addressing the sides of the dual nature. Whereas in the context of the dual nature theory, the question was raised whether one side is a precursor or more important, the cognitive sciences spent time researching how explanations from different stances result in different understanding. If (and only if!) these stances would be connected to addressing either the Architecture or the Relevance, our findings *may* support the hypothesis, that by categorizing utterances into either of the two code categories of the dual nature, in the case of *Quarto!*, we identified areas in which these stances were observable. If this was the case, interestingly, our dyads shifted between stances fairly quickly and multiple times throughout one explanation.

Potentials for XAI Research. We are convinced that researching how explanations of technological artifacts evolve naturally is an important prerequisite for formulating recommendations for the construction of synthetic explanations. The limitations of popular notions of explanations in XAI elaborated in [38] provide a good rationale for why research such as this is required for *real explanations* in XAI: understanding human-to-human (H-H) explanations and their co-constructive nature offers large potential for insights on what exactly is important for human understanding in explanations of artifacts. It is needed to establish a foundation of knowledge for formulating, for example, recommendations on whether and which of the artifacts properties should be referred to within a synthetic explanation and, for example, how that is related to certain knowledge needs or explaining intentions of EE and EX. Our findings, and our methodology, provide valuable tools for future research, too. To close this thought, we share an insight that surprised us in our studies: We did not anticipate the pattern that utterances about the Architecture were more present in the first third of the explanations. Instead, our intuition was guided by the idea of providing a source of motivation to create an atmosphere that is attractive, regularly referred to as good practice in teaching and education [4]. We intuitively expected that the EX would focus on the Relevance of *Quarto!* at the start of the explanations, to provide a source to stimulate the EE's motivation (e.g., by giving reasons why the game is exciting or fun to play). Yet, the opposite happened. Even if just anecdotal, we believe it provides an important learning: Simply sticking to popular beliefs or plausibility for constructing synthetic explanations already skips important steps. Instead, taking these steps in the shape of research on naturalistic, everyday H-H explanations of artifacts are an important prerequisite for the field and thus for *real* XAI.

Quality Standards. We adhered to a variety of recommendations [24, 33]. We documented the different stages of codings from initial codings until the final codings of the transcripts, especially for being able to reconstruct the interrater

data sessions in which the codings were discussed between the two coders and adjusted accordingly. Inter-rater and intra-rater reliability was examined for the analysis of the explanation content, intra-rater reliability was assessed for the video recalls. Additionally, three independent coders unfamiliar with the theoretical background used the coding manual for the explanation content analysis and coded three transcripts. We compared the results to our codings using inter-coder reliability tests ($k=0.53$). All reliability measures were good (see Sect. 3.4). Parts of the final codings were discussed within and outside the research project (peer debriefing). Earlier results were presented and discussed during a research retreat of the Transregional Collaborative Research Centre TRR 318 “Constructing Explainability”. To prevent false conclusions, and for a deeper understanding of the practical implications of the study design, all researchers involved in this publication were present during at least a quarter of the studies (experiencing the study). We triangulated the results from the content analysis with EX justifications stated in the video recalls to underpin, contrast or contradict our interpretations.

5.1 Limitations and Future Work

We found two commonalities across the explanations we analyzed. First, the majority of time was spent addressing the Architecture, especially in the first third of the explanation. Only later, the content was more balanced between utterances addressing the Architecture and Relevance. Whether this applies for other technological artifacts, too, is still an open research question that this study cannot answer. We believe that, especially if artifacts were more complex, describing the most important aspects of the Architecture at the beginning would maybe not be possible as (1) the amount of information could simply not be remembered (i.e., too much information) and (2) the amount of information would require explanations that would be far too long to be interesting for XAI. But, our insights open up new research opportunities and questions:

- Is *laying out the tools* a generally reasonable strategy when explaining technological artifacts? Do circumstances exist, in which it would make more sense to start with addressing the Relevance?
- How much focus on the Relevance is needed (and when) to provide a *good* explanation?
- What explanation strategies are used to start the explanation, if the explanandum is a very complicated, technological artifact (i.e., what are other strategies opposed to *laying out the tools*)?
- How is the level of understanding of the EE related to certain patterns or the ratio between the utterances addressing the two sides in the explanations?
- How is the EE ability to competently play the game connected to knowledge about both sides conveyed in the content of the explanations?

Based on our results, we now are in a position in which we claim that using the dual nature theory as an analytical tool to structure the content of explanations of technical artifacts is a worthwhile, and a viable approach. Coding the

content of the explanations is quick, and the structure provided by the codes allows quick insights into how the explanation was carried out. Based on this, further analyses can be carried out which ultimately may help for recommendations for the construction of synthetic explanations, too. In the theoretical background elaborated in Sect. 2, we addressed how different researchers have different perceptions regarding this question, yet more empirical support for the claims is needed. Our ongoing research will address such questions of which explanations, regarding their structure of addressing the dual nature of *Quarto!*, are better, that is, result in a deeper understanding of the EE. Currently, we are in the process of developing an instrument for such an assessment.

One remark that is necessary is, that in the case of games, the need for knowing the Architecture of the game is inherently clear. For example, if one does not know the shapes of the pieces in *Quarto!*, one would not be able to play it competently. Especially in the case of digital artifacts, this is not necessarily needed⁶. Therefore, if researching explanations of digital artifacts, another layer of complexity is added, which most likely also influences how easy it is to distinguish which side of the dual nature is addressed. Yet, as a first step, by investigating an example in which the need to understand the Architecture is so prominent, we gained insights that could be fundamental for a deeper understanding of explanations of more complex, and especially digital artifacts.

To further support, contrast or contradict our results, our study design could be changed in a way in which one participant is previously trained to influence the explanation by drawing the attention to either the Architecture or the Relevance, for example by instructing them to have certain knowledge needs (e.g., you want to program the game afterward vs. you want to know whether the game is a suitable present for a friend). Controlling that variable is promising to get a better understanding if one side may also be omitted completely or if it would always emerge naturally, regardless of what was instructed before.

The video recall method has some limitations, for example, the anxiousness of the participants, self-censorship and reduced visual hints through fixed perspective and technical format [6]. We addressed this by creating an atmosphere in the lab that reduced factors leading to stress or anxiety in the participants and gave room for self-exploration. By using this introspective method, we were able to better identify the reasons for explaining specific aspects of the game.

In some instances, the EX verbally remarked that they were surprised about the absence of the game during phase 2. Thus, the absence of the game is a potential factor influencing the beginning of the explanations quite drastically. But, as elaborated in 3.2, we believe this circumstance to be the norm rather than the exception, as it created a black box scenario. Yet, this factor needs to be addressed in future research. However, due to the low number of instances, we believe it to have only a minor effect on the naturalness of the explanations.

⁶ Research areas such as user experience and interface design are a good example in this regard. Put very briefly and naively, they are interested in circumventing the need to understand the Architecture by providing enough Relevance to be able to use things competently.

5.2 Conclusion

In the case of natural explanations of the board game *Quarto!*, EX referred to a carefully selected set of information about the Architecture of the game in the first third of the explanation. Their rationale was that such information (the *What*) is a prerequisite for information that they considered to be more complex (the *why*). They did this to make the more complex content of the explanation more accessible for the EE later in their explanations. At later stages of the explanations, the reasons for shifting from addressing the Architecture (the *what*) to addressing the Relevance (the *why*) were rooted in a larger variety of reasons, one of them being that they slowly emerged from the interaction between EX and EE. Our findings are a first step towards a more general understanding of extemporaneous explanations of technological artifacts, especially regarding when and how the two sides of the dual nature are addressed in explanations. Other researchers suggested that such an approach is not only viable, but necessary for the development of XIA systems which can explain *naturally* and thus for a broad audience [38]. A better conceptualization of these naturalistic explanations of technological artifacts can provide crucial insights into how XAI can explain its decision more naturally. Especially if certain patterns – such as adding why to what – are generalizable and thus could be embedded in an interactive explanation system, this could improve the interpretability of decisions, and ultimately support the agency of a large variety of people interacting with XAI.

Acknowledgements

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation): TRR 318/1 2021 - 438445824. We thank Vivien Lohmer and Friederike Kern for their input during discussions, as well as for creating the transcripts. We thank all research assistants for their support during coding, data acquisition and data analysis.

Ethics Statement

This study was approved by the Paderborn University Ethics Board. All participants participated voluntarily and provided written informed consent before the studies. The studies were conducted in concordance with local COVID-19 policies.

References

1. Attisano, E., Nancekivell, S.E., Denison, S.: Components and mechanisms: how children talk about machines in museum exhibits. *Front. Psychol.* **12**, 1737 (2021). <https://doi.org/10.3389/fpsyg.2021.636601>
2. Brennan, R.L., Prediger, D.J.: Coefficient kappa: some uses, misuses, and alternatives. *Educ. Psychol. Measur.* **41**(3), 687–699 (1981). <https://doi.org/10.1177/001316448104100307>
3. Brennan, S.E., Hanna, J.E.: Partner-specific adaptation in dialog. *Top. Cogn. Sci.* **1**(2), 274–291 (2009). <https://doi.org/10.1111/j.1756-8765.2009.01019.x>
4. Brophy, J.E.: *Motivating Students to Learn*. 3rd edn. Routledge, New York (2010). <https://doi.org/10.4324/9780203858318>
5. Buhl, H.M.: Partner orientation and speaker's knowledge as conflicting parameters in language production. *J. Psycholinguist. Res.* **30**(6), 549–567 (2001). <https://doi.org/10.1023/A:1014217421749>
6. Calderhead, J.: Stimulated recall: a method for research on teaching. *Br. J. Educ. Psychol.* **51**(2), 211–217 (1981). <https://doi.org/10.1111/j.2044-8279.1981.tb02474.x>
7. Cederqvist, A.M.: Pupils' ways of understanding programmed technological solutions when analysing structure and function. *Educ. Inf. Technol.* **25**(2), 1039–1065 (2020). <https://doi.org/10.1007/s10639-019-10006-4>
8. Cederqvist, A.M.: Seeing the parts, understanding the whole - A technology education perspective on teaching and learning in processes of analysing and designing programmed technological solutions. University of Gothenburg (2021)
9. Chi, M.T.H.: Self-explaining: the dual processes of generating inference and repairing mental models. In: *Advances in Instructional Psychology: Educational Design and Cognitive Science*, vol. 5, pp. 161–238. Lawrence Erlbaum Associates Publishers, Mahwah, NJ, US (2000)
10. Chi, M.T.H., Siler, S.A., Jeong, H.: Can tutors monitor students' understanding accurately? *Cogn. Instr.* **22**(3), 363–387 (2004). https://doi.org/10.1207/s1532690xci2203_4
11. Chin-Parker, S., Bradner, A.: A contrastive account of explanation generation. *Psychon. Bull. Rev.* **24**(5), 1387–1397 (2017). <https://doi.org/10.3758/s13423-017-1349-x>
12. Dennett, D.C.: *The Intentional Stance*. MIT Press, Cambridge (1987)
13. El-Assady, M., et al.: Towards XAI: structuring the processes of explanations. In: *Proceedings of the ACM Workshop on Human-Centered Machine Learning*, Glasgow, UK (2019)
14. Fox-Turnbull, W.: Autophotography. In: Benson, C., Lunt, J. (eds.) *International Handbook of Primary Technology Education: Reviewing the Past Twenty Years*, pp. 195–209. International Technology Education Studies, SensePublishers, Rotterdam (2011). https://doi.org/10.1007/978-94-6091-546-8_16
15. Gilbert, J.K., Boulter, C., Rutherford, M.: Models in explanations, Part 1: horses for courses? *Int. J. Sci. Educ.* **20**(1), 83–97 (1998). <https://doi.org/10.1080/0950069980200106>
16. Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., Yang, G.Z.: XAI-explainable artificial intelligence. *Sci. Robot.* **4**(37), eaay7120 (2019). <https://doi.org/10.1126/scirobotics.aay7120>
17. Hale, C.R., Barsalou, L.W.: Explanation content and construction during system learning and troubleshooting. *J. Learn. Sci.* **4**(4), 385–436 (1995). <https://doi.org/10.1207/s15327809jls0404.2>

18. Heider, F.: *The Psychology of Interpersonal Relations*. John Wiley & Sons Inc, Hoboken (1958). <https://doi.org/10.1037/10628-000>
19. Keil, F.C.: The birth and nurturance of concepts by domains: the origins of concepts of living things. In: *Mapping the Mind: Domain Specificity in Cognition and Culture*, pp. 234–254. Cambridge University Press, New York, NY, US (1994). <https://doi.org/10.1017/CBO9780511752902.010>
20. Keil, F.C.: Explanation and understanding. *Annu. Rev. Psychol.* **57**(1), 227–254 (2006). <https://doi.org/10.1146/annurev.psych.57.102904.190100>
21. Kroes, P., Meijers, A.W.M.: The dual nature of technical artifacts?: presentation of a new research programme. *Techné* **6**(2), 4–8 (2002)
22. Kroes, P.: technological explanations. *Phil. Tech.* **3**(3), 124–134 (1998). <https://doi.org/10.5840/techne19983325>
23. Kroes, P.: Design methodology and the nature of technical artefacts. *Des. Stud.* **23**(3), 287–302 (2002). [https://doi.org/10.1016/S0142-694X\(01\)00039-4](https://doi.org/10.1016/S0142-694X(01)00039-4)
24. Kuckartz, U.: *Qualitative Text Analysis: A Guide to Methods. Practice and Using Software*, SAGE (2014)
25. Landis, J.R., Koch, G.G.: The measurement of observer agreement for categorical data. *Biometrics* **33**(1), 159–174 (1977). <https://doi.org/10.2307/2529310>
26. Levelt, W.J.M., Le Page, R.B., Longuet-Higgins, H.C., Longuet-Higgins, H.C., Lyons, J., Broadbent, D.E.: The speaker’s linearization problem. *Philos. Trans. Royal Soc. London. B, Biol. Sci.* **295**(1077), 305–315 (1981). <https://doi.org/10.1098/rstb.1981.0142>
27. Lombrozo, T., Carey, S.: Functional explanation and the function of explanation. *Cognition* **99**(2), 167–204 (2006). <https://doi.org/10.1016/j.cognition.2004.12.009>
28. Lyle, J.: Stimulated recall: a report on its use in naturalistic research. *Br. Edu. Res. J.* **29**(6), 861–878 (2003). <https://doi.org/10.1080/0141192032000137349>
29. Mackey, A., Gass, S.M.: *Second language research: methodology and design. Methodology and design*, Lawrence Erlbaum Associates Publishers, Mahwah, NJ, US, Second Language Research (2005)
30. Malle, B.F.: *How the Mind Explains Behavior: Folk Explanations, Meaning, and Social Interaction*. MIT Press, Cambridge (2004). <https://direct.mit.edu/books/book/2642/How-the-Mind-Explains-BehaviorFolk-Explanations>
31. McNeill, D.: *Hand and Mind: What Gestures Reveal About Thought*. University of Chicago Press, Chicago (1992)
32. McNeill, D.: *Gesture and Thought*. The University of Chicago Press, Chicago (2008). oCLC: 781253715
33. Miles, M.B., Huberman, A.M., Saldaña, J.: *Qualitative Data Analysis: A Methods Sourcebook*. 3rd edn. SAGE Publications Inc, Thousand Oaks (2014)
34. Miller, T.: Explanation in artificial intelligence: insights from the social sciences. *Artif. Intell.* **267**, 1–38 (2019). <https://doi.org/10.1016/j.artint.2018.07.007>
35. O’Connell, D.C., Kowal, S.: Transcription systems for spoken discourse. In: *The Pragmatics of Interaction*, pp. 240–254 (2009)
36. Pennington, N.: Stimulus structures and mental representations in expert comprehension of computer programs. *Cogn. Psychol.* **19**(3), 295–341 (1987). [https://doi.org/10.1016/0010-0285\(87\)90007-7](https://doi.org/10.1016/0010-0285(87)90007-7)
37. de Ridder, J.: Mechanistic artefact explanation. *Stud. Hist. Philos. Sci. Part A* **37**(1), 81–96 (2006). <https://doi.org/10.1016/j.shpsa.2005.12.009>
38. Rohlfing, K.J., et al.: Explanation as a social practice: toward a conceptual framework for the social design of AI systems. *IEEE Trans. Cogn. Develop. Syst.* **13**(3), 717–728 (2021). <https://doi.org/10.1109/TCDS.2020.3044366>

39. Roscoe, R.D.: Self-monitoring and knowledge-building in learning by teaching. *Instr. Sci.* **42**(3), 327–351 (2014). <https://doi.org/10.1007/s11251-013-9283-4>
40. Salmon, W.C.: *Four Decades of Scientific Explanation*. University of Pittsburgh Press, Pittsburgh (1990)
41. Schreier, M.: *Qualitative Content Analysis in Practice*. SAGE, Los Angeles (2012)
42. Schulte, C.: Block model - an educational model of program comprehension as a tool for a scholarly approach to teaching. In: *Proceeding of the Fourth international Workshop on Computing Education Research*, pp. 149–160. ICER 2008, ACM, Sydney, Australia (2008). <https://doi.org/10.1145/1404520.1404535>
43. Schulte, C.: Duality reconstruction - teaching digital artifacts from a socio-technical perspective. In: Mittermeir, R.T., Sysło, M.M. (eds.) *Informatics Education - Supporting Computational Thinking*. ISSEP 2008. *Lecture Notes in Computer Science*, vol. 5090, pp. 110–121. Springer, (2008). https://doi.org/10.1007/978-3-540-69924-8_10
44. Selting, M., et al.: Gesprächsanalytisches Transkriptionssystem (GAT 2). *Gesprächsforsch* **10**, 152–183 (2009)
45. Soloway, E.: Learning to program = learning to construct mechanisms and explanations. *Commun. ACM* **29**(9), 850–858 (1986). <https://doi.org/10.1145/6592.6594>
46. Vermaas, P., Kroes, P., van de Poel, I., Franssen, M., Houkes, W.: A philosophy of technology: from technical artefacts to sociotechnical systems. *Synth. Lect. Eng. Technol. Soc.* **6**(1), 1–134 (2011). <https://doi.org/10.2200/S00321ED1V01Y201012ETS014>
47. Vermaas, P.E., Houkes, W.: Technical functions: a drawbridge between the intentional and structural natures of technical artefacts. *Stud. Hist. Philos. Sci. Part A* **37**(1), 5–18 (2006). <https://doi.org/10.1016/j.shpsa.2005.12.002>
48. van der Waa, J., Nieuwburg, E., Cremers, A., Neerincx, M.: Evaluating XAI: a comparison of rule-based and example-based explanations. *Artif. Intell.* **291**, 103404 (2021). <https://doi.org/10.1016/j.artint.2020.103404>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



3.3 Building Block 3: Monitoring of and Adapting to Explainees

Building Block 3

L. Terfloth et al. (n.d.). “Navigating the Dual Nature: Do Explainers Adapt to Explainee Interests When Explaining Technical Artifacts”. en. In: *International Journal of Technology and Design Education* () (accepted for publication)

3.3.1 Summary

Building on the naturalistic patterns identified in Building Block 2, this study investigates whether explainers actively adapt their explanations when confronted with explainees who express a clear interest in either the Architecture or the Relevance of a technical artifact. Using a controlled confederate design ($N=72$), trained student assistants posed as explainees and systematically displayed interest in either only Architecture or only Relevance while an unsuspecting explainer explained the board game Quarto. Three dependent variables were examined: (1) explainers’ mental representations of explainee interests and knowledge, assessed via pre-post questionnaires; (2) explanation content, coded as Architecture versus Relevance utterances using qualitative content analysis; and (3) gestural behavior, coded as iconic gestures (depicting physical properties of the artifact, such as tracing the shape of the board) versus pragmatic gestures (structuring the discourse, such as beats or pointing to mark transitions) in an exploratory subsample of 12 interactions. The manipulation was highly effective, with confederates maintaining distinct interest profiles across all phases of the explanation.

Results confirmed partial adaptation. Explainers successfully recognized explainee interest in Relevance but represented Architecture interest equally across conditions, suggesting Architecture is treated as a baseline assumption rather than an interest requiring active monitoring. Content analysis revealed that explainers adapted their explanations to explainee interests by increasing Relevance utterances from 28% to 40% in the Relevance condition while maintaining Architecture-dominant explanations (60–72%) regardless of condition. Both sides of the dual nature were addressed in all explanations, providing empirical support for the theory-based assumption that Architecture and Relevance need to be integrated for comprehensive understanding. Contrary to predictions, in an exploratory subsample analysis, explainers used significantly *fewer* iconic gestures in the Architecture condition (8.7%) than in the Relevance condition (13.7%), a finding tentatively interpreted as gestural compensation: when explainees showed disinterest in Architecture, explainers may have emphasized Architecture content more vigorously

through gesture. Pragmatic gestures remained stable across conditions. The paper discusses implications for dual nature theory, education, and the design of adaptive explanation systems.

3.3.2 Authors' Contributions

Lutz Terfloth developed the submitted manuscript from an initial 8-page working draft jointly written by Heike Buhl, Friederike Kern, Vivien Lohmer, Michael Schaffer, and himself, substantially restructuring, reframing, and extending it into its final form. He led the qualitative content analysis of the explanations, conducted the treatment check, and finalized all statistical analyses throughout the paper, building on preliminary analyses initially provided by Michael Schaffer. Lutz Terfloth wrote the majority of the final manuscript, including all theoretical framing sections, finalized all sections for submission, and revised the paper in response to peer review. Heike Buhl originated the conceptual idea for the study, supervised Michael Schaffer's contributions, and provided feedback on the manuscript with specific attention to the statistical analyses. Michael Schaffer provided the raw data of questionnaire instruments adapted for earlier studies (cf. Schaffer et al., 2024). Vivien Lohmer conducted the gesture coding and analysis and wrote the corresponding sections. Friederike Kern provided feedback on sections addressing linguistic, conversation-analytic, and ethnomethodological theory. Carsten Schulte supervised Lutz Terfloth's work, contributed to the theoretical framing within XAI and computing education research, and provided feedback on successive drafts.

Lutz Terfloth did greater than or equal to 70% of the work.

Acknowledgements

The experimental setting was designed as part of a research proposal by Heike Buhl, Friederike Kern, and Carsten Schulte within the TRR 318 *Constructing Explainability*. I thank all student assistants for their support during data acquisition and coding.

3.3.3 Author's Commentary

This paper picks up the central open question that Building Block 2 could not answer: Building Block 2 established that explainers naturally adopt an Architecture-first strategy and adjust toward Relevance over the course of an explanation, but it could not determine whether this shift was genuinely responsive to the explainee or simply an artifact of how explainers structure explanations regardless of who they are talking to. The confederate design here was constructed specifically to

force that question. By systematically manipulating explainee interest while holding everything else constant, it becomes possible to disentangle what explainers do by default from what they do in response to the person in front of them.

The answer is nuanced and, in some respects, surprising. Explainers do adapt: the 12 percentage point shift in Relevance content between conditions is both statistically reliable and practically meaningful. But the adaptation operates within a remarkably stable Architecture-dominant baseline. Even when explainees persistently steered the conversation toward Relevance, explainers maintained roughly 60% Architecture content. This suggests that the Architecture-first pattern identified in Building Block 2 is not merely a default that dissolves under co-constructive pressure, but something more deeply ingrained in how explainers conceptualize what an explanation of a technical artifact requires.

Theoretically, this finding is significant for the Co-Constructive Dual Nature Framework developed in this dissertation. It provides empirical evidence that the dual nature tension, the pull toward addressing both Architecture and Relevance, operates even when the collaborative dynamics of the explanation actively push in a single direction. Explainers are monitoring explainee interest and scaffold accordingly, but they are simultaneously maintaining a content structure they appear to regard as effective. Whether this reflects sound pedagogical intuition or a cognitive default that may sometimes work against optimal explanation is a question Building Block 4 is designed to address.

The gestural finding warrants particular attention because it contradicts the hypothesis derived from prior work. The expectation was that Architecture-focused explanations would produce more iconic gestures, given the representational potential of gesture for depicting physical structures. Instead, the opposite was found. The most plausible interpretation, that explainers compensate with more iconic gesturing precisely when explainees signal disinterest in Architecture, opens an interesting line of inquiry into how nonverbal behavior functions as a co-constructive tool in explanation, and points toward the kind of microanalytic follow-up that the multimodal methods developed in this project are positioned to pursue. A first step in this direction is reported in Lohmer and Kern (2024), who—in the same corpus—found that interactive gestures, particularly palm–vertical open–hand gestures and beats, are systematically deployed at transition points between explanation nodes, suggesting that the discursive structuring of Architecture and Relevance content has gestural correlates beyond the iconic–pragmatic distinction examined in the present paper.

Navigating the Dual Nature: Do Explainers Adapt to Explainee Interests When Explaining Technical Artifacts

Lutz Terfloth^{1*}, Heike M. Buhl², Vivien Lohmer³,
Friederike Kern³, Michael Schaffer², Carsten Schulte¹

^{1*}Computing Education Research, Paderborn University.

²Psychology, Paderborn University, Germany.

³Linguistics, Bielefeld University, Germany.

*Corresponding author(s). E-mail(s): lutz.terfloth@uni-paderborn.de;

Contributing authors: heike.buhl@uni-paderborn.de;

vivien.lohmer@uni-bielefeld.de; friederike.kern@uni-bielefeld.de;

michael.schaffer@uni-paderborn.de; carsten.schulte@uni-paderborn.de;

Abstract

Technical artifacts possess a dual nature comprising Architecture (how they are built) and Relevance (what they are for), both theoretically necessary for comprehensive understanding. This experimental study (N=72) investigated whether explainers (EXs) adapt their explanations to different explainees' (EEs) interests in these dimensions. Using confederates as EEs showing interest in either Architecture or Relevance, we analyzed EXs' mental representations of EE interests and knowledge, explanation content distribution, and gesture use when explaining the board game Quarto!. Results revealed that EXs successfully recognized EEs' interest in Relevance but represented Architecture interest equally across conditions, suggesting Architecture interest is assumed as baseline. While EXs adapted content to EE interests—increasing Relevance-focused utterances from 28% to 40% when EEs showed Relevance interest—they maintained Architecture-dominant explanations (60-72%) regardless of condition. This persistence suggests Architecture serves as an explanatory foundation that EXs deem necessary despite recognizing different EE preferences. Contrary to predictions, EXs used significantly fewer iconic gestures in the Architecture condition (8.7%) compared to the Relevance condition (13.7%), while pragmatic gestures remained stable across conditions. These findings provide empirical

support for the dual nature theory by revealing how EXs balance addressing both dimensions with responsive adaptation to EE interests. The stability of Architecture-dominant explanations across conditions raises questions about whether this reflects pedagogical intuitions, or the inherent structure of technical artifact understanding.

Keywords: Dual Nature Theory, Technical Artifacts, Adaptive Explanation, Grounding, Gestures

1 Introduction

Whenever one person explains something to someone else, the interlocutors adapt to each other in various ways (cf. [Garrod and Pickering 2004](#); [Dale et al. 2014](#)). In the explanation context, for example, a constant negotiation between what the explainer (EX) considers important and what the explainee (EE) wants to understand takes place. This means EXs can often tell whether the EE asking about a leaky faucet wants step-by-step instructions to fix it, or wants to know more about plumbing in general. The EXs adapt their explanation to match the desired level of detail and practical needs. The ongoing adaptation is based on evidence gathered by monitoring the recipients' reactions, questions, and apparent interests. As a result of this monitoring process, the EXs provide scaffolding by adjusting their approach accordingly. Indeed, both parties adapt to each other to establish a common ground in a process called grounding ([Clark and Brennan 1991](#)). But how do these processes unfold when the object of explanation itself has multiple inherent dimensions that could be addressed? Technical artifacts present exactly this challenge: Under the dual nature theory, every technical artifact¹ can be understood to possess two distinct dimensions: that is how it is built (its *Architecture*) and what it is for (its *Relevance*) ([Kroes 2009](#); [Winkelkemper et al. 2024](#))². Architecture encompasses implementation details, while Relevance includes subjective and objective purposes as well as potential societal impacts. According to the theory, both dimensions are necessary for a comprehensive understanding of the artifact and its design ([Kroes 1998, 2009](#)). While experts can competently bridge the gap between the two sides ([Kroes 1998](#); [de Ridder 2007](#)), students often struggle to make these connections ([Soloway 1986](#); [Spohrer et al. 1985](#)), and one-sided explanations risk being either technically detailed but meaningless or socially relevant but baseless ([Winkelkemper et al. 2024](#)). Despite the theoretical framework being well-established, surprisingly little is known empirically about how EXs navigate these dual dimensions in practice, particularly when learners show a preferred interest in one dimension over the other.

The theoretical framework of dual nature thus has direct implications for technology education, where instructors routinely encounter questions like "why do we

¹Henceforth we use the terms technical artifact and artifact interchangeably. Both refer to technical artifacts, not artifacts in general (e.g., they refer to a smartphone, but not a painting)

²Different terms are used throughout the literature. At its origin, the theory was framed through the duality of structure and function ([Kroes 2009](#); [de Ridder 2007](#)). In more recent publications, the terms were also adapted for digital artifacts, such as smartphones, which is why we decided to stick with the terms used and defined in ([Winkelkemper et al. 2024](#)).

need to learn this?”. When teaching programming, for example, educators face the challenge of navigating between technical implementation details and the artifact’s functions and purposes. Do instructors detect learner interests in one dimension, and how do they adapt their explanations accordingly? Do they lean into the dimension of main interest, maintain balance between both sides, or attempt to spark interest in the neglected side? When learners explicitly show interest in just one side, do EXs still feel compelled to address both dimensions of the dual nature?

These questions have not only practical implications for teaching, but they also have implications for the growing field of explainable AI where adaptive explanations are supposed to be tailored to user interests (Miller 2019; Rohlfing et al. 2021). In explainable AI systems, this translates to interface design choices: Should a recommendation system explain its neural network weights (Architecture), or focus on why it came to a certain loan approval decision based on credit data (Relevance)? A deeper understanding of how human EXs navigate these choices could inform teacher training, curriculum design, and explanation system engineering.

This study presents a controlled investigation of the possible effects such interests have on the course and make-up of an explanation. It uses the dual nature framework as a lens for understanding adaptive explanation: how EXs identify EEs’ interests in Architecture or Relevance, and how those interests influence the course and content of an explanation. We were also interested in whether naturalistic explanations actually address both dimensions as theoretically required, establishing an empirical baseline for understanding how EXs navigate the dual nature dimensions in practice. For our study, we deliberately used a board game as the explanandum. *Quarto!* as a game is a technical artifact with clearly observable features of Architecture (pieces, rules, board) and Relevance (strategies, visual design). Using *Quarto!* allowed us to isolate some core adaptive mechanisms without possible interfering variables present in explanations of more complex technological systems (e.g., uncertainty of EXs or limited transparency of the artifacts).

2 Related work

2.1 The Role of the Dual Nature in Explanations of Artifacts

When explaining technical artifacts, the dual nature theory provides two complementary perspectives useful for deciding what to explain (Kroes 2009; Winkelnkemper et al. 2024). One can explain its *Architecture*—roughly how it is built—or its *Relevance*—roughly what it is for (Kroes 2009; Winkelnkemper et al. 2024). Winkelnkemper et al. (2024) adapted this theoretical account to digital artifacts, thereby demonstrating how the *Architecture-Relevance* distinction applies to computational systems where the Architecture includes explaining algorithms, data structures, and technical implementations, while the Relevance encompasses purpose, user goals, societal impacts, and cultural significance. De Ridder (2007) argues that designers of a technical artifact are usually good at explaining that artifact because they can retrace the design process. In his view, by going through the design process backwards, designers can explain all the design choices they made regarding its Architecture or Relevance, and, importantly, how they relate. In the field of computer science education, students

regularly have to understand technical artifacts such as, for example, program code. There is some evidence that novices often are preoccupied with understanding how the program works (Architecture), while struggling to understand its functions (Relevance) [Soloway \(1986\)](#). A comparative case study indicated that successful program comprehension requires students to navigate between "problem level" and "program level", a distinction that directly corresponds to the *Architecture* and *Relevance* dimensions outlined above [Kallia \(2023\)](#). The results of the study indicated that proficient students systematically employ "Global Backward Explanations". These establish connections between the two levels by linking mechanistic code behaviors (Architecture) to functional problem goals (Relevance). In contrast, struggling students concentrated predominantly on implementation details without constructing the crucial interpretive bridges to Relevance features. This empirical finding supports the theoretical assumption that both dimensions must be integrated for a comprehensive understanding.

Another theoretical work by [Schulte et al. \(2010\)](#) provides a comprehensive framework that directly supports the Architecture-Relevance distinction in program comprehension. Their Block Model organizes program understanding into three dimensions: text surface (the actual code), program execution (how the code runs), and function (the goals and purposes of the program). This framework explicitly separates what the authors call "structure" (encompassing text surface and execution) from "function" thus mirroring the Architecture-Relevance duality. Importantly, their analysis of multiple program comprehension models reveals that expert programmers develop "navigational mental representations" that allow flexible movement between these dimensions, while novices often become trapped in isolated understanding of structural elements (i.e., Architecture) without connecting them to functional purposes (i.e., Relevance). The model reinforces the empirical findings from the study mentioned above ([Kallia 2023](#)) and supports the idea that effective technical artifact comprehension requires not just understanding both perspectives but developing the ability to dynamically navigate between them.

The dual perspectives align with a broader debate in cognitive sciences about mechanistic and functional explanations and understanding. [Lombrozo and Wilkenfeld \(2019\)](#), for example, describe two types of explanatory stances—mechanistic and functional—and explicitly connect them to two modes of understanding, mechanistic and functional understanding. Some empirical evidence suggests that the dual perspectives, understood here as predictive stances or modes of construal, correspond to different cognitive processes or representations ([Lombrozo 2012](#)). These ideas can therefore reveal how people naturally employ certain predictive stances when explaining technical artifacts to address their dual nature.

While research suggests that the distinction between Architecture and Relevance provides a conceptual approach useful for a deeper understanding of how everyday explanations of technical artifacts unfold, it remains unclear what role the dual nature distinction plays on the microlevel of explanations—on how EXs consider and negotiate between these perspectives when adapting to the interests of the EE.

2.2 EXs' Representation of EEs' Interests and Knowledge

If EXs tailor an explanation to specific EEs, their explanation is more successful; in other words, the understanding of the EE is higher (Nückles et al. 2005). Generally, speakers must adapt their communication to each other, their goals, and their knowledge, for example (Brennan et al. 2010; Brown-Schmidt et al. 2015; Clark 1996; Horton and Gerrig 2005), to create a shared common ground (Clark and Brennan 1991). To accomplish this, they must correctly recognize each other's characteristics. For explanations with the goal of understanding (Buschmeier et al. 2023), EXs should especially consider the EEs' interests, prior knowledge, and growing understanding (Kulgemeyer and Geelan 2024). Effective monitoring of their resulting needs requires EXs to continually assess and adapt to the EE's evolving knowledge state. However, as demonstrated by Chi et al. (2004), EXs frequently fail to do so, especially when the EE's underlying understanding or mental models differ substantially from their own. Incorrect monitoring can result in explanations that do not overcome misunderstandings, thus impeding effective knowledge acquisition. Interestingly, research on tutoring has shown that if tutors receive information regarding tutees' prior knowledge, they succeed in adapting their explanation accordingly. Such adaptation reduces the tutees' misunderstanding and supports their knowledge acquisition (Nückles et al. 2005). Thus, the successful adaptation of the explanation is dependent on EXs correctly monitoring the EE.

In a previous qualitative study, a game explanation was observed regarding utterances aimed at Architecture and Relevance. Afterward, the EXs were asked in interviews about the reasons for their focus and shifts from Architecture to Relevance and vice versa. They reported considering own goals as well as their perception of EEs' knowledge and interests regarding Architecture and Relevance (Terfloth et al. 2023). However, so far, it remains an open question whether EXs correctly perceive the interests of the EEs and how they take this into account in their explanations.

2.3 Adaptation of the Explanation Content to the Interests of the EEs

During conversations, people structure and organize the talk (Dale et al. 2014). This adaptation occurs based on many "channels of alignment" (Garrod and Pickering 2004). An example is the alignment of words used in conversation (cf. lexical alignment). In the context of explanation, processes such as monitoring, scaffolding, and the establishment of common ground are particularly interesting (Clark and Brennan 1991). Monitoring can be understood as estimating differences between one's own internal representation of knowledge and one's model of the interlocutor's understanding. Scaffolding can be understood as adapting to these differences to, for example, resolve misunderstandings. Through the iterative application of monitoring and scaffolding over the course of the interaction, interlocutors progressively establish common ground.

In the context of explaining artifacts, our previous study provided the first evidence that EXs argue, when asked after explaining, that they adapted their explanations spontaneously to the perceived interests and knowledge of EEs while still integrating content they themselves consider essential for understanding (Terfloth et al.

2023). Interestingly, EXs typically began with explaining Architecture—structural components, rules, and mechanisms—and argued it was important to provide it as a foundation before diving into more complex aspects of the game (Terfloth et al. 2023). Over the course of the explanation, however, EXs also shifted toward addressing the games’ Relevance—strategic or functional aspects—based on, often, perceived interest of the EEs that they monitored through their multimodal behavior.

For instance, statements like ”these pieces have four properties: they are either tall or small, bright or dark, [...]” or ”they will then be put alternately onto this board” introduce Architecture. In contrast, utterances such as ”what makes this game interesting is that if you place one piece, then you decide which piece your opponent places next” or ”the more pieces are put onto the board, the more you need to think, because there are multiple rows containing three pieces” reflect a turn toward Relevance.

These findings by Terfloth et al. (2023) demonstrate how explanations are not a unilateral knowledge transfer from EX to EE, but joint meaning-making and therefore, in a sense, co-constructed (Rohlfing et al. 2021). EXs initiate explanation with content they judge necessary but adapt the focus and depth in response to the EEs’ evolving understanding. This adaptive interplay between addressing Architecture and Relevance seems beneficial for effective explanations, and thus might have direct implications for education.

2.4 Adaptation of the EXs’ Nonverbal Behavior to the Interests of the EEs

While the previous section examined how EXs adapt their verbal content based on perceived EE interests, adaptation in explanations extends beyond spoken words and includes nonverbal behaviors as well. Regarding the relationship between common ground (Clark and Brennan 1991) and gestures, research indicates that gestures (as well as speech) provide less semantic information and are reduced in size when common ground is already established, as mutually shared knowledge exists (Holler and Wilkin 2009; Gerwing and Bavelas 2004). However, at least in narratives, it was also found that the amount of gestures increases with common ground while the number of words decreases, suggesting that gestures take on other important functions besides information transfer.

Relating to previous work on gestural behavior in naturalistic talk, two main types of gestures are important for game explanations: iconic, or representational, gestures and pragmatic, or non-representational, gestures (McNeill 1992). *Iconic* gestures are hand and arm movements that depict elements of an object or event being referred to in speech, thus bearing a close relation to its semantic content (McNeill 1992). In conversations, they can highlight or emphasize specific aspects of an object or event to bring it to the recipients’ attention. *Pragmatic* gestures are non-representational; their function is to display and highlight ”aspects of the communicative interaction” (Streeck and Mey 2005, p. 73). As such, they can frame the conversational situation, provide cues for interpretation, or mark the delivery of new information (Payrató and Teßendorf 2014). In sum, they display and highlight aspects of the communicative interaction itself (Payrató and Teßendorf 2014; Streeck and Mey 2005).

Considering the different conversational and interactional functions of the two types of gestures, we conducted a previous study that revealed EXs to use iconic gestures in naturalistic explanations predominantly when explaining Architecture (Lohmer et al. 2023). The result indicated that the representational nature of iconic gestures and their closely linked potential to refer to objects or events not available in the concrete speech situation make them valuable resources for explaining physical features of an absent explanandum by depicting them (Kern et al. 2024). A follow-up study revealed in more detail that pragmatic gestures are used in explanations when new semantic content is introduced by the EX, thus structuring the explanation as a communicative interaction and potentially drawing the EEs’ attention to the fact that some new information is about to be delivered (Lohmer and Kern 2024).

In sum, the literature suggests that EXs adapt their gestural behavior to the content of an explanation and that there is a link between the gesture type used and the Architecture and Relevance sides of the technical artifact explained.

3 Research questions and hypotheses

The study addresses the following question about the effect of the dual nature in everyday explanations of technical artifacts: Do EXs adapt their explanations to the EEs’ interests in the Architecture and Relevance dimensions? In an earlier study we acquired first evidence during post-interviews with the EXs (Terfloth et al. 2023). They reported having adapted their explanation, after the onset, to the interest of the EE and how they perceived it. As this result was based on a smaller number of explanations ($N=20$) and on qualitative post-interviews after the explanations and without systematically checking whether EXs actually had done what they had reported, we constructed the present study to have more control. It will provide causal evidence for the impact of EE interests.

In a role-playing experiment, we manipulate the interest of the EE. The EE was controlled to show interest in either only Architecture or only Relevance throughout the entire explanation. The central question examines how interests affect three key areas: (1) the EX’s mental representation of the EEs’ perceived interests and their presumed knowledge regarding Architecture and Relevance, (2) the content of explanations provided, and (3) gesture use (both iconic and pragmatic). We hypothesize:

- **H1 - EXs Mental Representation of EE:** Following the explanation, the EXs represent the interest of the EEs according to the interest shown in Architecture or Relevance.
- **H2.1 - Adaption of Content:** The EXs adapt the content of the explanation to the interests of the EEs.
- **H2.2 - Adaption of Content:** As both Architecture and Relevance are necessary for comprehensive understanding, EXs will address these in their explanations’ content regardless of EE interests.
- **H3 - Adaption of Gestures:** EXs use more iconic gestures in the condition in which the EEs are interested in Architecture than in the condition in which the EEs are interested in Relevance.

In addition, we explored two other variables without specific directional hypotheses. Regarding the EXs’ mental representation of the EEs, apart from EEs’ interest, EXs also took their knowledge into account. It is an open question how the assumed knowledge is influenced by expressed interest. If EEs showed strong interest in Architecture or Relevance, respectively, were they then assumed to have a correspondingly high or low level of knowledge? We therefore also considered the EE’s knowledge as represented by the EXs. Similarly, regarding EXs’ gestures, we studied the use of pragmatic gestures. Given the limited research on the functions of pragmatic gestures, particularly in explanations, we opted for an exploratory approach. The aim was to determine whether pragmatic gestures were used systematically at content-relevant moments—for example, when switching between Architecture and Relevance.

4 Methods

In a controlled experiment, student assistants took the role of the EEs and participated as confederates with manipulated interests. Half of them were instructed to show interest in the Architecture of *Quarto!* only, while the other half showed interest in Relevance of *Quarto!* only. The EXs were recruited participants and tasked to explain the game *Quarto!* they already knew before. The EXs were unaware of the experiment’s nature, unaware that the EEs were part of our team.

The explanations were analyzed qualitatively and quantitatively using a multi-method approach. The study measured three dependent variables: (1) EXs’ mental representations of EE interests and knowledge in Architecture vs. Relevance (assessed via pre-post questionnaires), explanation content (coded as Architecture vs. Relevance), and gestural behavior (coded as iconic and pragmatic gestures). These variables captured how EXs adapt their verbal content, perceptions, and nonverbal behavior when explaining technical artifacts to audiences with different interests.

4.1 Design and sample

A total of $N = 72$ participants recruited via advertisements using flyers, posters, and university newsletters, took part in the study. They booked an experiment appointment via a website. The number was limited by practical constraints of the laboratory setting and the related time reasons. Of these, 43 (59.7%) identified as female and 29 (40.3%) as male. Ages ranged from 18 to 60 years ($M = 23.75$, $SD = 6.40$). The majority of participants (86.1%) reported German as their first language. Most participants were students (95.8%). Half (50%) were enrolled in subjects to become future teachers; others majored in business studies, media studies, and computer science. 30 participants reported working in jobs involving explanatory activities.

Two conditions were implemented in the experiment: EE’s interests in either Architecture or Relevance (A-condition vs. R-condition). The EXs were randomly assigned to one of the two conditions, resulting in $N = 36$ participants per condition.

The explanandum *Quarto!* is a strategic board game involving 16 unique pieces, each with four binary attributes (tall/short, light/dark, square/round, solid/hollow). Players take turns placing pieces on a 4×4 board, with the goal of creating a line of four pieces sharing at least one attribute. Critically, players select which piece

their opponent must place next, creating some strategic complexity. This game is an interesting explanandum because it has a clear Architecture (e.g., pieces, board, rules) while also having important Relevance (strategic elements, tactical considerations, winning strategies, and common pitfalls). It is easy to learn, yet also hard to master.

4.2 Variables and instruments

Naturalness The EXs answered one item about their perception of the naturalness of the setting. "Explaining something in front of a running camera is an unusual situation for many people. This situation can feel more or less natural. We therefore ask you to assess how you experienced the situation". They rated the item *I found the situation ...* on a 7-point scale (1 = not at all natural, 7 = very natural).

EXs' Mental Representations of the EEs' Interests and Knowledge

The EXs answered standardized questionnaires to assess their perceptions of the EEs' interests and knowledge. We developed the questionnaires based on previous qualitative studies (cf. Schaffer et al. 2024) and empirical scale construction. EXs completed the same questionnaire before and after the explanation. The items assessed interest in and knowledge about the Architecture or Relevance of *Quarto!*. The questionnaire used a 7-point Likert scale, ranging from no to full agreement (1 = strongly disagree, 7 = strongly agree).

Table 1 Questionnaire Items Given to EXs for Assessing EEs' Interests in and Knowledge about the Architecture and Relevance of *Quarto!*

Condition	Interests Items	Knowledge Items
Architecture	... is interested in rules.	... knows the game rules.
Relevance	... is interested in the application of complex strategies.	... knows how to apply complex strategies in the game.

Note. Each item began with "My counterpart ..."

Analysis of the Explanation Content

To analyze the content of the explanation, we used qualitative content analysis. The coding manual was developed in a prior study (Terfloeth et al. 2023). All recordings were autotranscribed using Whisper³, afterward student assistants manually checked and adapted all autotranscripts to conform with rules 1, 3, 4, and 14 mentioned in Kuckartz and Rädiker (2019). All utterances were segmented into intonation units (cf. Selting et al. 2009), resulting in one transcript line corresponding to one intonational unit. Intonation units, in short, are prosodic units of spoken language, which helped improve the reliability of our coding in previous studies (Terfloeth et al. 2023, 2025).

We coded the explanation content using qualitative content analysis (Kuckartz 2014) using two code categories: (1) utterances that refer to the Architecture, and (2) utterances that refer to the Relevance of the game *Quarto!*. The coding manual elaborates more clearly when certain categories were annotated and what corner cases

³<https://github.com/openai/whisper>

are⁴. All utterances were also labeled according to the speaker (EX or EE). Two student assistants independently coded 25% of the data using ELAN (Wittenburg et al. 2006)⁵. Intercoder reliability showed almost perfect agreement between coders ($\kappa = .80$) (Landis and Koch 1977). The annotators were unaware of the condition of the experimental data to prevent influencing the coding. The coding manual demonstrated adequate reliability for the study’s objectives. Content analysis revealed that segments coded under both categories aligned with the theoretical framework, enabling reliable identification of explanation content that addresses the architecture or relevance of *Quarto!*. The coding distribution across 30868 words in the transcripts was as follows: 28.6% received no code (e.g., small talk), 40.5% received the code Architecture, 30.6% received the code Relevance and only 0.3% received both codes. Examples for the categories are: ”And these pieces have four properties: they are either tall or small, hollow or solid, dark or light, round or square” (Architecture) and ”so what makes this game interesting is that, if you placed one piece, then you decide which piece your opponent places next” (Relevance).

Analysis of Gesture Behavior/EXs’ and EEs’ Iconic and Pragmatic Gestures

Following the differentiation between iconic (i.e., representational) and pragmatic (i.e., non-representational) gestures, both types were coded using ELAN. Due to the resource-intensive nature of gesture annotation, we conducted an exploratory analysis of 12 random experiments—6 A-conditions and 6 R-conditions—focusing on the gestural behavior of both participants. Only hand movements considered as a ”visible action when it is used as an utterance or as part of an utterance” (Kendon 2004, p. 7) were annotated as gestures. A coding unit included only the stroke phase of a gesture; preparation and retraction phases were excluded. To avoid bias, annotators were unaware of the experimental condition during coding. Both annotators familiarized themselves with the categorization of hand gestures based on the literature (McNeill 2005, 1992) before coding the material.

The coding process was structured as follows: all strokes (i.e., hand movements) were identified and annotated without being assigned to the categories iconic or pragmatic. In a second pass, each stroke was then categorized into the two categories (iconic or pragmatic). This step involved deciding, for each individual annotation and by taking into account the interactive context, whether the gesture was to be classified as iconic, pragmatic, or both.

To assess coding reliability, 10% of the data was independently annotated by two coders experienced in gesture analysis (cf. Kendon 2004; McNeill 2005, 1992). Interrater reliability was almost perfect ($\kappa = .81$) (Landis and Koch 1977). The coding manual thus demonstrated adequate reliability for the study’s objectives and enabled reliable identification of iconic and pragmatic gestures used by EX and EE.

⁴https://osf.io/29pzu/overview?view_only=3d3ee33fb3af409c82d8b744138c5efa

⁵ELAN is an open-source multimedia annotation tool provided by the Language Archive at Max Planck Institute for Psycholinguistics (Nijmegen, the Netherlands) that enables precise temporal alignment of annotations with audio and video recordings.<https://archive.mpi.nl/tla/elan>

4.3 Procedure

The study consisted of three phases. After welcoming the participants (EXs) and thanking them for their participation, they first completed a questionnaire consisting of standardized items about their interest and knowledge of EEs (cf. Table 1). Afterward, they engaged in the explanation with the EE, for which the EE was guided into the room with them. Before they officially began, we started the recording of the cameras. Last, they completed a post questionnaire. After data collection, participants were fully debriefed about the EEs' role as confederates to ensure full transparency of the studies objectives to correspond to common ethical standards. This was done using a standardized information sheet that disclosed the details. Furthermore, participants were asked whether they had any further questions afterward. Participants were reimbursed for their participation monetarily (10€ per hour) or by receiving experiment credits relevant for their studies (only applicable for Psychology students).

EXs were instructed in a standardized manner: "Explain the board game *Quarto!* in such a way that the other person could win". All EE confederates were student assistants of the project. They had the task of expressing interest in Architecture or Relevance. The confederates were already familiar with the game due to previous studies. For their role as confederates, we instructed them equally through a roughly 30-minute workshop that elaborated the dual nature theory in the context of *Quarto!*. Content of the workshop was a 10 minute introduction to the dual nature theory and how it can be used for interpreting artifacts. Then, in a small data-session of around 10 minutes, participants looked through already coded transcripts of a prior study in which two strangers engaged in the explanation of the same explanandum in a natural setting (i.e., no confederate design). In the last 10 minutes, using a small quiz, the student assistants were asked to categorize different utterances into the two code categories (Architecture or Relevance) to test their understanding of the difference in addressing either of the two in *Quarto!* explanations. Their categorization was then discussed in the group and any misunderstandings resolved. In the workshop, they were also made aware that learning certain phrases by heart will not be a viable strategy for the studies, as the explanations should remain natural and unscripted. Therefore, their verbal and non-verbal behavior was *not* scripted but impromptu. We tested the setting and the effectiveness of the workshop over the course of five pilot studies to ensure EEs filled the role properly, the setting was working, and the explanation remained as naturalistic as possible. In the explanation context, whenever the EXs explained the less preferred side of the dual nature, the EEs tried redirecting the explanation, for example, by asking questions that steered the explanation back in the direction of Architecture or Relevance, respectively.

The duration of the explanation was not predetermined or fixed. Instead, the explanations ended when the EX considered the explanation to be completed and the EE had no open questions left. This was done to ensure an explanation ending that naturally occurs as a joint decision. The ending of the explanation was therefore not controlled in any way, which also translates to no time pressure of any sort.

4.4 Data Analysis

We analyzed the proportion of Architecture versus Relevance content in the explanations. To enable comparability across transcripts, we normalized them by standardizing their length. For that, we identified the first and last coded line in each transcript and normalized the time of the two such that it ranged from 0 to 1. Consequently, the first line was always the first utterance that addressed Quarto, while the last line was the last utterance addressing Quarto. Each code was then mapped to its corresponding segment along this normalized timeline (e.g., Architecture was coded from 0 to 0.03). This allowed us to analyze the distribution of coding categories as continuous proportions over the course of each explanation and compare differences across explanations, too. To verify the effectiveness of the manipulation, we additionally analyzed the proportion of Architecture versus Relevance utterances produced by confederates across conditions, too (reported in [subsection 6.3](#)).

5 Results

5.1 Treatment Check

To assess our experimental manipulation, we checked whether confederates (EEs) successfully maintained their prescribed interests. Additionally, we examined whether the conditions differ in terms of naturalness.

Confederates in the Architecture condition produced nearly 80% Architecture-focused utterances compared to only 51% in the Relevance condition—a substantial 29% percentage point difference. Relevance-focused utterances increased from 20% in the Architecture condition to 49% in the Relevance condition, representing a nearly 2.5-fold increase in Relevance content. These differences were substantial and highly reliable: Architecture utterances Welch’s $t(70.0) = 8.92, p < .001$; Relevance utterances Welch’s $t(70.0) = -8.92, p < .001$. Mann–Whitney U tests confirmed these findings (Architecture: $U = 1199.0, p < .001$; Relevance: $U = 97.0, p < .001$).

As shown in Figure 1, EEs in the Architecture condition produced approximately 80% Architecture-focused utterances versus 20% Relevance-focused utterances, while EEs in the Relevance condition produced 51% Architecture-focused versus 49% Relevance-focused utterances. The manipulation remained robust across all dialogue segments when explanations were divided into thirds by timestamp. Significant differences persisted throughout all three phases: first third (Welch’s $t(69.0) = 4.68, p < .001$ for Architecture; $t(69.0) = -4.68, p < .001$ for Relevance), second third (Welch’s $t(70.0) = 6.23, p < .001$ for Architecture; $t(70.0) = -6.23, p < .001$ for Relevance), and third third (Welch’s $t(69.0) = 4.19, p < .001$ for Architecture; Welch’s $t(69.0) = -4.19, p < .001$ for Relevance).

Despite these significant behavioral differences, EXs rated the interactions as similarly natural across conditions. Ratings were high in both the A-condition ($M = 5.17$, $SD = 1.50$, 95% CI [4.66, 5.68]) and the R-condition ($M = 4.89$, $SD = 1.70$, 95% CI [4.31, 5.47]), with no significant difference between conditions ($t(70) = .73$, $p = .466$, $d = 0.17$). This indicates that the manipulation did not create substantially different levels of perceived naturalness in the interactions. There was a slight

duration difference between conditions (Welch's $t(65) = -2.11$, $p = .038$, $d = -0.50$; R-condition: $M = 11.20$ minutes, $SD = 3.59$; A-condition: $M = 9.61$ minutes, $SD = 2.71$, 95% CI of difference $[-3.08, -0.09]$ minutes).

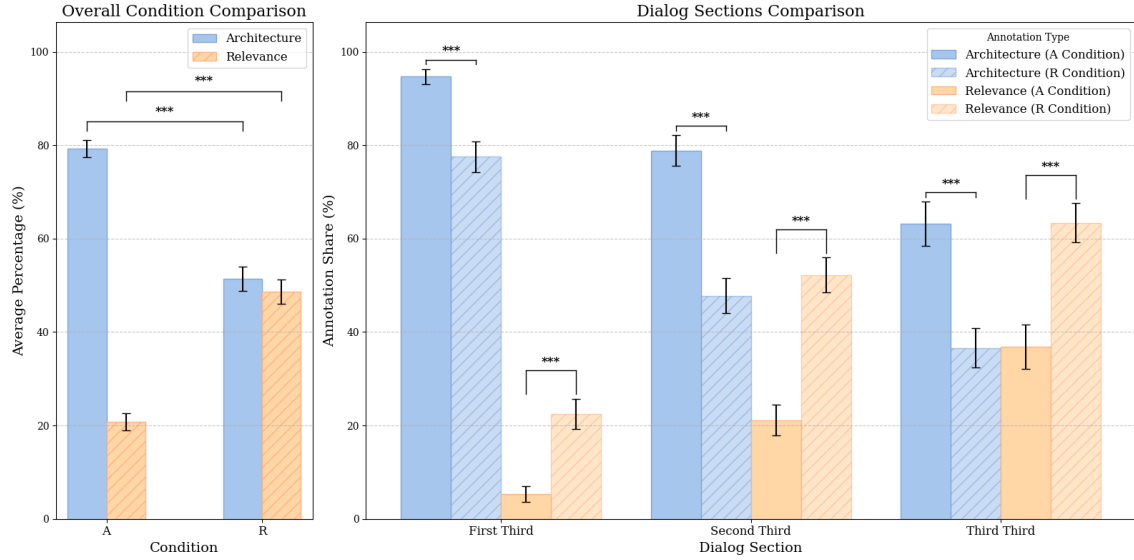


Fig. 1 Distribution of EEs' utterances addressing Architecture vs. Relevance by condition and dialogue section. The manipulation was highly effective, with confederates maintaining distinct interest patterns throughout all phases of the explanation ($p < .001$ for all comparisons).

Hypothesis 1: EXs' Mental Representations

We investigated Hypothesis 1 through a pre-post questionnaire design, revealing partial confirmation that (*H1*) *following the explanation, the EXs represent the interest and knowledge of the EEs according to the shown interest*. EXs successfully recognized EEs' interest in Relevance, but represented EEs' interest in Architecture equally across conditions, suggesting they treat Architecture as a baseline assumption rather than an interest dimension requiring monitoring.

We conducted 2×2 ANOVAs with two measurement points (before vs. after the explanation) and the two conditions (A-condition in which the EE showed interest in Architecture vs. R-condition in which the EE showed interest Relevance). The interaction between measurement point and condition was of particular interest for addressing the research question about whether changes in perceived EE interest and knowledge depend on the explanatory focus.

The results (see Table 2) demonstrate significant differences between the measurement points for all four items. The EXs perceived the knowledge as well as the interest of the EEs in the post-questionnaire as higher than before the explanation. There were almost no effects of the condition, and no significant interaction effects.

Interest in Relevance: Regarding EEs' interests in Relevance (Interests R), a non-significant interaction effect between time and condition emerged ($p < .10$), coupled

Table 2 ANOVA Results (F Values) for the EXs’ Mental Representation of the Interests and Knowledge of the EEs by Condition and Time Point

Item	Condition	Time	Interaction
Interests A	0.63 ($\eta^2 = .009$)	21.12* ($\eta^2 = .232$)	1.64 ($\eta^2 = .023$)
Interests R	1.66 ($\eta^2 = .023$)	10.84* ($\eta^2 = .134$)	0.50 ($\eta^2 = .007$)
Knowledge A	0.17 ($\eta^2 = .002$)	125.84* ($\eta^2 = .643$)	0.07 ($\eta^2 = .001$)
Knowledge R	0.01 ($\eta^2 = .000$)	7.36* ($\eta^2 = .095$)	0.36 ($\eta^2 = .005$)

Note. All $df = (1, 70)$. * $p < .01$. η^2 = partial eta-squared.

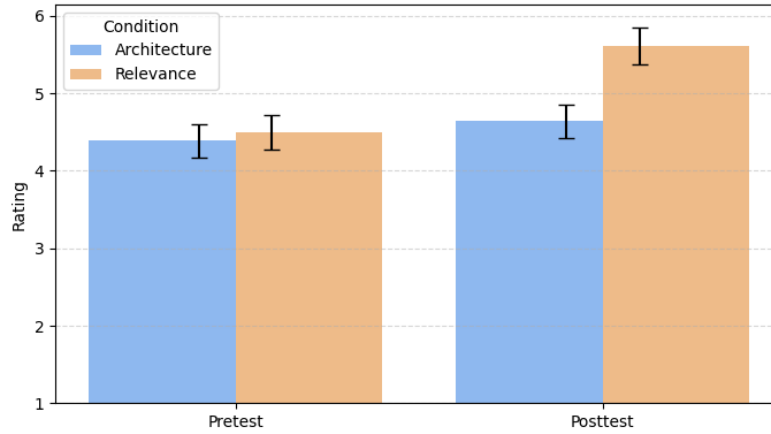


Fig. 2 EXs’ representation of EEs’ interest in Relevance by measurement point and condition.

with a significant condition effect (see Figure 2). Thus, there was a trend for the EXs in the R-condition to rate EEs as having higher interest in Relevance compared to the A-condition.

Interest in Architecture: In contrast, the EXs represented the EEs’ interest in architecture equally in both conditions. No significant condition effects or interactions were found for Architecture interest. This pattern suggests that explainers may assume Architecture interest is universal and/or necessary, whereas they actively monitor for and recognize preferences regarding Relevance. This is coherent with previous findings (Schaffer and Buhl 2025; Terfloth et al. 2023; Schaffer et al. 2024).

Explorative analyses of knowledge representation: No hypothesis was formulated regarding the effect of EE interest on the EX’s assessment of EE’s knowledge. Thus, we conducted explorative analyses. No significant interactions between time and condition were found for either Architecture or Relevance knowledge, indicating that EXs did not differentiate their knowledge assessments based on the manipulated interests. The most pronounced effect emerged in EXs’ representation of EE knowledge about the game’s architecture (Knowledge A), which showed the strongest time effect ($F = 125.84, p < .01$). In contrast, while Relevance knowledge (Knowledge R) also improved significantly over time ($F = 7.36, p < .01$), the effect was considerably more modest (see Figure 3).

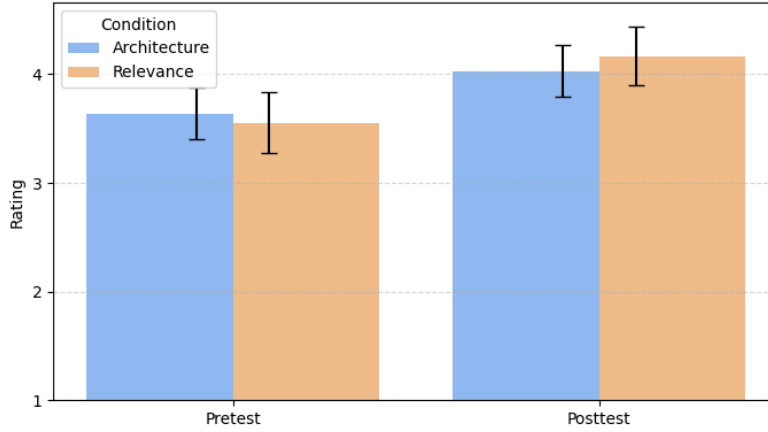


Fig. 3 EXs' representation of EEs' knowledge about Relevance by measurement point and condition.

Hypothesis 2: Adaption of Explanation Content to EE Interests

We examined whether EXs adapted their explanation content to match EE interests through qualitative analysis of the 72 explanations. The results provide clear evidence supporting our hypothesis that explanations adapt to EE interests in terms of content (H2).

EXs demonstrated significant content adaptation while maintaining an Architecture-dominant explanation structure. Consistent with previous findings (Terfloth et al. 2023), EXs produced more Architecture-related content than Relevance content regardless of condition. However, the ratio of Architecture to Relevance content differed significantly between conditions. In the Relevance condition, EXs produced more Relevance-focused utterances compared to the Architecture condition, as shown in Figure 4.

The differences in content adaptation are also significant. When EEs showed interest in Relevance, EXs increased their Relevance content from 28% to 40%, a substantial 12 percentage point shift. Welch's t -tests confirmed these differences were reliable: overall Architecture content differed significantly between conditions ($t(70.0) = 4.34, p < .001$), with identical but opposite patterns for Relevance content ($t(70.0) = -4.34, p < .001$). Mann-Whitney U tests confirmed these findings (Architecture: $U = 999.0, p < .001$; Relevance: $U = 297.0, p < .001$). The adaptation effect remained strong early in explanations (first third: $t(70.0) = 3.54, p < .001$) but weakened somewhat in the second third ($t(70.0) = 3.44, p < .001$) and were only marginally significant in the final third ($t(69.0) = 2.21, p = .031$), suggesting EXs make their most substantial adjustments when first assessing what EEs are eager to learn about.

Despite clear adaptation, we observed substantial Architecture-focused content across both conditions (Architecture condition: 72% Architecture vs. 28% Relevance; Relevance condition: 60% Architecture vs. 40% Relevance).

Hypothesis 3: Gestural Adaptation

In our exploratory analysis of Hypothesis 3, using chi-square tests of independence, we found no indication for the proposition that (H3) EXs use more iconic gestures in

the condition in which the EEs are interested in Architecture than in the condition in which the EEs are interested in Relevance (see Figure 5). In fact, we found the opposite pattern: EXs employed significantly fewer iconic gestures in the A-condition (8.7%) compared to the R-condition (13.7%) ($\chi^2 = 15.72, p < .001$, Cramer's $V = 0.079$).

For EEs, we found a non-significant numerical tendency toward more iconic gestures in the A-condition (5.7%) compared to the R-condition (3.1%) ($\chi^2 = 1.99, p = .158$, Cramer's $V = 0.05$). However, this difference was not statistically significant.

For pragmatic gestures, the distributions remained relatively stable across conditions for both participant groups.

With regard to pragmatic gestures, no differences could be found in our explorative investigation, neither between the EXs and EEs nor between the conditions. These findings should, however, be interpreted as preliminary and require replication with larger samples before firm conclusions can be drawn.

6 Discussion

6.1 Summary of Main Findings

We asked how EXs identify EEs' interests in Architecture or Relevance and how the interests of the EEs influence the explanation. Additionally, we were interested in whether the theoretical imperative to address both dimensions of the artifacts occurs in semi-naturalistic everyday explanations to establish an empirical baseline for understanding how EXs navigate these dual nature tensions in practice. We used the dual nature theory as a theoretical framework for interpreting adaptivity in board game explanations.

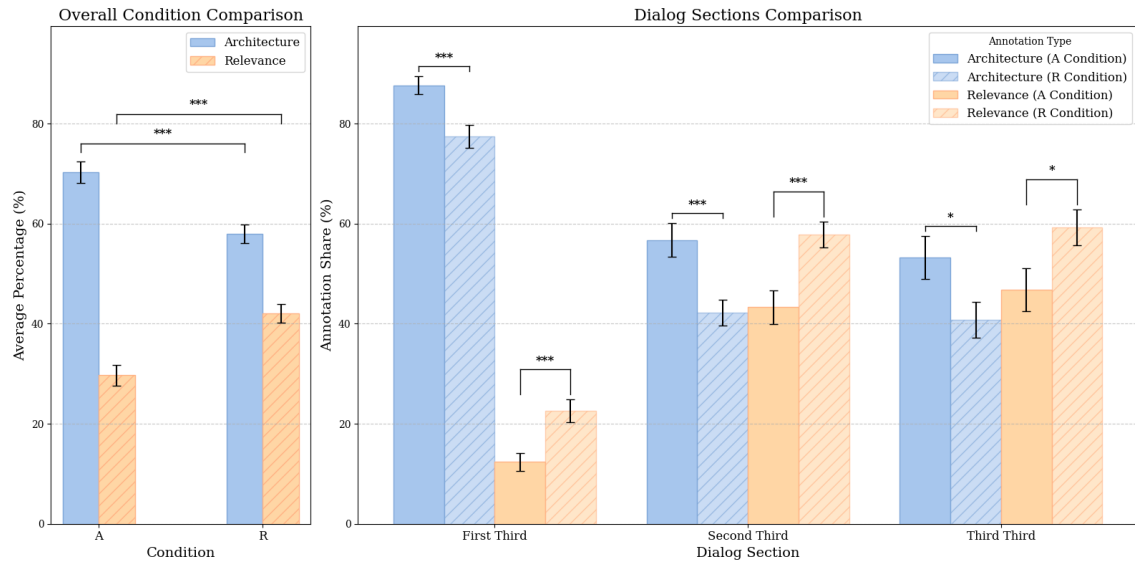


Fig. 4 Distribution of EXs' utterances addressing Architecture versus Relevance by condition and dialogue section. EXs significantly adapted content to match EE interests while maintaining Architecture-dominant explanatory structures.

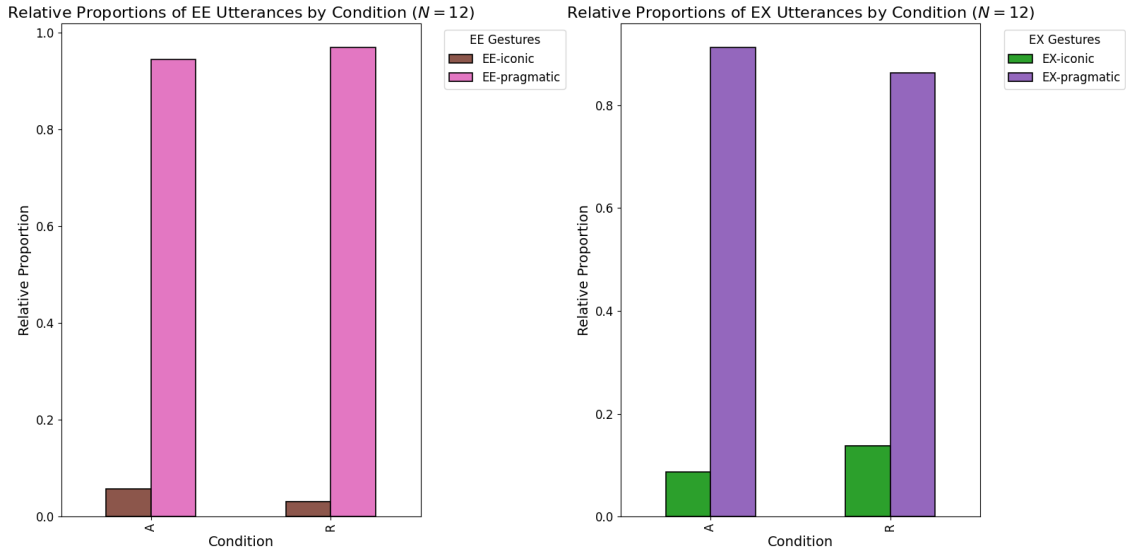


Fig. 5 Relative amount of iconic and pragmatic gestures by conditions shown for each speaker.

The investigation was targeted at the adaptation of the EXs to the interests of the EEs in the context of an everyday explanation of the board game *Quarto!*. We were interested in the effects of EEs’ interests on the EXs’ (1) mental representation of EEs’ interests and knowledge regarding Architecture and Relevance, (2) the content of the explanation, and (3) occurrences of iconic and pragmatic gestures. Understanding these adaptive mechanisms is important because it reveals how (and whether) EXs naturally navigate the theoretical requirement to address both dimensions of technical artifacts while simultaneously responding to EE needs—insights that can inform both educational practice and the design of adaptive explanation systems.

The key findings across the three questions are:

- **Mental representation - H1:** EXs successfully recognized EEs’ interest in Relevance. In contrast, EXs represented EEs’ interest in Architecture equally in both conditions. This underlines our previous assumption (Terfloth et al. 2023) that the interest in Architecture is always expected as the basis of explanations.
- **Content adaptation - H2.1:** EXs adapted explanation content to match EE interests while maintaining an Architecture-dominant structure throughout their explanations.
- **Content adaptation - H2.2:** EXs addressed both sides of the dual nature of *Quarto!* throughout their explanations, regardless of EE interests.
- **Gesture patterns - H3 (exploratory):** Contrary to predictions, EXs used significantly fewer iconic gestures in the Architecture condition compared to the Relevance condition, while pragmatic gestures remained stable across conditions. The underpowered gesture analysis suggests this is an area warranting further investigation with larger sample sizes.

6.2 Limitations

Our sample consisted primarily of students (95%) explaining to confederates, which differs substantially from experienced educators explaining to genuine learners with authentic interests and stakes in understanding, thus limiting the ecological validity. The confederate approach, while enabling experimental control, creates performed rather than authentic interest expression. Real student interests emerge from genuine confusion, personal curiosity, or even career goals, rather than experimental instruction. The board game context may amplify Architecture-dominant explanations due to the physical absence of the game, potentially overestimating the prevalence of Architecture-first strategies in everyday artifact explanation contexts.

The, albeit detailed and time-intensive qualitative, evaluation of gestures based on only 12 interactions and thus bears the risk of random effects. More detailed analyses of the explanation process have to be conducted to provide precise information about how the respective interests in Architecture or Relevance influence the explanation, and how the aspects studied in this paper—EX’ mental representation of EEs’ interest, content regarding Architecture and Relevance, and use of iconic and pragmatic gestures—are intertwined.

6.3 Experimental Validity and Manipulation Success

The experimental design prioritizes establishing basic behavioral patterns through controlled manipulation of EE interests using trained confederates. This approach constrains the realism of the study environment compared to naturalistic everyday explanations, but it enables systematic isolation of the explanation mechanisms that would be difficult to interpret due to confounding factors such as, for example, assessment concerns.

The design successfully established the necessary conditions for investigating explanation adaptation and the influence of EE interests. The confederates maintained their prescribed roles consistently throughout all phases of the explanations, demonstrating that they did not drift from their assigned roles throughout the explanations. This consistency established internal validity for our experimental design, as the confederates successfully created and maintained distinct interests across conditions, providing the necessary foundation to attribute any observed changes of EX adaptation as responses to different EE interests rather than experimental phenomena.

The finding that EXs rated interactions as equally natural across conditions indicates that the manipulation did not create different artificial interactions, supporting the ecological validity of our approach. The observed difference in explanation duration between conditions can be interpreted as evidence that Architecture-focused explanations involve more straightforward, systematic content, whereas relevance-focused explanations require more nuanced, interpretative elaboration. This pattern further supports the manipulation’s effectiveness and provides insight into the cognitive demands of different explanatory approaches.

Importantly, the explanation of the game, irrespective of its focus, resulted in a more favorable evaluation of the EEs by EXs over time, which is to be expected in successful explanatory interactions where mutual understanding develops.

6.4 Theoretical Implications

6.4.1 Architecture as Explanatory Foundation?

Architecture appears to serve as a foundational layer in our data, as it was consistently addressed across conditions, regardless of EE interests. EXs reflected this interest equally in both conditions, suggesting that Architecture may function as a baseline for explanations rather than something actively detected. It might also provide necessary grounding for understanding Relevance (Terfloth et al. 2023, cf.).

Even when confederates were instructed to exclusively show interest in Relevance, confederates produced 75% Architecture-focused utterances, compared to 90% in the Architecture condition. While this 15-percentage-point difference is statistically significant ($p < .001$), the high proportion of Architecture content in the Relevance condition suggests it may be an unavoidable component of technical artifact discourse.

Several mechanisms may explain this. Relevance discussions are linguistically tied to architectural details. For example, when confederates asked questions like "What makes this strategy effective?", EXs grounded their answers in Architecture ("It works because you control which piece your opponent gets, and the pieces have four properties..."). Architecture thus functions as scaffolding for Relevance explanations, making complete separation difficult. Architecture may provide necessary grounding early in explanations (the "laying out the tools" phase, Terfloth et al. (cf. 2023)) and is then supplemented with Relevance elements, consistent with longer interactions in the Relevance condition.

Alternatively, the 75% baseline may reflect a certain explanatory default. Explaining technical artifacts may inherently involve discussing Architecture, even when focusing on purpose or strategies. EXs shifted from 72% Architecture in the Architecture condition to 60% in the Relevance condition. This modest 12-percentage-point change suggests some sensitivity to EE interest, though Architecture remained dominant, indicating an EX motivated Architecture-first strategy.

6.4.2 Adaptation in Explanations

Regarding the EXs' mental representation of the EEs' interests and knowledge, the interests expressed by the EEs (Architecture vs. Relevance) influenced their according mental interest representation by the EX only slightly and did not substantially influence how EXs revised their mental representation of the EEs' knowledge. Consistent with the partial confirmation of H1, EXs recognized EEs' interest in Relevance to a small degree, but did not differentiate their representation of Architecture interest across conditions. However, as we will discuss in the next paragraph, the EXs consider the EEs' interests. Thus, it remains an open question how the adaptation works. We cannot assume that the EX reacts to the interests of the EE without noticing them. On the other hand, the explanations were too complex for a purely linguistic alignment. Further analyses that relate the specific verbal content of EX and EE and the resulting perception by EXs in detailed analyses are therefore needed for understanding the processes (Buhl et al. 2024). Additionally, a detailed view into the interaction, as it is possible with retrospective methods, might be informative (Terfloth et al. 2023). Regarding the EEs' knowledge, EXs may have assumptions about general knowledge

requirements in explanations. In sum, after the explanations, the EX reported a higher knowledge than before the explanation, especially regarding Architecture, while EXs were less certain about their success in conveying Relevance information. However, it remains unclear how the assumed EEs’ knowledge is influenced by the expressed interest. If EEs show strong interest in Architecture or Relevance, respectively, are they then assumed to have a correspondingly high or low level of knowledge? Thus, in conclusion, there was no difference after the explanations.

The results provide evidence supporting our hypothesis that explanations adapt to EE interests in terms of content. Consistent with previous findings (Terflath et al. 2023), EXs produced more Architecture-related content than Relevance content regardless of condition. However, the differences in content adaptation are also practically meaningful, with EXs making their most considerable adjustments when first assessing what EEs are eager to learn about. The results suggest that EXs do not blindly adapt, but follow a personal explanatory strategy by emphasizing Architecture especially in the beginning, but also later in the explanation. When EEs showed interest in either Architecture and Relevance, EXs consistently adapted their explanations accordingly, though adaptation occurred within the constraint of an overall Architecture-dominant explanation content.

EXs actually used significantly fewer iconic gestures when EEs showed interest in Architecture (8.7%) compared to Relevance (13.7%). One possible explanation for this contradictory finding could be located in the study design, where the EEs as confederates were already familiar with the explanandum. Therefore, they may have shown fewer multimodal signals of non-understanding. Because there were fewer signals of non-understanding, highlighting important Architecture aspects through iconic gestures might have been less necessary in the A-condition. Additionally, one could argue that in the R-condition, the perceived interest prompted EX to use more iconic gestures to emphasize important Architecture aspects even more clearly, resulting in the significantly higher amount of iconic gestures. To confirm such a form of compensation — despite the EE’s lack of interest in Architecture, EX still highlighted and underscored the important points very explicitly — a larger empirical foundation and additional methods are required. For pragmatic gestures, the distributions remained relatively stable across conditions for both participant groups. Because pragmatic gestures, like beat gestures or palm-up/palm-down open-hand gestures have mainly a pragmatic function, this result is not surprising.

This interpretation aligns with our previous findings that important Architecture aspects are typically supported by iconic gestures in naturalistic explanations (Lohmer and Kern 2024). The confederate design may have disrupted this natural pattern differently across conditions, with EX compensating through increased gestural emphasis when they perceived EE disinterest in foundational Architecture content.

For EEs themselves, there was a non-significant numerical tendency toward more iconic gestures in the A-condition, though this difference was not statistically reliable. To sum up, we were therefore not only able to disprove (H3), but also found the exact opposite pattern. Further microanalytic analyses are needed to reveal the exact timing, positioning, and function of iconic gestures in explanatory dialogue and their relation to the perceived interests and understanding signals of the interaction partners.

6.4.3 Foundational Implications for Dual Nature Theory

This study provides some empirical evidence that suggests that EXs intuitively attempt to bridge Architecture and Relevance dimensions even when recipients signal clear preferences for just one dimension. The persistence of Architecture-dominant content in the explanations (60-72% across conditions) combined with systematic adaptation to Relevance interests suggests that dual nature bridging might happen as both an automatic tendency and a responsive process during the co-construction of the explanation ([Rohlfing et al. 2021](#)). It can also be interpreted as Architecture being the important matter, to which Relevance can be connected, which would be coherent with the theory.

These findings suggest two important principles: first, Architecture content served as an explanatory foundation regardless of EE interests. Second, the theoretical imperative to address both dimensions naturally manifests in the explanation content regardless of the interest by the EE. This provides empirical insights for dual nature theory's core assumption while revealing first dynamics through which bridging occurs in practice.

6.5 Practical Implications

The persistent inclusion of both Architecture and Relevance content, despite recognized EE preferences, suggests patterns potentially relevant for educational practice. However, these implications must be interpreted cautiously given our experimental context of students explaining a board game to confederates, and should be understood as motivating further investigation rather than directly informing practice.

For teacher training, our findings raise the question of whether explainers more broadly — including professional educators — implicitly prioritize Architecture as foundational knowledge that transcends immediate learner interests. If this pattern extends to trained teachers, developing awareness of this tendency could enable more deliberate pedagogical choices. Educators might benefit from recognizing when their assumption of Architecture importance may not align with learner needs, and developing strategies to make Architecture content more engaging when learners primarily seek Relevance. Research indicates that teachers struggle to connect these dimensions effectively ([Frederik et al. 2011](#)), suggesting this skill may require explicit development.

The finding that explainers maintain Architecture-dominant content despite recognizing Relevance interests raises the question of whether effective artifact education involves ensuring comprehensive coverage rather than simply mirroring learner preferences. These explanatory interactions reveal complex negotiations between what explainers deem foundational and what learners find relevant. Whether creating environments that enable dynamic adaptation between teacher and student perspectives supports learning outcomes remains an open empirical question, though it is consistent with research on high-impact learning activities ([Hattie 2012](#)).

For technology and design education specifically, these patterns raise questions worth investigating in the context of developing technological literacy and capability as part of general education. The dual nature framework maps onto what the field describes as "knowledge in action": understanding how artifacts work (Architecture)

may not be fully separable from understanding what they are for and what they enable (Relevance), and navigating between these dimensions may itself be a component of technological capability. Our finding that explainers spontaneously bridge both dimensions — even when recipients show clear preferences for just one — suggests this integration may reflect a deep explanatory tendency rather than a deliberate pedagogical strategy. Whether making this tendency explicit and reflective in teacher preparation would support more effective technology education warrants empirical investigation. Similarly, it remains an open question whether curricula structured around the navigation between technical understanding and broader purposes, rather than treating these as separate course elements, would better support technological capability — and how assessment designs might reflect such integrated understanding rather than evaluating each dimension in isolation.

Regarding explanation systems and intelligent tutoring, our observed adaptation patterns — partial content adjustment while maintaining Architecture emphasis — may offer tentative inspiration for partner modeling mechanisms (Mousavinasab et al. 2021; Robrecht and Kopp 2023), though direct translation from a board game explanation study requires caution. The patterns observed here are consistent with the idea that effective systems might benefit from recognizing the dual-natured character of technological understanding, rather than tracking only a general knowledge level. What does seem clear is that static assumptions about prior knowledge and interests are insufficient; systems would need to respond to interests emerging during explanation itself (Buhl et al. 2024; Schaffer et al. 2024).

Our results are consistent with challenges to one-sided explanatory frameworks that focus exclusively on mechanistic aspects (de Ridder 2006; Salmon 2020). Even when explainees preferred one dimension, explainers frequently incorporated both. Whether this spontaneous integration enhances comprehension by grounding technical knowledge in meaningful contexts remains an empirical question that this study was not designed to answer.

For assessment, the implications remain speculative. When learners explain artifacts, they may similarly draw on both dimensions regardless of the assessment focus, but this has not been tested. Awareness of these dual perspectives could potentially inform more nuanced evaluation approaches, and understanding how these dimensions interact in learner explanations warrants dedicated investigation.

6.6 Future Directions

This study is a first step towards empirical foundations regarding adaptation of explanations to EE interest in Architecture and Relevance. In that sense it establishes a baseline that requires systematic extension. The controlled experiment conditions provide some empirical foundation, and raise important questions about how these patterns operate under different constraints.

Various directions for future work are sensible. The results of this study provide important information about how EX address the dimensions of Architecture and Relevance adaptively. Thus, based on these insights, creating an experiment with flipped roles, in which the EXs act as confederates to control which side of the dual nature is explained, could be useful. In combination with a post understanding test,

it could provide useful information on whether addressing both sides is always the most effective approach in terms of learning. Another possible direction could be the systematic examination of whether experienced educators show similar bridging tendencies as student EXs in teacher-student conversations. Additionally, it could be useful to explore specific variations of dual nature explanation preferences. To this end, it could make sense to study the explanations of different kinds of explananda — from complex digital artifacts such as fitness trackers to more abstract concepts like shortest path algorithms, which, by their nature, can never be fully represented in physical form. Finally, it remains open whether EXs relied on Architecture-focused explanations because of the experimental setup where the object being explained was not physically present, or because it serves as an important foundation, as suggested by previous studies (Terfloth et al. 2023).

Such progression from controlled experimental studies to the investigation of more natural settings (e.g., during one-on-one tutor sessions) represents a systematic research program that builds empirical understanding of dual nature explanation dynamics. Longer-term, classroom-based studies in technology education settings could examine whether our laboratory findings transfer to authentic pedagogical contexts where technological literacy is the goal. Multi-site observations across different national curricula could reveal whether Architecture-first tendencies persist across diverse approaches to technology education and how teachers navigate between technical detail and broader technological capability when students express divergent interests. Additionally, intervention studies could compare instructional approaches that explicitly foreground dual nature navigation versus separated approaches, measuring students’ integrated technological understanding rather than isolated technical or contextual knowledge.

For intelligent, AI-based explanation systems, our findings suggest partner models that recognize technological understanding as inherently dual-natured rather than simply tracking a general “knowledge level.” Systems could adaptively provide Architecture-focused scaffolding when learners struggle with mechanisms while maintaining connections to purpose, or offer Relevance elaboration when learners grasp implementation but miss broader implications. Such explanation systems must support learners’ developing ability to independently navigate these dimensions for unfamiliar artifacts, not just master some particular technical content. Developing assessment instruments that measure this navigational competency, in other words students’ ability to flexibly shift between how artifacts work (Architecture) and what they are for (Relevance) depending on context, would enable systematic evaluation of both human teaching and system design approaches to supporting technological literacy.

Declarations

Funding

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation): TRR 318/1 2021 – 438445824.

Ethics Statement

This study was approved by the Paderborn University Ethics Board. All subjects participated voluntarily and provided written informed consent before the studies.

Acknowledgements

We thank our colleagues for feedback on the first drafts, as well as all research assistants for their support during coding, analyses, and data acquisition.

Conflict of Interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

- Brennan SE, Galati A, Kuhlen AK. Two Minds, One Dialog: Coordinating Speaking and Understanding. In: Ross BH, editor. *Psychology of Learning and Motivation* vol. 53 of *The Psychology of Learning and Motivation: Advances in Research and Theory*. Academic Press; 2010. p. 301–344. <https://www.sciencedirect.com/science/article/pii/S0079742110530081>. <https://www.sciencedirect.com/science/article/pii/S0079742110530081>.
- Brown-Schmidt S, Yoon SO, Ryskin RA. People as Contexts in Conversation. In: Ross BH, editor. *Psychology of Learning and Motivation* vol. 62. Academic Press; 2015. p. 59–99. <https://www.sciencedirect.com/science/article/pii/S0079742114000048>. <https://www.sciencedirect.com/science/article/pii/S0079742114000048>.
- Buhl HM, Fisher JB, Rohlfing K. Changes in Partner Models – Effects of Adaptivity in the Course of Explanations. In: *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 46 Cognitive Science Society; 2024. <https://escholarship.org/uc/item/5bb7p5pt>.
- Buschmeier H, Buhl HM, Kern F, Grimminger A, Beierling H, Fisher J, et al.: Forms of Understanding of XAI-Explanations. arXiv; 2023. <http://arxiv.org/abs/2311.08760>, arXiv:2311.08760 [cs].
- Chi MTH, Siler SA, Jeong H. Can Tutors Monitor Students’ Understanding Accurately? *Cognition and Instruction*. 2004 Sep;22(3):363–387. <https://doi.org/10.1207/s1532690xci2203.4>.
- Clark HH. *Using Language*. Cambridge: Cambridge University Press; 1996.
- Clark HH, Brennan SE. Grounding in communication. In: Resnick LB, Levine JM, Teasley SD, editors. *Shared Cognition: Thinking as Social Practice*. Washington D.C.: APA Books; 1991. p. 127–149. <https://doi.org/10.1037/10096-006>.
- Dale R, Fusaroli R, Duran N, Richardson DC. The Self-Organization of human interaction. *Psychology of Learning and Motivation - Advances in Research and Theory*. 2014;59:43–95. <http://www.scopus.com/inward/record.url?scp=84877981347&partnerID=8YFLogxK>, <https://doi.org/10.1016/B978-0-12-407187-2.00002-2>.

- Frederik I, Sonneveld W, De Vries MJ. Teaching and learning the nature of technical artifacts. *International Journal of Technology and Design Education*. 2011 Aug;21(3):277–290. <http://link.springer.com/10.1007/s10798-010-9119-3>, <https://doi.org/10.1007/s10798-010-9119-3>.
- Garrod S, Pickering MJ. Why is conversation so easy? *Trends in Cognitive Sciences*. 2004 Jan;8(1):8–11. <https://doi.org/10.1016/j.tics.2003.10.016>.
- Gerwing J, Bavelas J. Linguistic Influences on Gesture’s Form. *Gesture*. 2004 Dec;4:157–195. <https://doi.org/10.1075/gest.4.2.04ger>.
- Hattie J. *Visible Learning for Teachers: Maximizing Impact on Learning*. London: Routledge; 2012.
- Holler J, Wilkin K. Communicating common ground: How mutually shared knowledge influences speech and gesture in a narrative task. In: Kita S, editor. *Speech Accompanying-Gesture 1st edition ed*. Psychology Press; 2009. p. 23–45. Num Pages: 23.
- Horton WS, Gerrig RJ. The impact of memory demands on audience design during language production. *Cognition*. 2005 Jun;96(2):127–142. <https://doi.org/10.1016/j.cognition.2004.07.001>, <https://doi.org/https://doi.org/10.1016/j.cognition.2004.07.001>.
- Kallia M. The Search for Meaning: Inferential Strategic Reading Comprehension in Programming. In: *Proceedings of the 2023 ACM Conference on International Computing Education Research - Volume 1*, vol. 1 of ICER ’23 New York, NY, USA: Association for Computing Machinery; 2023. p. 1–14. <https://doi.org/10.1145/3568813.3600135>.
- Kendon A. *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press; 2004. <http://ebooks.cambridge.org/ref/id/CBO9780511807572>.
- Kern F, Boden U, Nemeth A, Koutalidis S, Abramov O, Kopp S, et al. Preschool children’s discourse competence in different genres and how it relates to iconic gestures. *Journal of Child Language*. 2024;51(3):1–25. <https://pub.uni-bielefeld.de/record/2986707>, <https://doi.org/10.1017/S030500092300065X>.
- Kroes P. Technological Explanations. *Phil & Tech*. 1998;3(3):124–134. <https://doi.org/10.5840/techne19983325>.
- Kroes P. Engineering and the dual nature of technical artefacts. *Cambridge journal of economics*. 2009;34(1):51–62. Tex.ids= kroesEngineeringDualNature2010, peterkroesEngineeringDualNature2010, <https://doi.org/https://doi.org/10.1093/cje/bep019>.

- Kuckartz U. Qualitative Text Analysis: A Guide to Methods, Practice and Using Software. 1st edition ed. Thousand Oaks, CA: SAGE; 2014.
- Kuckartz U, Rädiker S. Transcribing Audio and Video Recordings. In: Kuckartz U, Rädiker S, editors. Analyzing Qualitative Data with MAXQDA: Text, Audio, and Video. Cham: Springer International Publishing; 2019. p. 41–49. https://doi.org/10.1007/978-3-030-15671-8_4. https://doi.org/10.1007/978-3-030-15671-8_4.
- Kulgemeyer C, Geelan D. Towards a constructivist view of instructional explanations as a core practice of science teachers. Science Education. 2024;108(4):1034–1050. <https://onlinelibrary.wiley.com/doi/abs/10.1002/sce.21863>, eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/sce.21863>, <https://doi.org/10.1002/sce.21863>.
- Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. Biometrics. 1977;33(1):159–174. <https://doi.org/10.2307/2529310>.
- Lohmer V, Kern PDF. The role of interactive gestures in explanatory interactions. In: International Multimodal Communication Symposiums (MMSYM) Frankfurt am Main, Germany; 2024. p. 157–158. <http://mmsym.org/wp-content/uploads/2024/09/BookOfAbstractsMMSYM2024-3.pdf>.
- Lohmer V, Terfloth L, Kern PDF. Explaining the Technical Artifact Quarto!: How Gestures are used in Everyday Explanations. In: 1st International Multimodal Communication Symposium Barcelona Spain; 2023. p. 133–134. http://mmsym.org/wp-content/uploads/2023/04/Book_of_abstracts_MMSYM.pdf.
- Lombrozo T. Explanation and abductive inference. In: The Oxford handbook of thinking and reasoning Oxford library of psychology. New York: Oxford University Press; 2012. p. 260–276.
- Lombrozo T, Wilkenfeld D. Mechanistic versus Functional Understanding. In: Grimm SR, editor. Varieties of Understanding: New Perspectives from Philosophy, Psychology, and Theology 1 ed. New York: Oxford University Press; 2019. p. 209–230. <https://academic.oup.com/book/35263/chapter/299833282>. <https://academic.oup.com/book/35263/chapter/299833282>.
- McNeill D. Hand and mind: what gestures reveal about thought. Chicago: University of Chicago Press; 1992.
- McNeill D. Gesture and thought. Chicago: University of Chicago press; 2005.
- Miller T. Explanation in artificial intelligence: Insights from the social sciences. Artificial Intelligence. 2019 Feb;267:1–38. <https://doi.org/https://doi.org/10.1016/j.artint.2018.07.007>.

- Mousavinasab E, Zarifsanaiey N, Niakan Kalhori SR, Rakhshan M, Keikha L, Ghazi Saeedi M. Intelligent tutoring systems: A systematic review of characteristics, applications, and evaluation methods. *Interactive Learning Environments*. 2021;29(1):142–163. <https://doi.org/https://doi.org/10.1080/10494820.2018.1558257>.
- Nückles M, Wittwer J, Renkl A. Information About a Layperson’s Knowledge Supports Experts in Giving Effective and Efficient Online Advice to Laypersons. *Journal of Experimental Psychology: Applied*. 2005;11(4):219–236. Place: US, <https://doi.org/10.1037/1076-898X.11.4.219>.
- Payrató L, Teßendorf S. Pragmatic gestures. In: Müller C, Cienki A, Fricke E, Ladewig SH, McNeill D, Bressemer J, editors. *Body–language–communication: An international handbook on multimodality in human interaction vol. 2*. Berlin: De Gruyter Mouton; 2014. p. 1531 – 1539. Publisher: De Gruyter Mouton Berlin & Boston.
- de Ridder GJ. Reconstructing design, explaining artifacts: philosophical reflections on the design and explanation of technical artifacts. No. vol. 4 in Simon Stevin series in the philosophy of technology, Delft; 2007.
- de Ridder J. Mechanistic artefact explanation. *Studies in History and Philosophy of Science Part A*. 2006 Mar;37(1):81–96. <https://doi.org/10.1016/j.shpsa.2005.12.009>.
- Robrecht AS, Kopp S. SNAPE: A Sequential Non-Stationary Decision Process Model for Adaptive Explanation Generation. In: *Proceedings of the 15th International Conference on Agents and Artificial Intelligence (ICAART 2023)* Lissabon, Portugal: SCITEPRESS – Science and Technology Publications; 2023. p. 48–58. <https://www.scitepress.org/Papers/2023/116713/116713.pdf>.
- Rohlfing KJ, Cimiano P, Scharlau I, Matzner T, Buhl HM, Buschmeier H, et al. Explanation as a Social Practice: Toward a Conceptual Framework for the Social Design of AI Systems. *IEEE Transactions on Cognitive and Developmental Systems*. 2021 Sep;13(3):717–728. <https://doi.org/10.1109/TCDS.2020.3044366>.
- Salmon WC. *Scientific Explanation and the Causal Structure of the World*. Princeton University Press; 2020. <https://www.degruyterbrill.com/document/doi/10.1515/9780691221489/html>.
- Schaffer ME, Buhl HM.: Do Explainers Perceive the Explainees’ Specific Interests and Knowledge? *OSF*; 2025. https://doi.org/10.31219/osf.io/tph6d_v1.
- Schaffer ME, Terfloth L, Schulte C, Buhl HM. Perception and Consideration of the Explainees’ Needs for Satisfying Explanations. In: *xAI-2024 Late-breaking Work, Demos and Doctoral Consortium Joint Proceedings* Valletta, Malta; 2024. <https://ceur-ws.org/Vol-3793/paper.3.pdf>.

- Schulte C, Clear T, Taherkhani A, Busjahn T, Paterson JH. An introduction to program comprehension for computer science educators. In: Proceedings of the 2010 ITiCSE working group reports ITiCSE-WGR '10, New York, NY, USA: Association for Computing Machinery; 2010. p. 65–86. <https://doi.org/10.1145/1971681.1971687>.
- Selting M, Auer P, Barth-Weingarten D, Bergmann J, Bergmann P, Birkner K, et al. Gesprächsanalytisches Transkriptionssystem (GAT 2). Gesprächsforschung: Online-Zeitschrift zur verbalen Interaktion. 2009 Jan;10:152–183. <https://hdl.handle.net/10993/4358>.
- Soloway E. Learning to program = learning to construct mechanisms and explanations. Communications of the ACM. 1986 Sep;29(9):850–858. <https://doi.org/10.1145/6592.6594>, <https://doi.org/10.1145/6592.6594>.
- Spohrer JC, Soloway E, Pope E. A Goal/Plan Analysis of Buggy Pascal Programs. Human–Computer Interaction. 1985 Jun;1(2):163–207. http://www.tandfonline.com/doi/abs/10.1207/s15327051hci0102_4, tex.ids: spohrer1985goal, spohrerGoalPlanAnalysis, https://doi.org/10.1207/s15327051hci0102_4.
- Streeck J, Mey J. Pragmatic aspects of gesture. In: Brown K, editor. International encyclopedia of languages and linguistics vol. 5, elsevier ed. Oxford: Elsevier Oxford, UK; 2005. p. 71–76.
- Terfloth L, Lohmer V, Kern F, Schulte C. Transcription in Computing Education Research: A Review and Recommendations. Informatics in Education. 2025 Apr;<https://www.infedu.vu.lt/journal/INFEDU/article/805>, <https://doi.org/10.15388/infedu.2025.09>.
- Terfloth L, Schaffer M, Buhl HM, Schulte C. Adding Why to What? Analyses of an Everyday Explanation. In: Longo L, editor. Explainable Artificial Intelligence vol. 1903. Cham: Springer Nature; 2023. p. 256–279. https://link.springer.com/10.1007/978-3-031-44070-0_13. https://link.springer.com/10.1007/978-3-031-44070-0_13.
- Winkelkemper F, Höper L, Schulte C. ARIadne – An Explanation Model for Digital Artefacts. Informatics in Education. 2024 Jun;23(2):479–505. <https://infedu.vu.lt/journal/INFEDU/article/772>, <https://doi.org/10.15388/infedu.2024.09>.
- Wittenburg P, Brugman H, Russel A, Klassmann A, Sloetjes H. ELAN : a professional framework for multimodality research. Genoa, Italy; 2006. p. 1556–1559. <https://hdl.handle.net/11858/00-001M-0000-0013-1E7E-4>.

3.4 Building Block 4: Influence of Dual Nature Focus on Understanding

Building Block 4

L. Terfloth et al. (Mar. 2026). *Bridging the Dual Nature: How Integrated Explanations Enhance Understanding of Technical Artifacts*. arXiv:2603.24325 [cs]. DOI: [10.48550/arXiv.2603.24325](https://doi.org/10.48550/arXiv.2603.24325) (submitted to *Informatics in Education*, revised version)

3.4.1 Summary

The previous building blocks established that explainers naturally adopt an Architecture-first strategy when explaining Quarto, and that they partially adapt this strategy in response to explainee interests while maintaining an Architecture-dominant baseline. What neither of the studies so far could address was the question of causality: does the particular way an explainer addresses the dual nature of an artifact actually affect how well the explainee understands it? This building block reports a controlled experiment ($N=104$) designed to answer precisely this question. Using trained confederate explainers assigned to one of three conditions — Architecture-focused (A), Relevance-focused (R), or Integrated (AR) — the study examines the theoretically grounded claim that different explanatory approaches affect two dimensions of understanding differently: comprehension (knowing that) and enabledness (knowing how), following Buschmeier et al. (2025). A 34-item questionnaire was developed through an iterative process involving interview-based item generation and two rounds of exploratory factor analysis on independent samples, and was evaluated in the present study to operationalize these dimensions. Confederate fidelity was confirmed through qualitative content analysis of all 106 coded explanations, showing that the three conditions produced distinctly different Architecture-Relevance profiles (A: 71.7% Architecture, 5.9% Relevance; R: 26.0% Architecture, 57.1% Relevance; AR: 44.4% Architecture, 38.7% Relevance). Intercoder reliability for the coding manual had been established in two prior studies using the same coders and comparable designs ($\kappa = .80$).

Contrary to the hypotheses that Architecture-focused and Relevance-focused explanations would produce characteristic but opposing understanding profiles, the two focused conditions produced virtually identical outcomes on all measures. The integrated condition, however, produced significantly higher enabledness scores than the focused conditions combined ($F(1, 102) = 4.83, p = .030, \eta_p^2 = .045$), while producing no advantage on comprehension. The paper further documents a set of theoretically significant secondary findings. Even Architecture-focused explainers

showed a natural drift toward Relevance over the course of the explanation; explainees consistently gravitated toward Architecture-oriented content regardless of condition; and the introduction of any Relevance content into the explanation appeared sufficient to shift explainees toward a more balanced engagement pattern, an asymmetry the paper terms a ratchet-like dynamic. The paper discusses implications for dual nature theory, education, and XAI design. The manuscript has been submitted to Informatics in Education and simultaneously uploaded to arXiv.

3.4.2 Authors' Contributions

Lutz Terfloth conceived and designed the paper, conducted the content analysis of the explanations with the support of student assistants, automated the transcription of the data, performed all statistical analyses, and wrote the manuscript. The understanding instrument was developed in early collaboration with Michael Schaffer, Heike Buhl, and Carsten Schulte; the majority of subsequent development and validation across multiple independent samples was carried out by Lutz Terfloth under the guidance of Heike Buhl. Michael Schaffer contributed questionnaire data, in particular regarding naturalness ratings and sociodemographic items. Vivien Lohmer contributed to intonation phrasing analyses with the support of student assistants and was involved in the organisation of the study. Friederike Kern was involved in the Phase 1 grant proposal, contributed to the conception of the work package, and secured funding for student assistants. Heike Buhl originated the conceptual idea for the study and provided feedback on the manuscript with specific attention to the statistical analyses. Carsten Schulte supervised Lutz Terfloth and contributed to the theoretical framing within XAI and computing education research. Both reviewed and provided feedback on the final draft.

Lutz Terfloth did greater than or equal to 85% of the work.

Acknowledgements

The experimental setting was designed as part of a research proposal by Heike Buhl, Friederike Kern, and Carsten Schulte within the TRR 318 *Constructing Explainability*. Vivien Lohmer applied the intonation-phrase segmentation to the transcripts and co-organized the data collection. Michael Schaffer co-organized the data collection and contributed questionnaire instruments originally developed for related studies and the corresponding data. Friederike Kern contributed to the study design through the research proposal. I thank all student assistants for their support during data acquisition and coding.

3.4.3 Author's Commentary

This paper closes the causal loop that the earlier building blocks left open. Building Block 2 described what explainers naturally do; Building Block 3 showed that this behavior adapts, within certain boundaries, to explainee interests. Neither study, however, could say whether the Architecture-first strategy, or any particular approach to the dual nature, actually makes a difference for how well an explainee understands the artifact. This paper was designed to force that question by removing the natural variation in explainer behavior and replacing it with experimentally controlled conditions in the form of confederate explainers.

The most surprising finding is not the result confirming the hypothesis but the result contradicting it. The Architecture-focused and Relevance-focused conditions produced identical outcomes, despite receiving explanations with radically different content. This null result is arguably as theoretically significant as the positive finding: it suggests that within the constraints of a well-bounded explanandum, the content focus of an explanation is less decisive for comprehension than one might intuitively assume. What does appear to matter is integration — the explicit bridging of Architecture and Relevance — but its effect is specific to enabledness (knowing how) rather than comprehension (knowing that). Put plainly: to comprehend the artifact, either perspective seems to suffice, at least given the rather low complexity of Quarto. To be able to use—or in this case play—it competently, however, both perspectives need to be connected.

This distinction carries direct implications for the Co-Constructive Dual Nature Framework developed in this dissertation. The framework posits that genuine understanding of a technical artifact requires engaging with both Architecture and Relevance and, crucially, with the integration relations between them. The present data provide the first direct experimental support for this claim, while refining it: the integration imperative appears to be specifically relevant for enabledness (knowing how), making it the dimension of understanding that enables agency over the artifact (see discussion on agency in Buschmeier et al., 2025). This is also the dimension most immediately relevant to the dissertation's broader context. Explanations in informatics education and synthetic explanations in XAI systems also frequently need support users not merely in just knowing about a system but in being able to act with and upon it (see Chapter 1).

Two secondary findings are worth highlighting for their theoretical implications. First, even confederate explainers instructed to maintain an Architecture focus showed a natural drift toward Relevance content over time, suggesting that the Architecture-first pattern identified in Building Block 2 is not merely a default that can be maintained indefinitely but is unstable once sufficient foundation has been laid. Second, the ratchet-like asymmetry in explainee behavior, whereby the introduction of Relevance content frequently shifts explainees toward more balanced engagement, regardless of how much Architecture accompanies it, suggests that

the two sides of the dual nature are not symmetrically positioned in explanatory dialogue. Architecture may be the necessary foundation, but Relevance is the dimension that, once introduced, reshapes the co-constructive dynamic of the explanation.

Bridging the Dual Nature: How Integrated Explanations Enhance Understanding of Technical Artifacts

Anonymous Anonymous^{1*}

Corresponding author(s). E-mail(s): anonymous@anonymous.org;

Abstract

Purpose: Understanding a technical artifact requires grasping both its internal structure (Architecture) and its purpose and significance (Relevance), as formalized by the Dual Nature Theory. But *understanding* itself is not unidimensional: it encompasses comprehension (knowing that) and enabledness (knowing how), whose combination prerequisites agency. This study investigates whether the way explainers address Architecture and Relevance affects the forms of understanding.

Methods: In a between-subjects experiment, 104 participants received face-to-face explanations of the board game *Quarto!* from trained confederates in one of three conditions: Architecture-focused (A), Relevance-focused (R), or Integrated (AR). A questionnaire assessed comprehension and enabledness; qualitative content analysis coded explanations for Architecture and Relevance.

Results: Omnibus comparisons yielded no significant differences for comprehension or enabledness. However, the Integrated condition produced significantly higher enabledness than the two focused conditions combined ($F(1, 102) = 4.83$, $p = .030$, $\eta_p^2 = .045$). Notably, the A and R conditions produced equivalent outcomes despite significantly different explanation content.

Conclusion: The findings suggest that comprehension and enabledness respond differently to explanation structure. Factual comprehension appears robust to variation in dual-nature emphasis, whereas enabledness—the capacity to apply comprehension in practice—benefits specifically from explanations that explicitly bridge Architecture and Relevance. For informatics education, this implies that fostering agency with technical artifacts requires more than teaching either how systems work or what they are for; it requires helping learners connect both perspectives.

Keywords: dual nature theory, explanation, understanding, enabledness, agency, informatics education

1 Introduction

Computing curricula routinely teach students *how* algorithms, data structures, and systems work—their *Architecture*—but connecting this structural knowledge to questions of *why* certain designs are chosen, what purposes they serve, and how they can be critically evaluated—their *Relevance*—remains a persistent pedagogical challenge that benefits from deliberate choices (Schulte 2008; Tedre and Apiola 2013). For example, students frequently acquire the ability to trace code or describe system components without developing the inferential comprehension needed to construct the underlying mechanisms while being able to explain their purpose (Soloway 1986; Kallia 2023). A clearer picture of whether bridging this gap between knowing how an artifact is built and what it is for is beneficial for understanding outcomes motivates the present study.

The Dual Nature Theory provides a principled way to describe this distinction. It posits that technical artifacts, be they smartphones, sorting algorithms, or recommendation systems, possess two essential characteristics: their *Architecture* (how they are built) and their *Relevance* (what they are for) (Kroes 2010; Winkelnkemper et al. 2024). Winkelnkemper et al. (2024) formalize these as distinct but related perspectives on the artifact: the perspective on the *Architecture* emphasizes the physical and formal structures (hardware, algorithms, protocols) while *Relevance* emphasizes functions, societal purposes, discourses, and values. This distinction is not merely techno-philosophical, it has concrete consequences for how explanations of technical artifacts are structured and, as this paper argues, for how well those explanations are understood.

The question of how well an artifact is understood requires a second distinction, this time about understanding itself. Following Buschmeier et al. (2025), understanding can be decomposed into comprehension (knowing that) and enabledness (knowing how), with agency emerging only when both are achieved—a learner not only knows what and how an artifact works but can also act with and upon it. In these terms, the pedagogical problem described above can be restated: students often achieve comprehension without enabledness, and thus lack agency. Whether this gap can be addressed through the way explanations integrate *Architecture* and *Relevance* is the central question of this paper.

Explanation in informatics education is often a face-to-face, dialogic activity (lessons, workshops, pair programming), so research on verbal explanation provides a natural starting point. Research on face-to-face explanations¹ has consistently shown that understanding is shaped by interactive explanatory processes, as speakers adapt explanations to learners’ comprehension, monitor comprehension, and support the construction of coherent mental representations (Chi and Wylie 2014; Fox Tree 1999; Clark and Brennan 1991; Graesser et al. 1995). During explanations, the explainer selects, frames, and sequences information, while the explainee interprets it against their prior knowledge and interests, asks questions, and communicates confusion (Rohlfing et al. 2021; Fisher et al. 2023).

Previous work began to characterize this dynamic process empirically by operationalizing the dual nature perspectives as codes for qualitative content coding.

¹Henceforth the term explanations and face-to-face explanations is used interchangeably.

A study of board game explanations (Terfloeth et al. 2023) indicated that explainers naturally draw on both the Architecture and the Relevance perspective when explaining technical artifacts, with an Architecture-first ordering that corresponds with findings from program comprehension research (Pennington 1987). A previous study (REMOVED FOR REVIEW)The source is currently in press; as a citation would reveal authorship, it has been omitted to preserve double-blind review. revealed that explainers adapt their explanations to the explainees’ interests in Architecture or Relevance while maintaining a broadly Architecture-dominant structure. Together, these findings paint a picture of explanation as a flexible activity in which the two perspectives are navigated jointly.

While these studies observed explanation dynamics on a content level, they never experimentally varied the explanatory approach and did not draw causal conclusions about the explanation effects on understanding outcomes. But, the form and depth of understanding that develops may differ depending on whether explainers emphasize Architecture, foreground Relevance, or explicitly bridge between both. If certain patterns of dual nature addressing consistently produce better understanding, such patterns could inform teacher training and curriculum design in informatics education, with possible further implications for the development of social explainable AI (sXAI) systems, where structuring explanations to support not only comprehension but also user agency is an active area of concern (Miller 2019).

This paper makes a primarily conceptual contribution: it argues that understanding of technical artifacts should be analyzed along two distinguishable dimensions—comprehension and enabledness—and that the way an explanation addresses the dual nature of the artifact may affect these dimensions differently. We use a controlled experiment as a tool for sharpening this conceptual claim, systematically varying whether explanations foreground Architecture, foreground Relevance, or explicitly integrate the two. The study was designed as foundational research: the goal is to test theoretically grounded hypotheses under conditions that minimize confounding variables, using a proxy artifact (the board game *Quarto!*) accessible enough to recruit a broad participant pool without requiring domain-specific prior knowledge. Its findings are intended to inform and contrast claims about explanation in informatics education.

2 Related Work

A large background of literature is distributed in a large array of disciplines. In the next paragraphs, these are presented.

2.1 The Dual Nature of Technical Artifacts

When we try to understand a technical artifact, we frequently encounter two distinct but interrelated questions: *how does it work* and *what is it for*? The Dual Nature Theory, developed primarily through a research program at Delft University of Technology, formalizes this intuition by positing that technical artifacts possess a dual nature. They have a physical structure operating according to natural laws and a function rooted in human intentionality (Kroes 2010). Neither aspect alone captures

the artifact fully. While the physical conceptualization may account for how the artifact works, without its function, the object loses its status as a technical artifact. In contrast, purely functional descriptions effectively "black box" the physical structure (Kroes 2010, p. 56). A holistic understanding therefore requires engaging with both perspectives, and their integration and being able to flexibly move between the two (cf. de Ridder 2007). Following Winkelkemper et al. (2024); Budde (2021), who tailored the dual nature framework to digital artifacts, this paper conceptualizes these two aspects as perspectives on the *Architecture* (how the artifact is constructed and works) and on the *Relevance* (what it is intended for, including purposes, values, and societal significance). This terminology deliberately extends narrow notions of function (see, e.g., Kroes 2010), since complex digital artifacts in particular often have societal impact well beyond their originally intended purposes, requiring a richer notion that situates them within broader human practices, goals, and values (Schulte and Budde 2018; Winkelkemper et al. 2024).

2.2 Understanding Technical Artifacts

If understanding a technical artifact requires engaging with both its Architecture and its Relevance, a natural follow-up question is: what does such understanding actually look like? Research in computing education has long recognized that understanding technical artifacts involves more than being able to recite how something works. Soloway (1986), in the context of learning programming, emphasized that novices often learn to construct working programs (mechanisms) while struggling to understand *why* those mechanisms are appropriate. They implement working code without comprehending the goal-plan decompositions (see Spohrer et al. 1985) that expert programmers employ as organizing frameworks for design decisions. More recently, Register and Ko (2020) found that teaching algorithmic machine learning Architecture with personally meaningful data (and therefore expanding the Relevance) led to greater learner agency than explanations with ambiguous data.

This divide has been elaborated in several complementary frameworks. The Block Model (Schulte 2008) organizes program comprehension into three dimensions (text surface, program execution, and functional goals) across multiple levels of granularity. Expert programmers develop flexible mental representations that allow movement between these dimensions and levels, whereas novices tend to become trapped in some areas without the flexibility to, for example, bridge program execution and function (Schulte et al. 2010). A closely related observation appears in Pennington's (1987) work, which differentiates between *program models* (representing Architecture: syntax, control flow, implementation details) and *domain models* (representing Relevance in the form of the problem being solved in a specific context). Pennington found that effective understanding requires not only the construction of both models but their cross-referencing in a way that connects program parts to domain functions: that connects Architecture to Relevance. This distinction is not specific to computing: cognitive research has consistently shown that the two modes of construal, a mechanistic and a functional mode, target different features of the world and afford different inferential advantages (Lombrozo and Wilkenfeld 2019; Keil 2006; Dennett 2009).

At last, these frameworks address what understanding *targets*—Architecture and Relevance as different features of the artifact—but leave open the complementary question of what *form* understanding takes, in other words, whether a person merely knows facts about the artifact or can act competently with it. [Buschmeier et al. \(2025\)](#) address this dimension by differentiating between *comprehension* (knowing that) and *enabledness* (knowing how to use and adapt a technical artifact). The distinction draws on established foundations in cognitive psychology, notably [Anderson’s 1983](#) account of declarative and procedural knowledge, refined in knowledge typologies such as [de Jong and Ferguson-Hessler \(1996\)](#) and reflected in the knowledge dimension of the revised Bloom taxonomy ([Anderson and Krathwohl 2001](#); [Krathwohl 2002](#)). [Buschmeier et al. \(2025\)](#) elaborate these distinctions between declarative “knowing-that” from procedural “knowing-how” for the analysis of explanations, integrating perspectives from computer science, (psycho-)linguistics, psychology, sociology, philosophy, and education. In their account, comprehension and enabledness are not orthogonal but intertwined dimensions whose joint achievement constitutes agency: the capacity to act with and upon a, for example, technical artifact. Each dimension is further graded from shallow to deep, ranging from recognition of concepts or execution of routine actions to grasping relations between concepts and flexibly applying skills to novel contexts. These constructs translate directly into measurable outcomes: comprehension corresponds to knowledge that can be queried through factual items, while enabledness corresponds to performance in applying that knowledge to concrete scenarios.

2.3 Explanatory Strategies and Understanding Outcomes

The dual nature framework provides a foundation for describing *what* learners need to understand about a technical artifact. It does not, however, describe *how* understanding is established in face-to-face explanations. Such explanations can foreground Architecture, foreground Relevance, or address both, and whether these strategic choices produce different understanding outcomes is an empirical question that has not yet been addressed.

Work on explanation effectiveness provides a general context: explanations adapted to learners’ current understanding consistently produce stronger outcomes than generic approaches ([Wittwer et al. 2010](#); [Nückles et al. 2006](#)), reinforcing the view that understanding requires active construction, not passive reception ([Rohlfing et al. 2021](#)). In that sense, explanations are not unilateral knowledge transfers but co-constructive processes in which explainers and explainees jointly build understanding through dialogue ([Rohlfing et al. 2021](#)). This involves continuous monitoring of mutual understanding, scaffolding of emerging comprehension, and the establishment of common ground ([Clark and Brennan 1991](#)). For technical artifacts specifically, this implies that how Architecture and Relevance are addressed is not necessarily predetermined but emerges through the interaction, as explainers, for example, adapt focus and depth in response to the explainee’s evolving interests and understanding. [Terfloth et al. \(2023\)](#) found that explainers naturally draw on both perspectives, with

Architecture dominating the early phases before Relevance content increases progressively. (REMOVED FOR REVIEW)² revealed that explainers detect and adapt to explainees’ interests in Architecture or Relevance, a process interpretable as monitoring the explainee’s orientation toward one perspective and scaffolding accordingly, while maintaining a broadly Architecture-dominant structure. Together, these findings suggest that the dual nature is navigated somewhat systematically, with both content emphasis and interactional dynamics shaping how the explanation unfolds.

2.4 The Bridge Between Architecture and Relevance

The theoretical case for integrating both perspectives goes beyond simply presenting each one. Kroes (2010) argues that neither Architecture nor Relevance descriptions alone capture the design of a technical artifact. Instead, genuine understanding requires grasping how the two relate, even though a “logical gap” between both exists (Kroes 2006). De Ridder (2007) elaborates that explaining technical artifacts is, in a sense, the reverse of the design process: designers bridge from intended function to physical structure, while explainers trace design decisions backward from structure to function. On this account, expert understanding is characterized not only by depth in one dimension but especially by the ability to flexibly move between them.

Empirical evidence supports this claim at several levels. Kallia (2023) found that successful programmers employ “Global Backward Explanations” that systematically link code behaviors (Architecture) to problem goals (Relevance), while struggling students focus on implementation details without constructing the crucial interpretive bridges. Regarding non-digital artifacts, Miyake (1986), studying dyads making sense of sewing machine mechanics, showed that deep understanding emerged through collaborative exploration that iteratively cycled between identifying functions and questioning mechanisms. At the level of text comprehension, Kintsch’s (1991) distinction between text-base representations (surface information processing, analogous to Architecture) and situation models (meaning construction through integration with prior knowledge, analogous to Relevance) similarly suggests that understanding requires building interpretive bridges between levels of representation.

The Architecture-first ordering reported by Terfloth et al. (2023) adds a temporal dimension to this integration question. Explainers consistently began with structural components and mechanisms before shifting toward Relevance, a pattern that converges with Pennington’s (1987) finding that the program model is typically constructed before the domain model. This Architecture-first pattern may appear to conflict with evidence that functional information is cognitively privileged (McCarthy and Keil 2023; Kelemen 1999), but the tension dissolves on closer inspection: an Architecture-first strategy can serve as an epistemic scaffold, ensuring that when Relevance is addressed, sufficient structural comprehension exists to make the integration meaningful. These ordering findings raise a further question that prior work has not addressed: how does the balance between Architecture and Relevance develop over the course of an explanation, and does the temporal content structure matter for what the explainee understands?

²The source is currently in press; as a citation would reveal authorship, it has been omitted to preserve double-blind review.

Together, these lines of evidence suggest that bridging Architecture and Relevance may be crucial for deeper understanding (in the sense of [Buschmeier et al. 2025](#)). However, while the theoretical arguments and observational evidence may be persuasive, direct experimental evidence for an effect of integrated dual-nature explanations on understanding outcomes does, to our knowledge, not exist. It is this gap the present study addresses.

3 Research Questions and Hypotheses

The preceding sections established three building blocks that jointly motivate the present study: the Dual Nature framework, which posits Architecture and Relevance as two perspectives neither of which alone captures the artifact fully; the distinction between comprehension and enabledness as different forms understanding takes ([Buschmeier et al. 2025](#)); and observational work showing that explainers naturally navigate both perspectives with Architecture-dominant openings and progressive shifts toward Relevance (REMOVED FOR REVIEW)³, whose causal consequences for understanding remained untested.

The inferential logic connecting these building blocks is as follows: if the Architecture perspective and the Relevance perspective target genuinely different features of understanding (cf. mechanistic and functional mode [Lombrozo and Wilkenfeld 2019](#)), and if bridging them is necessary for deeper understanding ([de Ridder 2007](#); [Winkelkemper et al. 2024](#)), then explanations that explicitly integrate both perspectives should produce understanding outcomes that neither focused approach achieves alone. The central theoretical question is therefore not merely whether integration helps, but *for which form of understanding*—comprehension, enabledness, or both—integration matters. Figure 1 summarises how this theoretical layer maps onto the hypotheses tested below, and how the hypotheses will in turn be operationalised in the experiment described in Section 4.

This study therefore addresses whether and how the explainer’s approach to addressing the dual nature affects the depth and form of understanding achieved by the explainees:

RQ1: How do different explanatory approaches (Architecture-focused, Relevance-focused, or Integrated) affect the degree and form of understanding achieved by explainees, and which approach best promotes it?

We hypothesize:

- **H1:** Architecture-focused explanations will produce better comprehension but lower enabledness than Relevance-focused explanations.
- **H2:** Relevance-focused explanations will produce better enabledness but lower comprehension than Architecture-focused explanations.
- **H3:** Integrated explanations that explicitly bridge Architecture and Relevance will produce higher levels of both comprehension and enabledness than either focused condition.

³The source is currently in press; as a citation would reveal authorship, it has been omitted to preserve double-blind review.

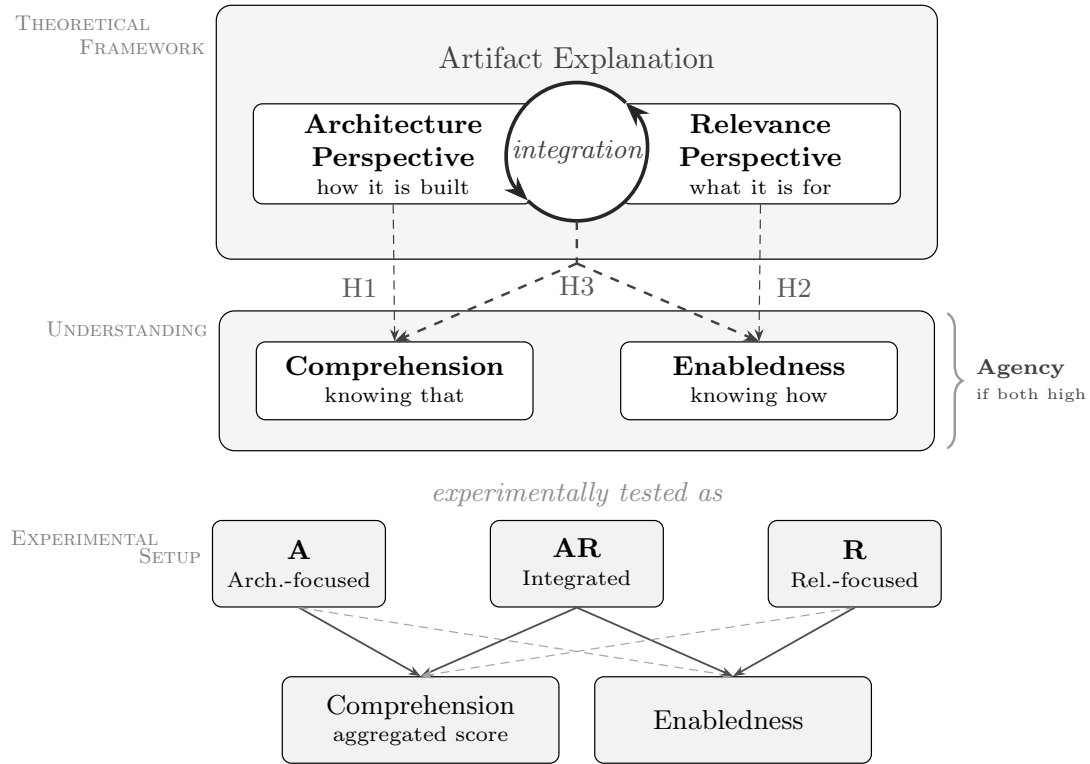


Fig. 1 Inferential structure of the study. The Dual Nature framework posits Architecture and Relevance as two perspectives on technical artifacts (top row). H1 predicts that the Architecture perspective more strongly affords comprehension; H2 predicts that the Relevance perspective more strongly affords enabledness. The central claim (H3) is that only *integration*—explicitly bridging the two modes—fully affords both forms of understanding, and thereby agency (Buschmeier et al. 2025). The experiment (bottom) operationalises these theoretical distinctions by varying explanatory emphasis across three conditions. Predicted outcome patterns are shown as solid (strong prediction) and dashed (weak prediction) arrows. Architecture and Relevance comprehension could not be, as planned, empirically separated and are therefore analysed as their aggregated comprehension score (see Section 4.4).

A note on testability is warranted. H1 and H2 ideally require separable measures of Architecture-specific and Relevance-specific comprehension. As detailed in Section 4.4, the measurement instrument developed for this study did not achieve clean empirical separation of these subdimensions: Architecture and Relevance items loaded on shared latent factors, leading us to analyse only their aggregate comprehension score rather than two distinct subscales as planned. H1 and H2 therefore cannot be tested in their strongest form. We retain them because the predicted contrast between focused conditions—regardless of which subdimension drives it—remains testable at the level of aggregated comprehension and enabledness, and because the theoretical reasoning behind them informs the interpretation of whatever pattern emerges. The primary confirmatory test of this study is H3, which does not depend on subscale separability.

Scope of claims. Because the present study uses a proxy artifact (the board game Quarto!), its claims operate at three distinguishable epistemic levels. *Direct evidence* concerns the experimental effects of explanation conditions on comprehension and enabledness for the specific explanandum. *Conceptual generalisations* concern what these effects imply about the relationship between explanation structure and forms of understanding, grounded in the Dual Nature framework, which applies to technical artifacts broadly. *Transfer hypotheses* concern implications for informatics education, which are theoretically motivated but require empirical testing with complex digital artifacts in authentic educational settings. This distinction structures the interpretation of results in Section 6 and, in particular, the scope we claim for them.

In addition to the understanding-focused hypotheses above, we address the following question:

RQ2 (exploratory): How does the distribution of Architecture and Relevance content develop over the course of the explanation across conditions and speakers?

This question is motivated by evidence that explainers tend to adopt an Architecture-first strategy (Terfloth et al. 2023) and by the theoretical expectation that integration is not merely a matter of overall content balance but of how perspectives are sequenced and connected over time.

4 Method

The hypotheses were tested in a controlled confederate experiment that operationalises the theoretical distinctions introduced above as follows. The *manipulation* varies how the dual nature is addressed, yielding three conditions: Architecture-focused (A), Relevance-focused (R), and Integrated (AR). The *outcome measures* operationalise understanding as comprehension (knowing that) and enabledness (knowing how), following Buschmeier et al. (2025). The *inferential chain* runs from the content manipulation through these two outcome measures to the theoretical claim formulated in H3: whether, and for which form of understanding, integrated explanations outperform focused ones. Figure 1 in Section 3 shows this mapping; the remainder of this section specifies the concrete experimental realisation.

In each session, a confederate explainer trained to deliver condition-specific explanations explained the board game Quarto! to a naive explainee ($N = 104$). Understanding was assessed using a questionnaire developed and evaluated through an iterative process, measuring comprehension (knowing that) and enabledness (knowing how) (Buschmeier et al. 2025). Additionally, the recorded explanatory dialogues were transcribed and coded for Architecture and Relevance content using qualitative content analysis, serving both as a manipulation check and as data for temporal analysis of how dual nature content unfolded across the interaction.

4.1 Participants

4.1.1 Examinee Participants

A total of $N = 104$ examinees (EEs) were recruited from [University].⁴ Participants ranged in age from 18 to 60 years ($M = 23.58$, $SD = 5.05$), with 63.21% identifying as female and 34.91% as male, with 1.89% identifying as diverse. All participants were native German speakers with no prior knowledge of the game Quarto!.

All participants were university students across various degree programs. The largest group studied to become future teachers (Lehramt, $n = 46$), followed by computer science (Informatik, $n = 12$) and business-related programs (Wirtschaftswissenschaften, $n = 7$). Thirty-three percent of participants indicated that their occupation or studies involved explaining ($n = 35$), while 66.98% reported no such focus ($n = 71$).

4.1.2 Confederate Explainers

The EX confederates were five student assistants working in the project very familiar with the explanandum and different explanatory approaches. As preparation, they completed a standardized 30-minute workshop elaborating the dual nature theory in the context of Quarto!: a 10-minute introduction to the dual nature theory and its application to artifacts, a 10-minute data-session reviewing coded transcripts from a prior naturalistic study of explanation, and a 10-minute categorization quiz where confederates classified utterances into Architecture or Relevance categories, with discussion and resolution of misunderstandings.

In training sessions, they tested explaining in the different conditions, too. Confederates did not memorize their explanations, as their verbal and non-verbal behavior had to remain impromptu and unscripted to maintain naturalness. The setting and confederate training effectiveness were qualitatively assured across five pilot studies. All confederates explained in all conditions to control for confederate-specific effects. Confederate fidelity to condition instructions was additionally assessed via content analysis of post-session recordings (see Treatment Check in Section 4.5).

4.2 Materials

4.2.1 Explanandum

We deliberately chose Quarto! because it exemplifies a technical artifact with distinct and observable Architecture (physical components, governing rules, board layout) and Relevance (tactical approaches, aesthetic choices). Its structure parallels that of algorithmic artifacts: the pieces and 4×4 board form a bounded state space, the placement rules are deterministic state transitions, the victory condition acts as a termination

⁴The total set contained 107 EEs. For the content analysis, one EE was excluded because the participant discontinued the experiment, resulting in 106 coded dyads. For the understanding questionnaire, two participants were excluded based on inattentiveness assessed through the instructional manipulation check (IMC, see Section 4.4), and the discontinued participant was likewise absent, yielding the analyzed sample of $N = 104$. Note that the two IMC-excluded participants remain in the content analysis and sociodemographic data, as the IMC assessed questionnaire attentiveness rather than engagement during the explanation phase. Group sizes for the understanding analyses: $n_A = 36$, $n_R = 35$, $n_{AR} = 33$.

predicate, and the strategic choice of which piece to hand to the opponent is a decision procedure under incomplete information about future states. Quarto! thus preserves the Architecture/Relevance structure that motivates the study while minimising confounds present in more complex computing artifacts (explainer uncertainty, partial opacity, domain-specific prior knowledge).

Quarto! features 16 distinct pieces, each defined by four binary attributes (tall/short, light/dark, square/round, solid/hollow). Players alternate placing pieces on a 4×4 grid, aiming to complete a line of four pieces sharing at least one common attribute. A distinctive rule allows players to determine which piece their opponent must place, introducing strategic depth. Despite being straightforward to learn, mastering it proves challenging.

4.2.2 Explanation Conditions

To examine how deliberate foci in explanatory emphasis affect understanding, these experimental conditions deliberately exaggerated the focus on one side of the dual nature to create experimentally separable treatments:

- *Architecture-focused (A)*: Predominantly emphasized game mechanics, rules, piece properties, and structural relationships. Confederate EXs were instructed to allocate the majority of their explanatory effort to Architecture content, while including only minimal-but-essential Relevance content necessary to ensure basic coherence and conversational naturalness.
- *Relevance-focused (R)*: Predominantly emphasized strategies, goals, tactical considerations, and gameplay dynamics. Confederate EXs were instructed to allocate the majority of their explanatory effort to Relevance content, while including only minimal-but-essential Architecture content necessary to ensure basic coherence and conversational naturalness.
- *Integrated (AR)*: Balanced presentation of both Architecture and Relevance with *explicit* bridging statements. Confederate EXs were instructed to address both Architecture and Relevance approximately equally, with deliberate emphasis on how design choices (Architecture) enable or constrain strategic possibilities (Relevance).

4.3 Procedure

Upon arrival, participants (EEs) were welcomed, guided to the explanation room, and asked to read and sign an informed consent form. They then provided payment information for reimbursement (10 € per hour or via study credits applicable to Psychology students) and completed a sociodemographic questionnaire. Before the explanation began, standard recording preparations were completed and camera recordings were initiated.

The EX confederates decided for a condition (based on a plan) and received a standardized instruction: "*Explain the board game Quarto! in such a way that the other person could win*". The EE received the instruction to take part in the explanation actively. The explanation ended whenever the EE had no further questions and both participants considered it complete, ensuring no predetermined duration or time pressure.

The board game Quarto! was *not* present during the explanation. We excluded the physical game for several reasons. We anticipated that having the game present would discourage explainers from explicitly articulating Architecture details (such as describing the piece shapes). By withholding the game, EXs must consciously select which elements warrant explanation, revealing the reasoning strategies they employ, for instance, how they abstract or generalize specific features. On the theoretical premise that gestures reflect underlying cognitive activity (McNeill 1992; Streeck 2009), we also aimed for a dataset suitable for gesture analysis in future work. The table was therefore kept clear of physical objects.

Following the explanation, EEs completed a post-explanation questionnaire including a naturalness rating (1–7), were offered the opportunity to play two rounds of Quarto!, and received a standardized debriefing revealing the confederate design.

4.4 Measures

We operationalized understanding through two forms, following the distinction previously introduced (see Figure 1): *comprehension* (knowing that) as knowledge of both architectural and relevance aspects and their integration, and *enabledness* (knowing how) as the ability to apply that understanding in practice (cf. Buschmeier et al. 2025). Comprehension comprises two sub-dimensions: Architecture comprehension (knowing that regarding game mechanics, rules, components, and structural relationships) and Relevance comprehension (knowing that regarding strategies, goals, and functional significance of design choices). Enabledness captures the ability to apply that comprehension in practical game scenarios. Understanding outcomes were operationalized as three mean scores: a *comprehension score* (mean of correct responses on Architecture and Relevance items combined), an *enabledness score* (mean of correct responses on scenario-based Enabledness items), and a *total score* (mean of both). Mean scores were used rather than sum scores to facilitate interpretability on the original item metric (0–1 scale, representing proportion of correctly answered items).

To measure understanding along these dimensions, we developed a questionnaire through an iterative process involving interview-based item generation, two rounds of exploratory factor analysis on independent samples ($N = 157$, $N = 134$), and a final evaluation. The full process is documented in an OSF repository⁵. The final instrument comprised 34 items across three subscales: 13 Architecture comprehension items, 12 Relevance comprehension items, and 6 Enabledness items, plus 1 Instructional Manipulation Check item assessing inattentiveness and 2 items for participant identification across questionnaires. Architecture and Relevance items use a binary Agree/Disagree format; Enabledness items present visual game scenarios requiring participants to select the correct answer from multiple-choice options (between 3 and 8 options).

Item assignments follow the conceptual structure established during scale development. However, in the present study the Architecture and Relevance subscales did not function as separable dimensions: some Architecture and Relevance items loaded together on shared latent factors in both exploratory factor analyses during development, and confirmatory factor analysis in the present sample ($N = 104$) did not yield admissible solutions, likely due to ceiling effects produced by effective explanations and

⁵https://osf.io/w39dc/overview?view_only=517962d50d314c529b6d0aafe20cf650

condition effects. Internal consistency for the combined comprehension subscale was $\alpha = .65$, and $\alpha = .49$ for the Enabledness subscale. The correlation between Architecture and Relevance comprehension subscores was moderate ($r = .35$, $p < .001$), consistent with the EFA pattern of partial overlap between the two dimensions in this domain. We therefore do not treat Architecture and Relevance comprehension as separable outcomes in the present analyses. Instead, we report a single aggregated comprehension score (Architecture + Relevance items combined) alongside the enabledness score.

Alongside understanding, a post questionnaire assessed naturalness through a single item ("For many people, giving a statement in front of rolling cameras is an unusual situation. This situation can feel more or less natural. We therefore ask you to describe how you experienced the situation.") rated on a 1-7 likert scale.

4.5 Data Analysis

Explanation duration varied considerably across conditions (overall $M = 410.1s$, $SD = 176.5$; 95% CI [376.1, 444.1]): Architecture explanations were shortest ($M = 265.7s$, $SD = 89.9$), followed by Integrated ($M = 454.6s$, $SD = 181.8$) and Relevance ($M = 514.1s$, $SD = 139.8$). A one-way ANOVA confirmed a significant condition effect, $F(2, 103) = 29.81$, $p < .001$; Tukey HSD showed Architecture differed from both Relevance and Integrated (both $p < .001$), while Relevance and Integrated did not differ ($p = .190$). Explanation length is therefore a potential confound for understanding outcomes; we return to this in the Discussion.

4.5.1 Content Analysis

Recordings of the explanatory dialogues were automatically transcribed using Whisper, with manual verification by student assistants following established transcription conventions (Kuckartz and Rädiker 2019, rules 1, 3, 4, 14). All utterances were segmented into intonation units (cf. Selting et al. 2009), prosodic units of spoken language that enhanced coding reliability in previous work. This resulted in each transcript line corresponding to a single intonational unit.

We employed qualitative content analysis to examine explanation content, drawing on a validated coding manual⁶. Two steps structured the coding scheme. First, we classified utterances into two content categories: those addressing the Architecture of Quarto! and those addressing its Relevance. Second, we identified the speaker for each utterance (EXs or EEs). Two student assistants coded the data using ELAN (Wittenburg et al. 2006)⁷. The coding manual's adequacy was confirmed ($\kappa = .81$) in prior work, in settings with the same coders and almost equal design (i.e., one uncontrolled explanation setting, one where EEs were confederates); therefore intercoder reliability was not assessed in the present study. All coders were blind to condition assignment throughout annotation. Across the 115,155 coded segments in the transcripts, the code distribution was: 43.2% Architecture, 36.3% Relevance, 20.3% no code (e.g., conversational filler), and 0.3% received both codes. Representative examples are: "And these

⁶https://osf.io/29pzu/overview?view_only=0b4e0a7546c0435b8b1ded9c8ecd43f7

⁷ELAN is an open-source multimedia annotation tool provided by the Language Archive at Max Planck Institute for Psycholinguistics (Nijmegen, the Netherlands) that enables precise temporal alignment of annotations with audio and video recordings. <https://archive.mpi.nl/tla/elan>

pieces have four properties: they are either tall or small, hollow or solid, dark or light, round or square" (Architecture), "so what makes this game interesting is that, if you placed one piece, then you decide which piece your opponent places next" (Relevance), "hey I am name" (none) and "Quarto! is similar to connect four" (both).

4.5.2 Temporal Analysis of Explanation Content

To compare how explanations distributed content across Architecture and Relevance despite varying lengths, we developed a temporal normalization approach. We identified each transcript's substantive boundaries, specifically, the first utterance mentioning Quarto and the last one, and rescaled these intervals to a 0–1 timeline. This normalization enabled us to represent each coded segment as a proportion of the explanation's overall temporal span (e.g., an Architecture code occurring early might occupy the 0–0.03 range). By treating coding categories as continuous distributions across this normalized time axis, we could characterize how explanations unfolded and directly compare patterning across different speakers and experimental conditions. We further validated whether the experimental conditions successfully shaped the balance of Architecture and Relevance content by examining the proportion of each category produced by confederates in each condition (see Section 5). To capture phase-dependent dynamics, we divided each normalized timeline into three equal segments (first, second, and third third), representing the opening, middle, and closing phases of the explanation. This division is coarse enough to yield robust per-phase estimates even for shorter dialogues, while fine enough to capture the opening-to-closing progression. For each dyad, we computed the proportion of Architecture and Relevance content within each third, separately for EX and EE utterances and for both.

4.5.3 Understanding Outcomes

We first tested for differences across the three conditions (A, R, AR) using one-way ANOVAs and non-parametric Mann–Whitney U tests reported alongside as a robustness check. Because H3 specifically predicts that integrated explanations outperform single-focus explanations regardless of which perspective is emphasised, we additionally compared the AR group against the A and R conditions collapsed into a single "focused" group. This contrast is orthogonal to the omnibus test and follows directly from the theoretical claim that integration, rather than any particular content emphasis, drives enabledness (see Section 2.4). We note that the decision to foreground this contrast in the reporting was reinforced by the observed equivalence of A and R. Effect sizes are reported as partial eta-squared (η_p^2) for ANOVAs and as Hedges' g and rank-biserial correlations for pairwise comparisons.

4.5.4 Statistical Approach

The primary understanding outcomes are derived from binary and multiple-choice items. Because such scores can violate distributional assumptions of parametric tests, we adopt the following strategy throughout the understanding analyses: parametric tests (one-way ANOVAs, Welch's t -tests) serve as the primary analyses, supplemented by non-parametric equivalents (Kruskal–Wallis H , Mann–Whitney U) as robustness

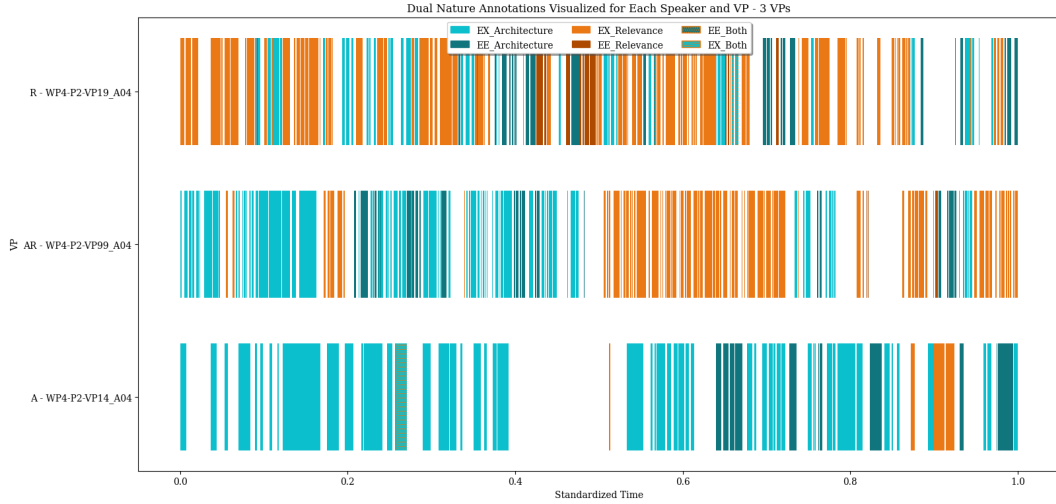


Fig. 2 Illustrative examples of dual nature annotation timelines for one representative dyad per condition. Each bar represents the normalized dialog timeline (0–1), with segments colored by coding category: Architecture (blue), Relevance (orange), and uncoded content (white), with different shading per speaker. Dyads were selected as the closest to their condition mean in Euclidean distance on Architecture and Relevance proportions (A: arch = 92.4%, rel = 7.6%; R: arch = 36.5%, rel = 63.5%; AR: arch = 55.5%, rel = 44.5%).

checks. Where parametric and non-parametric results converge, we report only the parametric statistics and note convergence; where they diverge, we report both and interpret conservatively. For the treatment check, which relies on content proportions rather than participant-level outcome scores, parametric and non-parametric tests consistently converged (all $ps < .001$); we therefore report only parametric results there. Dialog duration, and naturalness ratings are analyzed with parametric tests only.

5 Results

5.1 Confederate Treatment Check

Content analysis of coded EX confederate utterances in the explanation transcripts confirmed successful manipulation across conditions ($N = 95,590$ coded EX utterances across 106 dyads). All pairwise comparisons between conditions were highly significant for both Architecture and Relevance proportions (all $ps < .001$ by Welch’s t -tests and Mann–Whitney U tests; see Table 1).

Participants rated the interactions as similarly natural across all three conditions. Ratings were high in the A-condition ($M = 5.06$, $SD = 1.74$, 95% CI [4.49, 5.62]), the AR-condition ($M = 5.40$, $SD = 1.42$, 95% CI [4.93, 5.87]),

Content	Condition		
	A	R	AR
Architecture	71.7%	26.0%	44.4%
Relevance	5.9%	57.1%	38.7%
Both	0.2%	0.3%	0.2%
None (uncoded)	22.1%	16.6%	16.7%

Table 1 Observed proportions of Architecture and Relevance content by condition, based on coded EX utterances. “None” denotes utterances classified as neither category (e.g., backchannels, fillers, smalltalk).

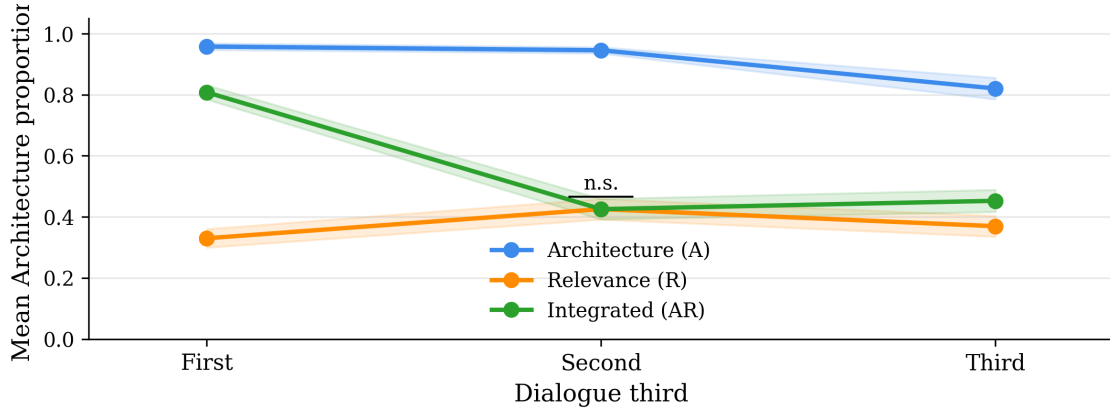


Fig. 3 Mean Architecture proportion across dialogue thirds by condition, averaged over both EX and EE utterances ($\pm$ SEM). The A and AR conditions open with Architecture-dominant content, while the R condition is Relevance-oriented from the outset ($M_{\text{arch}} = .33$ in the first third). The AR condition shifts toward more Relevance-oriented content across the dialogue, converging with R in the second third ($p = .998$) and remaining statistically indistinguishable from R in the third ($p = .100$). The A condition remains Architecture-dominant throughout, declining gradually from $M = .96$ to $M = .82$.

and the R-condition ($M = 4.97$, $SD = 1.52$, 95% CI [4.47, 5.48]). A one-way ANOVA revealed no significant differences in naturalness ratings across conditions ($F(2, 103) = 0.74$, $p = .482$). Pairwise comparisons confirmed no significant differences (all $p > .22$, all $|d| < .30$).

To verify that the experimental conditions were maintained consistently throughout the explanations, we examined how EX Architecture proportions varied across conditions and over the course of the dialogue (divided into thirds). A mixed ANOVA confirmed that the conditions differed strongly in their Architecture content ($\eta_p^2 = .789$, a large effect), and that this difference shifted systematically across dialogue thirds ($\eta_p^2 = .333$), meaning that the three conditions did not simply differ in overall amount but also in how their content emphasis developed over time.

5.2 Temporal Patterns in Explanation Content

Having established that the experimental conditions produced distinct content profiles (Section 5.1), we turn to RQ2: how did the balance between Architecture and Relevance develop over the course of the explanation? We first report patterns across both speakers’ utterances before examining EX and EE contributions separately. Analysis of how content was distributed across dialogue thirds revealed condition-specific temporal patterns (see Table 2 and Figure 3). Overall, all three conditions differed significantly from one another (A vs. AR: $t(69) = 16.22$, $p < .001$; A vs. R: $t(69) = 18.44$, $p < .001$; AR vs. R: $t(68) = 5.24$, $p < .001$). Within this overall difference, however, a temporal pattern emerged: Architecture content was dominant in the opening phase for the A and AR conditions, while the R condition was Relevance-oriented from the outset ($M_{\text{arch}} = .33$ in the first third). AR and R converged in the second third ($p = .998$) and remained statistically indistinguishable in the third ($p = .100$).

Table 2 Pairwise differences in EX Architecture proportions by dialogue third (Welch’s t -tests).

Third	Comparison	t	p
First	A vs. AR	6.30	< .001
First	A vs. R	21.78	< .001
First	AR vs. R	13.25	< .001
Second	A vs. AR	15.13	< .001
Second	A vs. R	15.82	< .001
Second	AR vs. R	1.22	.23
Third	A vs. AR	7.66	< .001
Third	A vs. R	10.64	< .001
Third	AR vs. R	2.80	< .01

At the confederate EX level, the first third showed the expected ordering: A highest ($M = .96$), followed by AR ($M = .81$) and R ($M = .31$), with all pairwise comparisons significant (all $ps < .001$; see Table 2). By the second third, AR and R EXs converged ($t = 1.22$, $p = .23$): AR EXs shifted toward Relevance content, producing a profile statistically indistinguishable from R EXs. In the third third, a modest divergence re-emerged ($t = 2.80$, $p = .007$). The A condition remained Architecture-dominant throughout, though with a gradual decline from .96 to .82, suggesting that even under explicit instructions to maintain an Architecture focus, some drift toward Relevance occurred.

EE content patterns complemented the EX findings but are reported more briefly here. A corresponding ANOVA on EE Architecture proportions revealed a significant Condition effect ($\eta_p^2 = .254$). The central pattern was that A-condition EEs contributed distinctly more Architecture content than either AR-condition EEs ($t(35.81) = 5.38$, $p < .001$, $g = 1.51$) or R-condition EEs ($t(36.60) = 4.64$, $p < .001$, $g = 1.24$), while AR and R EEs were statistically indistinguishable ($t(47.80) = -0.26$, $p = .793$, $g = -0.07$). This divergence was not present in the first third ($p = .433$), but grew progressively across the second and third thirds (both $p < .001$). This pattern suggests that the effect of Relevance content accumulated over time, rather than causing an immediate shift in the other conditions.

Having established the content structure of the explanations, both overall and across the three phases, we now turn to the central question: did the different explanatory approaches lead to different outcomes in understanding?

5.3 Understanding Outcomes

5.3.1 Three-Condition Comparison

Shapiro–Wilk tests indicated departures from normality for enabledness in all conditions and comprehension in A and AR (all $p \leq .002$), consistent with ceiling effects and motivating the dual-testing strategy described in Section 4.5.4.

One-way ANOVAs comparing the three conditions (A, R, AR) revealed no significant differences for comprehension, $F(2, 101) = 0.70$, $p = .500$, $\eta_p^2 = .014$

($N = 104$; two participants excluded based on IMC, see Section 4.4), or for enabledness, $F(2, 101) = 2.50$, $p = .087$, $\eta_p^2 = .047$. Non-parametric Kruskal–Wallis tests converged: comprehension, $H(2) = 1.78$, $p = .411$; enabledness, $H(2) = 5.27$, $p = .072$.

Pairwise comparisons confirmed this pattern. For comprehension, no pair differed significantly (all $p > .15$ by Mann–Whitney U ; all $p_{\text{Holm}} > .76$ by Welch’s t). For enabledness, the only noteworthy comparison was AR vs. R, which reached significance in the uncorrected Mann–Whitney test ($U = 759.0$, $p = .020$) but was marginal after Holm correction ($p_{\text{Holm}} = .064$). No other enabledness pair approached significance (A–AR: $U = 479.5$, $p = .156$; A–R: $U = 694.0$, $p = .446$).

5.3.2 Focused vs. Integrated Comparison

To test H3, given that A and R conditions produced equivalent outcomes and that the theoretical framework predicts a contrast between single-focus and integrated explanations, we conducted a comparison collapsing A and R into a single “focused” group (A or R) versus the AR group. Results are summarized in Table 3.

Table 3 Comparison of Focused (A or R) vs. Integrated (AR) Conditions

Measure	$F(1, 102)$	p	η_p^2	U	p_U	Hedges’ g
Comprehension	0.88	.349	.009	1000.0	.228	−0.20
Enabledness	4.83	.030	.045	875.5	.032	−0.46

Note. F and η_p^2 from one-way ANOVA; U and p_U from Mann–Whitney U test; Hedges’ g from Tukey HSD pairwise comparison. Negative Hedges’ g values indicate higher mean scores in the AR group. Both parametric and non-parametric tests converge for comprehension (both non-significant) and for enabledness (both significant at $\alpha = .05$).

Participants in the AR condition achieved significantly higher enabledness scores than participants in the focused conditions. This effect was consistent across parametric and non-parametric tests: $F(1, 102) = 4.83$, $p = .030$, $\eta_p^2 = .045$; $U = 875.5$, $p = .032$; Hedges’ $g = -0.46$. No significant difference was observed for comprehension by either test: $F(1, 102) = 0.88$, $p = .349$; $U = 1000.0$, $p = .228$.

We note that for total score, parametric and non-parametric results diverged: the ANOVA was non-significant, $F(1, 102) = 3.25$, $p = .075$, $\eta_p^2 = .031$, while the Mann–Whitney U test reached significance, $U = 887.5$, $p = .046$. Following our pre-stated strategy of interpreting conservatively where tests diverge, we do not claim a significant total score difference but note that the marginal parametric result ($p = .075$) is consistent with the interpretation that the enabledness signal is diluted when combined with 25 comprehension items that show no condition difference.

5.3.3 Exploratory Analyses

As an exploratory complement to the temporal analysis, we computed the rate of Architecture–Relevance content transitions per dyad (i.e., the proportion of consecutive utterance pairs in which the content category switched) and examined its

correlation with understanding outcomes. The overarching idea was that more frequently bridging the perspectives might be a strategy that generally outperforms other approaches. But no significant associations were observed (all Spearman $r < .17$, all $p > .12$).

As a further exploratory complement, we examined whether the degree of Architecture emphasis in the opening phase of AR explanations predicted understanding outcomes. Within the AR condition ($n = 33$), the EX Architecture proportion in the first dialogue third was not significantly correlated with enabledness ($r_s = .061$, $p = .738$), comprehension ($r_s = -.108$, $p = .550$), or total score ($r_s = -.084$, $p = .642$). This null result complements the switch-rate finding: within integrated explanations, neither transition frequency nor opening emphasis predicted understanding, suggesting that the benefit of integration is not reducible to a particular sequencing strategy. However, the small subsample limits the detectable effect size, and this finding should be treated as hypothesis-generating.

6 Discussion

We interpret the results below at the three epistemic levels distinguished in Section 3. The summary of findings reports the *direct evidence* obtained for Quarto!; the theoretical implications develop *conceptual generalisations* grounded in the Dual Nature framework; and the practical implications formulate *transfer hypotheses* for informatics education that remain to be tested empirically with complex digital artifacts.

6.1 Summary of Main Findings

This study investigated whether the explainer’s approach to addressing the dual nature of technical artifacts—emphasizing Architecture, emphasizing Relevance, or explicitly integrating both—produces different forms of understanding. The results partially support our hypotheses but reveal an unexpected pattern that refines the theoretical picture.

Contrary to H1 and H2, the Architecture-focused and Relevance-focused conditions did not produce the predicted differential profiles. H1 predicted that Architecture-focused explanations would yield better comprehension but limited enabledness, while H2 predicted that Relevance-focused explanations would yield better enabledness but limited comprehension. Instead, the A and R conditions produced virtually identical outcomes on all measures, despite the treatment check confirming vastly different explanation content.

H3 predicted that integrated explanations would produce the highest levels of comprehension and enabledness. This hypothesis was partially supported: the AR condition produced significantly higher enabledness scores than the focused conditions. The predicted advantage on comprehension did not materialize. The integration advantage was thus specific to the enabledness dimension of understanding.

	Predicted		Found			
	Comp.	Enab.	Comp.	Enab.		
A	high	low	—	—	H1	<i>not supported</i>
R	low	high	—	—	H2	<i>not supported</i>
<i>A ≈ R on all measures (p = .786)</i>						
AR	highest	highest	n.s.	AR > focused	H3	<i>partial</i>
<i>p = .030, g = 0.46</i>						

Fig. 4 Predictions versus findings and conceptual inference. The left columns show the predicted outcome patterns (cf. Figure 1); the right columns show what the experiment found. Dashes indicate no differential effect: A and R conditions produced virtually identical outcomes. The key finding (shaded) is that integrated explanations specifically enhanced enabledness relative to both focused conditions, while comprehension remained unaffected. The inference row states the conceptual result: bridging Architecture and Relevance is important for the capacity to act with the artifact (enabledness), not for factual knowledge about it (comprehension).

6.2 Evaluation of Hypotheses

H1 and H2: Not Supported. The most striking finding is the complete equivalence of the A and R conditions. We had expected that each focused condition would produce a characteristic understanding profile: Architecture explanations fostering comprehension, Relevance explanations fostering enabledness. The absence of any differential effect has several possible explanations. First, the comprehension items in the current instrument do not differentiate finely enough between architecture-specific and relevance-specific knowledge, which may have masked condition-specific advantages. Second, within the constrained domain of a board game, even single-focus explanations may convey enough implicit information from the other dimension to produce comparable outcomes, particularly given that even the R condition included 29.7% Architecture content (both speakers). This suggests that some minimal Architecture content may function as a necessary scaffold even in Relevance-focused explanations, consistent with prior work documenting that Architecture serves as a foundation upon which Relevance content is built (Terfloth et al. 2023). Third, the ceiling effects observed in several comprehension items may have compressed variance, making condition differences undetectable.

H3: Partially Supported. The prediction that integrated explanations would produce the highest enabledness was confirmed. The AR condition produced significantly better enabledness than either focused condition. However, the predicted advantage on comprehension was not found. This partial support suggests that integration of Architecture and Relevance is specifically relevant for the enabledness dimension, the capacity to apply knowledge in practical scenarios, rather than for comprehension per se.

6.3 Theoretical Implications

Why Integration Enhances Enabledness but Not Comprehension. The selective effect on enabledness, while not predicted in this specific form, is interpretable within the dual nature framework. Comprehension may be achievable through exposure to either side of the duality. Whether an EX emphasizes Architecture or Relevance, the factual content conveyed may be sufficient for EEs to answer knowledge-based questions. In contrast, enabledness, the capacity to actually use the artifact competently in novel situations, plausibly requires connecting knowledge about Architecture and Relevance. The scenario-based enabledness items required participants to make tactical decisions in concrete game situations, which demands integrating knowledge about *how the game works* (Architecture) with knowledge about *what matters strategically* (Relevance). Only the AR condition provided explicit bridges between these perspectives. This post-hoc interpretation aligns with [Schulte and Budde \(2018\)](#) and [de Ridder \(2007\)](#) who argue that genuine understanding of technical artifacts requires bridging both sides of the duality, while refining this claim: such bridging appears to be specifically necessary for the enabledness dimension rather than for comprehension and is thus a prerequisite for agency.

The Unexpected Equivalence of Focused Conditions. The $A \approx R$ equivalence challenges the intuitive assumption that explanation content directly determines what is learned. Despite radically different content profiles (71.7% vs. 26.0% Architecture; 5.9% vs. 57.1% Relevance), participants achieved virtually identical understanding outcomes. Several factors may explain this. First, Architecture and Relevance may be sufficiently interconnected that EEs spontaneously infer one from the other—and even R-condition explanations still contained 26.0% Architecture content, suggesting some minimal structural scaffolding persists even under Relevance-focused instruction. Second, the comprehension measure aggregates Architecture and Relevance items into a single score (motivated by the EFA-based factor structure), which may mask condition-specific advantages at the subscale level. Third, ceiling effects on several items may have compressed variance, making differences undetectable. A more complex explanandum would likely reduce both the ceiling effects and the cross-inferability between perspectives, potentially revealing the differential profiles that H1 and H2 predicted.

Reconciling Natural and Optimal Patterns. [Terflath et al. \(2023\)](#) found that EXs naturally employ Architecture-first strategies, and the present data confirm that even experimentally controlled explanations tend to open with Architecture-dominant content across all conditions. The present findings do not challenge the utility of this natural pattern for comprehension, as the A condition achieved comprehension on par with all other conditions. However, the data suggest that the Architecture-first strategy, if not supplemented by explicit integration with Relevance, is insufficient for promoting enabledness. To promote agency, understood here as the condition in which both comprehension and enabledness are achieved (cf. [Buschmeier et al. 2025](#)), EXs need to progress beyond Architecture toward explicit integration with Relevance. The exploratory correlation analysis within the AR conditions first third reinforces this interpretation. The degree of Architecture emphasis in the opening phase was unrelated to any understanding outcome ($r_s = .061$ for enabledness, $r_s = -.108$

for comprehension, both $p > .55$; $n = 33$). Combined with the null result for content switch rates, this suggests that the benefit of integrated explanations is not attributable to a particular sequencing strategy—neither an Architecture-first opening nor frequent perspective switching predicted better outcomes. What appears to matter is instead the presence of explicit integration between Architecture and Relevance in a qualitative, meaningful sense, regardless of how the explanation is temporally organized. In the present study, producing integrated explanations required targeted confederate training, including explicit instruction on the dual nature framework and practice connecting Architecture to Relevance. Frederik et al. (2011) provide independent evidence that this capacity cannot be taken for granted: in their study, both experienced and inexperienced teachers had difficulty even differentiating structural properties from functional tasks, let alone systematically connecting them. Whether untrained explainers spontaneously produce such bridging remains an open question, but it likely is not intuitively achieved, implying that without explicit curricular emphasis on connecting Architecture and Relevance, it might not become part of teachers’ explanatory repertoire.

6.4 Exploratory Findings: Temporal Dynamics and Co-Constructive Patterns

The temporal and speaker-level patterns reported under RQ2 were not predicted by the hypotheses and should be interpreted as exploratory. They do, however, offer suggestive evidence for how dual-nature content shapes the co-constructive dynamic between explainer and explainee.

Architecture Drift in the A Condition. The Condition $\times$ Third interaction revealed that even EXs instructed to maintain an Architecture focus showed a gradual decline in Architecture proportion from .96 to .82 across dialogue thirds. This suggests that pure Architecture explanations may be inherently unstable: as enough structural foundation is laid, the conversational context pulls explanations toward strategic and functional content. The AR condition, rather than fighting this natural pull, channels it deliberately through bridging statements that connect Architecture to Relevance. That the AR condition produced the only significant advantage on enabledness is consistent with the interpretation that deliberate integration potentially outperforms both arbitrary perspective alternations and forced single-perspective explanations.

The Asymmetric Influence of Relevance on Explainee Engagement. EEs contributed predominantly Architecture-oriented content across all conditions (49.1% Architecture vs. 17.9% Relevance overall), suggesting that when learners spontaneously engage with a technical artifact, they gravitate toward structural features. However, this overall Architecture dominance conceals a significant condition effect. A-condition EEs contributed distinctly more Architecture content than EEs in either the AR or R conditions (both $p < .001$, $g > 1.2$), while AR and R EEs were statistically indistinguishable ($g = -0.07$). This pattern suggests a ratchet-like dynamic: Architecture content alone produces Architecture-focused EEs, but the introduction of any sustained Relevance content, whether in the R or AR condition, shifts EEs toward a more balanced engagement pattern. The shift was not immediate—A and

AR EEs did not differ in the first third ($p = .433$)—but emerged cumulatively through the second and third thirds (both $p < .001$), indicating a progressive reorientation rather than an instant response to the explanation’s opening.

For dual nature theory, this asymmetry is notable. The two perspectives do not appear to be symmetrically positioned in explanatory dialogue. Architecture content, even when dominant, does not pull EEs who have been exposed to Relevance back toward a purely structural orientation. Relevance, once introduced, reshapes the co-constructive dynamic in a way that Architecture alone does not. This is consistent with the cognitive preference for functional reasoning discussed in the Related Work (cf. [Kelemen 1999](#); [Lombrozo and Wilkenfeld 2019](#)). From a practical standpoint, the finding implies that Relevance-oriented engagement may require active scaffolding from the EX, as it does not emerge spontaneously from Architecture-focused explanation. Achieving integrated understanding during the co-constructive process may therefore depend on the EX actively introducing Relevance perspectives. A more detailed analysis of EE temporal dynamics, including speaker-level interaction patterns across conditions, is planned for future work.

6.5 Practical Implications

The implications below are *transfer hypotheses* in the sense of Section 3: theoretically motivated and consistent with the present evidence, but involving digital artifacts of substantially greater complexity, opacity, and social contestedness than Quarto!

For Informatics Education. The equivalence of the A and R conditions suggests that informatics educators need not worry about choosing the “right” content focus for promoting comprehension, as both perspectives achieve comparable results. However, if the goal is to give students more than comprehension—the ability to use, evaluate, and adapt technical tools (i.e., agency)—then explicitly connecting Architecture and Relevance should be preferred over teaching one at the expense of the other. This aligns with a broader disciplinary observation. [Tedre and Apiola \(2013\)](#) identify three traditions that have shaped computing and its teaching: a theoretical tradition concerned with formal structures, an engineering tradition concerned with producing working implementations, and a scientific tradition concerned with empirical investigation. The theoretical and scientific traditions are both primarily propositional in their epistemic products (know-that, differing in whether it is established by proof or by empirical observation), while the engineering tradition uniquely integrates propositional with procedural knowledge (know-that and know-how), since designing a working system requires both knowing what principles govern it and knowing how to enact them in a specific implementation. Our findings map onto this distinction: focused explanations produce equivalent comprehension regardless of content emphasis, but only integrated explanations produce enabledness. In [Tedre and Apiola](#)’s terms, single-perspective explanations may produce understanding of the kind the theoretical and scientific traditions privilege, whereas integration fosters the engineering tradition’s epistemic stance of connecting what an artefact is for with how it is built. To the extent that mainstream computing education continues to foreground the foundational disciplinary knowledge shared across all three traditions rather than the integrative work of connecting structure to purpose (cf. [Winkelkemper et al.](#)

2024), the kind of integration our data show to be necessary for enabledness may be systematically undertaught.

In concrete informatics-education terms, the distinction applies across topics where how an artifact is built is routinely taught more thoroughly than why its particular design fits its purpose: algorithm instruction, where control flow and complexity analysis tend to dominate over the problem-structural reasons a given algorithm fits a given problem; data structure instruction, where operation mechanics tend to dominate over the access patterns that motivate the choice of structure; and, increasingly, the teaching of AI and machine learning systems, where training mechanics can crowd out the stakeholder-facing consequences of model deployment. Our findings suggest that in each of these areas, focused single-perspective instruction may produce adequate comprehension while failing to produce the enabledness required for students to actually use, evaluate, and critique the artifacts they are learning about. Curricula should therefore incorporate explicit bridging activities—for example, “explanation generation” tasks in which students explain an artifact to a peer, or design-rationale reflections in which students reconstruct the connection between an implementation choice and the problem it was meant to solve.

A further consequence concerns teacher education in informatics. As noted in the preceding section, [Frederik et al. \(2011\)](#) observed that both experienced and inexperienced teachers struggle to differentiate, let alone connect, structural and functional properties of technical artifacts. If integrated explanation does not emerge spontaneously even among trained teachers, it cannot be assumed as a byproduct of subject-matter expertise and would need to be addressed explicitly in teacher-education curricula—as a practised explanatory habit rather than as incidental content.

Beyond the classroom. The same finding has analogues wherever explanations of digital systems need to support decision-making rather than knowledge alone. In explainable AI, for example, [Miller \(2019\)](#) argues that evaluation has focused too narrowly on the researcher’s perspective, neglecting how people actually process and use explanations; our data give this concern concrete shape, since comprehension is largely insensitive to how Architecture and Relevance are balanced while enabledness—the capacity to act with and upon the system—is not (cf. [Sokol and Flach 2020](#)). The central claim—that bridging matters specifically for agency—therefore generalises naturally to contexts where users of digital systems, not only students of them, need to act on what they understand.

6.6 Limitations

Several limitations should be noted when interpreting these findings. Regarding the understanding measure, the questionnaire was developed through a three-phase process involving interview-based item generation and two rounds of exploratory factor analysis on independent samples. One Relevance item (R13) was excluded post-hoc on substantive grounds (both response options were strategically defensible). The instrument showed modest internal consistency in its final application ($\alpha = .49$ for enabledness), expected given the content heterogeneity and ceiling effects introduced by effective face-to-face explanations. Since measurement error attenuates effects, the

significant enabledness result is best read as a conservative estimate. Across both EFA phases, Architecture and Relevance items consistently loaded on shared factors rather than separating along the theoretical distinction, supporting the decision to aggregate them into a single comprehension score—though this aggregation may mask the condition-specific effects that H1 and H2 predicted. Future work with greater item difficulty, feasible with a more complex explanandum, could improve subscale separability.

The total sample ($N = 104$) provides moderate statistical power for detecting medium-sized effects. Two participants who failed the Instructional Manipulation Check (Oppenheimer et al. 2009) were excluded from the understanding analyses. When included, Cronbach’s α inflates from .60 to .80 and distorts normality in the AR group (W dropping from .87 to .69), both consistent with patterned responding. When retained, the enabledness effect is no longer significant ($F(1, 104) = 2.59$, $p = .111$). We report both analyses for transparency.

The controlled experimental setting with confederate EXs maximized internal validity but necessarily reduced ecological validity compared to, for example, natural teaching contexts. Naturalness ratings, which were uniformly high and did not differ across conditions, provide some reassurance that the confederate design did not introduce perceptible artificiality, though we acknowledge that ratings cannot capture all aspects of ecological validity and item formulation did not explicitly point at a lack of naturalness of the explanation itself.

Explanation duration varied significantly across conditions, with Architecture-focused explanations being substantially shorter than the other two conditions. While this reflects a genuine property of the conditions (less content to cover when one perspective is minimized) it means that duration is confounded with condition. That said, the AR and R condition produced comparably long explanations ($p = .190$), without the enabledness benefit in the AR condition, which argues against a pure time-on-task interpretation.

The physical absence of Quarto! during the explanation was a deliberate design choice intended to encourage explicit verbalization and to support future gesture analysis. However, it also means that the explanations unfolded without the support that a physical artifact provides. In naturalistic settings, explainers frequently point to, manipulate, or demonstrate with the artifact itself, which may alter both the structure of the explanation and the understanding it produces. The present findings therefore characterize explanation dynamics in a “black box” scenario and may not generalize directly to contexts where the artifact is available for joint inspection. This potentially influences the amount of Architecture related utterances in the explanation, as the material of the game needs description. With a game that is present on the table, we expect less Architecture utterances and therefore the balance might shift throughout conditions. Future work should assess how and if that would influence the results of the temporal patterns.

Finally, the explanandum itself constrains generalizability. This study was designed as foundational research: the goal was to test theoretically grounded hypotheses about the effect of dual-nature emphasis on understanding under conditions that minimize

confounding variables, using an explanandum accessible enough to recruit a broad participant pool without requiring domain-specific prior knowledge. Quarto! serves this purpose well, exhibiting clear dual-nature characteristics that make it suited for isolating core explanatory mechanisms. However, it is a board game of limited complexity. Its Architecture can be grasped quickly, and its Relevance, though strategically rich, does not approach the contested, multi-stakeholder character of Relevance in complex digital artifacts such as AI systems or algorithmic decision-making tools. Whether the integration advantage on enabledness holds—or becomes even more pronounced—for artifacts where the Architecture is partially opaque and the Relevance is socially contested remains an open and important question.

Future research should examine whether integrated explanation strategies transfer to complex digital artifacts. Replication with larger samples and improved measurement instruments, particularly for the enabledness construct and with separable comprehension subscales, would strengthen confidence in the present findings. Investigating how learners can be trained to generate bridging connections autonomously, rather than relying on explainer-provided bridges, represents a promising direction for educational practice.

7 Conclusion

This study found that the content focus of an explanation—whether predominantly Architecture or Relevance—had no detectable effect on understanding outcomes, while integrated explanations specifically enhanced enabledness. Comprehension was comparable across all conditions. This pattern differed from our original predictions. We had hypothesized that each focused condition (A-focus, R-focus) would produce a characteristic understanding profile (H1, H2) and that integrated explanations would produce the highest levels of both comprehension and enabledness (H3).

But the partial hypothesis support is, on reflection, very informative. The null effect of content focus (disconfirming H1 and H2), given it really can be confirmed with more complex artifacts in future research, suggests that learners are more resilient to variation in explanation content than we assumed. The integrated focus’ advantage on enabledness (partially confirming H3) suggests that what integration provides is not more comprehension but the capacity to use it in action. This is consistent with claims around the dual nature framework’s stating that genuine understanding of technical artifacts requires bridging Architecture and Relevance ([Schulte and Budde 2018](#); [Winkelkemper et al. 2024](#)), while specifying that this bridging is specifically necessary for enabledness rather than for comprehension.

For the broader goals of the research project, if these findings generalize, they provide empirical grounding for the claim that promoting agency, the ability to understand, evaluate, and adapt technical artifacts to one’s needs, requires more than conveying comprehension about either Architecture and Relevance. It requires helping learners build inferential bridges between these perspectives.

Declarations

Funding

This research was funded by REMOVED FOR REVIEW

Ethics Statement

This study was approved by REMOVED FOR REVIEW Ethics Board. All subjects participated voluntarily and provided written informed consent before the studies.

Acknowledgements

We thank our colleagues for feedback on the first drafts, as well as all research assistants for their support during coding, analyses, and data acquisition.

Disclosure

AI-assisted tools were used for checking linguistic correctness (LanguageTool, Claude), and during data analysis in Python (Copilot in VSCode). All recommendations were systematically double-checked to meet a high quality standard.

Conflict of Interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

- Anderson JR. The architecture of cognition. The architecture of cognition, Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc; 1983. Pages: xi, 345.
- Anderson LW, Krathwohl DR. A taxonomy for learning, teaching, and assessing: a revision of bloom's taxonomy of educational objectives. Longman; 2001. <https://books.google.de/books?id=EMQIAQAIAAJ>.
- Budde L. Entwicklung und Rekonstruktion einer interaktionsgeprägten Sichtweise auf das komplementäre Mensch-Maschine-Verhältnis. PhD thesis; 2021.
- Buschmeier H, Buhl HM, Kern F, Grimminger A, Beierling H, Fisher J, et al. Forms of understanding for XAI-Explanations. Cognitive Systems Research. 2025 Dec;94:101419. <https://www.sciencedirect.com/science/article/pii/S1389041725000993>, <https://doi.org/10.1016/j.cogsys.2025.101419>.
- Chi MTH, Wylie R. The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes. Educational Psychologist. 2014 Oct;49(4):219–243. <https://doi.org/10.1080/00461520.2014.965823>, <https://doi.org/10.1080/00461520.2014.965823>.
- Clark HH, Brennan SE. Grounding in communication. In: Perspectives on socially shared cognition. Washington, DC, US: American Psychological Association; 1991. p. 127–149.
- Dennett D. Intentional Systems Theory. In: Beckermann A, McLaughlin BP, Walter S, editors. The Oxford Handbook of Philosophy of Mind. Oxford University Press; 2009. p. 339–350. <https://doi.org/10.1093/oxfordhb/9780199262618.003>.

0020. <https://doi.org/10.1093/oxfordhb/9780199262618.003.0020>.
- Fisher J, Robrecht A, Kopp S, Rohlfing K. Exploring the Semantic Dialogue Patterns of Explanations - a Case Study of Game Explanations; 2023.
- Fox Tree JE. Listening in on monologues and dialogues. *Discourse Processes*. 1999;27(1):35–53. <https://doi.org/10.1080/01638539909545049>.
- Frederik I, Sonneveld W, De Vries MJ. Teaching and learning the nature of technical artifacts. *International Journal of Technology and Design Education*. 2011 Aug;21(3):277–290. <http://link.springer.com/10.1007/s10798-010-9119-3>, <https://doi.org/10.1007/s10798-010-9119-3>.
- Graesser AC, Person NK, Magliano JP. Collaborative dialogue patterns in naturalistic one-to-one tutoring. *Applied Cognitive Psychology*. 1995;9(6):495–522. <https://onlinelibrary.wiley.com/doi/abs/10.1002/acp.2350090604>, <https://doi.org/10.1002/acp.2350090604>.
- de Jong T, Ferguson-Hessler MGM. Types and qualities of knowledge. *Educational Psychologist*. 1996 Apr;31(2):105–113. https://doi.org/10.1207/s15326985ep3102_2, https://doi.org/10.1207/s15326985ep3102_2.
- Kallia M. The Search for Meaning: Inferential Strategic Reading Comprehension in Programming. In: *Proceedings of the 2023 ACM Conference on International Computing Education Research V.1* Chicago IL USA: ACM; 2023. p. 1–14. <https://dl.acm.org/doi/10.1145/3568813.3600135>.
- Keil FC. Explanation and Understanding. *Annual Review of Psychology*. 2006;57(1):227–254. <https://doi.org/10.1146/annurev.psych.57.102904.190100>, <https://doi.org/10.1146/annurev.psych.57.102904.190100>.
- Kelemen D. Function, goals and intention: children’s teleological reasoning about objects. *Trends in Cognitive Sciences*. 1999 Dec;3(12):461–468. <https://linkinghub.elsevier.com/retrieve/pii/S1364661399014023>, [https://doi.org/10.1016/S1364-6613\(99\)01402-3](https://doi.org/10.1016/S1364-6613(99)01402-3).
- Kintsch W. The Role of Knowledge in Discourse Comprehension: A Construction-Integration Model. In: Stelmach GE, Vroon PA, editors. *Advances in Psychology* vol. 79 of Text and Text Processing. North-Holland; 1991. p. 107–153. <https://www.sciencedirect.com/science/article/pii/S0166411508615514>. <https://www.sciencedirect.com/science/article/pii/S0166411508615514>.
- Krathwohl DR. A Revision of Bloom’s Taxonomy: An Overview. *Theory Into Practice*. 2002 Nov;41(4):212–218. https://doi.org/10.1207/s15430421tip4104_2, [_eprint: https://doi.org/10.1207/s15430421tip4104_2](https://doi.org/10.1207/s15430421tip4104_2), https://doi.org/10.1207/s15430421tip4104_2.

- Kroes P. Coherence of structural and functional descriptions of technical artefacts. *Studies in History and Philosophy of Science Part A*. 2006 Mar;37(1):137–151. <https://www.sciencedirect.com/science/article/pii/S0039368105001184>, <https://doi.org/10.1016/j.shpsa.2005.12.015>.
- Kroes P. Engineering and the dual nature of technical artefacts. *Cambridge journal of economics*. 2010;34(1):51–62. <https://doi.org/10.1093/cje/bep019>.
- Kuckartz U, Rädiker S. Transcribing Audio and Video Recordings. In: Kuckartz U, Rädiker S, editors. *Analyzing Qualitative Data with MAXQDA: Text, Audio, and Video*. Cham: Springer International Publishing; 2019. p. 41–49. https://doi.org/10.1007/978-3-030-15671-8_4. https://doi.org/10.1007/978-3-030-15671-8_4.
- Lombrozo T, Wilkenfeld D. Mechanistic versus Functional Understanding. In: Grimm SR, editor. *Varieties of Understanding: New Perspectives from Philosophy, Psychology, and Theology*. Oxford University Press; 2019. .
- McCarthy AM, Keil FC. A right way to explain? Function, mechanism, and the order of explanations. *Cognition*. 2023 Sep;238:105494. <https://linkinghub.elsevier.com/retrieve/pii/S0010027723001282>, <https://doi.org/10.1016/j.cognition.2023.105494>.
- McNeill D. *Hand and mind: what gestures reveal about thought*. Chicago: University of Chicago Press; 1992.
- Miller T. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*. 2019 Feb;267:1–38. <https://www.sciencedirect.com/science/article/pii/S0004370218305988>, <https://doi.org/10.1016/j.artint.2018.07.007>.
- Miyake N. Constructive Interaction and the Iterative Process of Understanding. *Cognitive Science*. 1986 Apr;10(2):151–177. https://onlinelibrary.wiley.com/doi/10.1207/s15516709cog1002_2, https://doi.org/10.1207/s15516709cog1002_2.
- Nückles M, Wittwer J, Renkl A. Information About a Layperson’s Knowledge Supports Experts in Giving Effective and Efficient Online Advice to Laypersons. *Journal of experimental psychology Applied*. 2006 Jan;11:219–36. <https://doi.org/10.1037/1076-898X.11.4.219>.
- Oppenheimer DM, Meyvis T, Davidenko N. Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*. 2009 Jul;45(4):867–872. <https://www.sciencedirect.com/science/article/pii/S0022103109000766>, <https://doi.org/10.1016/j.jesp.2009.03.009>.
- Pennington N. Comprehension strategies in programming. In: Olson GM, Sheppard S, Soloway E, editors. *Empirical studies of programmers: second workshop*. Ablex Publishing Corp.; 1987. p. 100–113. <http://portal.acm.org/citation.cfm?id=54968.54975>. <http://portal.acm.org/citation.cfm?id=54968.54975>.

- Register Y, Ko AJ. Learning Machine Learning with Personal Data Helps Stakeholders Ground Advocacy Arguments in Model Mechanics. In: Proceedings of the 2020 ACM Conference on International Computing Education Research Virtual Event New Zealand: ACM; 2020. p. 67–78. <https://dl.acm.org/doi/10.1145/3372782.3406252>.
- de Ridder J. Reconstructing design, explaining artifacts: philosophical reflections on the design and explanation of technical artifacts; 2007. <https://repository.tudelft.nl/record/uuid:d67f6fe7-59c5-4357-903e-e3c3891e2721>.
- Rohlfing KJ, Cimiano P, Scharlau I, Matzner T, Buhl HM, Buschmeier H, et al. Explanation as a Social Practice: Toward a Conceptual Framework for the Social Design of AI Systems. *IEEE Transactions on Cognitive and Developmental Systems*. 2021 Sep;13(3):717–728. <https://ieeexplore.ieee.org/document/9292993/>, <https://doi.org/10.1109/TCDS.2020.3044366>.
- Schulte C. Block Model - an Educational Model of Program Comprehension as a Tool for a Scholarly Approach to Teaching. In: Proceeding of the Fourth international Workshop on Computing Education Research ICER '08, Sydney, Australia: ACM; 2008. p. 149–160. <http://doi.acm.org/10.1145/1404520.1404535>.
- Schulte C, Budde L. A Framework for Computing Education: Hybrid Interaction System: The need for a bigger picture in computing education. In: 18th Koli Calling International Conference on Computing Education Research (Koli Calling '18), vol. 18 Koli, Finland: ACM; 2018. p. 10.
- Schulte C, Clear T, Taherkhani A, Busjahn T, Paterson JH. An introduction to program comprehension for computer science educators. In: Proceedings of the 2010 ITiCSE working group reports ITiCSE-WGR '10, New York, NY, USA: Association for Computing Machinery; 2010. p. 65–86. <https://doi.org/10.1145/1971681.1971687>.
- Selting M, Auer P, Barth-Weingarten D, Bergmann J, Bergmann P, Birkner K, et al. Gesprächsanalytisches Transkriptionssystem (GAT 2). *Gesprächsforsch*. 2009 Jan;10:152–183.
- Sokol K, Flach P. One Explanation Does Not Fit All. *KI - Künstliche Intelligenz*. 2020 Jun;34(2):235–250. <https://doi.org/10.1007/s13218-020-00637-y>, <https://doi.org/10.1007/s13218-020-00637-y>.
- Soloway E. Learning to program = learning to construct mechanisms and explanations. *Communications of the ACM*. 1986 Sep;29(9):850–858. <https://doi.org/10.1145/6592.6594>, <https://doi.org/10.1145/6592.6594>.
- Spohrer JC, Soloway E, Pope E. A Goal/Plan Analysis of Buggy Pascal Programs. *Human–Computer Interaction*. 1985 Jun;1(2):163–207. http://www.tandfonline.com/doi/abs/10.1207/s15327051hci0102_4, https://doi.org/10.1207/s15327051hci0102_4.

- Streeck J. *Gesturecraft: The manu-facture of meaning*, vol. 2 of *Gesture Studies*. Amsterdam: John Benjamins Publishing Company; 2009. <http://www.jbe-platform.com/content/books/9789027289827>.
- Tedre M, Apiola M. Three computing traditions in school computing education. In: Kadijevich DM, Angeli C, Schulte C, editors. *Improving computer science education*. Routledge New York, NY and London; 2013. p. 100–116. 00013.
- Terfloeth L, Schaffer M, Buhl HM, Schulte C. Adding Why to What? Analyses of an Everyday Explanation. In: Longo L, editor. *Explainable Artificial Intelligence* vol. 1903. Cham: Springer Nature; 2023. p. 256–279. https://link.springer.com/10.1007/978-3-031-44070-0_13. https://link.springer.com/10.1007/978-3-031-44070-0_13.
- Winkelinkemper F, Höper L, Schulte C. ARIadne – An Explanation Model for Digital Artefacts. *Informatics in Education*. 2024 Jun;23(2):479–505. <https://infedu.vu.lt/journal/INFEDU/article/772>, <https://doi.org/10.15388/infedu.2024.09>.
- Wittenburg P, Brugman H, Russel A, Klassmann A, Sloetjes H. ELAN : a professional framework for multimodality research. Genoa, Italy; 2006. p. 1556–1559. <https://hdl.handle.net/11858/00-001M-0000-0013-1E7E-4>.
- Wittwer J, Nückles M, Landmann N, Renkl A. Can Tutors Be Supported in Giving Effective Explanations? *Journal of Educational Psychology*. 2010 Feb;102:74–89. <https://doi.org/10.1037/a0016727>.

4 Interpretation and Discussion

THE four Building Blocks have moved from a methodological foundation for dual-nature coding (BB1) through observation of explanatory defaults (BB2) and their behavior under co-constructive pressure (BB3) to an experimental test of whether the approach to the dual nature matters for understanding outcomes (BB4). Each Author's Commentary has situated its paper's contribution within the dissertation's overall argument. What remains is to examine what the findings, taken together, mean for the Co-Constructive Dual Nature Framework developed in Chapter 2 and for the broader questions about computer science education and explainable AI that motivated the investigation. The findings substantiate several of the framework's core claims but also complicate and refine others relevant for both theory and application. Before reexamining each of the three criteria in detail, this chapter provides a brief overview of the central findings and their interrelations.

Across all empirical studies, two overarching patterns emerge. First, the dual nature of technical artifacts can be operationalized to be an empirically tractable feature of the content of everyday explanations. Perspectives on the dual nature are used to address the artifact in virtually all explanations, and the distinction between them can be coded with great intercoder reliability ($\kappa = .80$). Explainers consistently adopt an Architecture-first strategy, producing roughly 60–70% of their utterances taking the Architecture perspective. This pattern is robust: it initially presented itself under naturalistic conditions (BB2), was present under systematic pressure from confederate explainees' interest (BB3), and (though to a lesser extent) showed even when confederate explainers are trained to maintain a Relevance or integrated focus (BB4). The Architecture-first pattern appears to reflect a genuine explanatory intuition rather than a contextual default, one that explainers rationalize as providing the necessary foundation before Relevance can be meaningfully introduced (BB2).

Second, and more consequentially, the way explainers navigate the dual nature has measurable effects on explainees' understanding of the artifact, but not in the way one might expect. The integrated condition in BB4 produced significantly higher enabledness than either focused condition, while all three conditions produced equivalent comprehension outcomes. Thus, what matters for the kind of understanding that enables agency over the artifact is not how much Architecture or Relevance content is provided, but whether the two are meaningfully connected.

The Architecture-first strategy that explainers find most natural is sufficient for basic comprehension but might be inappropriate for enabledness. At the same time, the ratchet-like dynamic observed in BB4, whereby the introduction of Relevance content shifted explainees toward more balanced engagement, suggests that the two perspectives may not be symmetrically positioned in explanatory dialogue: Architecture may be the necessary foundation, but Relevance is the dimension that, once introduced, reshapes the co-constructive dynamic and interest dynamics of the explainees. The following sections examine these findings through the lens of each criterion, elaborating what is empirically supported, what is complicated or only partially addressed, and what remains open for future investigation.

4.1 The Dual-Perspective Criterion

The dual-perspective criterion identified the integration of Architecture and Relevance as an important element for adequate artifact explanation. The empirical findings substantially satisfy this while complicating what integration requires in practice.

Both perspectives appear in virtually all explanations across BB2, BB3, and BB4, replicating the basic prediction under naturalistic, partially controlled, and fully controlled conditions respectively. The dual nature is empirically tractable as an analytical lens: Architecture and Relevance utterances are distinguishable with “almost perfect” Landis and Koch, 1977, p. 165 intercoder reliability ($\kappa = .80$), validating the coding approach based on intonation phrase segmentation established in BB1 and confirming that the theoretical distinction maps onto observable explanatory behavior. The framework’s central claim that neither perspective alone suffices for genuine understanding receives experimental support in BB4: Architecture-focused and Relevance-focused conditions both fall short of the integrated condition on enabledness outcomes. To the author’s knowledge, this constitutes the first experimental test of this claim, which dual nature theory has maintained primarily on philosophical grounds (Kroes, 2010; Vermaas and Houkes, 2006).

However, the null result on comprehension in BB4 complicates the integration imperative. For low-complexity artifacts such as Quarto, either focused perspective appears sufficient for “knowing-that”, even though neither suffices for deep “knowing-how” (Buschmeier et al., 2025). This suggests that the necessity of integration is outcome-sensitive as integration matters specifically for enabledness and does not appear to be a prerequisite for basic comprehension of Quarto. The instrument’s psychometric limitations bear on how firmly this interpretation can be held, as discussed in Section 4.6: the comprehension null result could reflect either a genuine theoretical finding or measurement insensitivity, and the current data do not cleanly disentangle the two.

However, the null result on comprehension in BB4 complicates the integration imperative. For low-complexity artifacts such as Quarto, either focused perspective appears sufficient for “knowing-that”, even though neither suffices for deep “knowing-how” (Buschmeier et al., 2025). This suggests that the necessity of integration is outcome-sensitive as integration matters specifically for enabledness and does not appear to be a prerequisite for basic comprehension of Quarto. The instrument’s psychometric limitations bear on how firmly this interpretation can be held, as discussed in Section 4.6: the comprehension null result could reflect either a genuine theoretical finding or measurement insensitivity, and the current data do not cleanly disentangle the two.

The Architecture-dominant baseline of 60–70% across all conditions in BB2 and BB3 raises a further question the dual nature framework does not directly address. The theory posits two equally important perspectives; the data consistently show one is treated as more foundational by explainers. Whether this reflects a cognitive default toward mechanistic reasoning in the explainer role, a cultural convention in artifact explanation, or a pragmatic judgment that Architecture content must come first to give both parties something concrete to build on is a question the current data cannot exhaustively answer.

Converging evidence for this Architecture-as-baseline pattern comes from how explainers attribute knowledge. Schaffer and Buhl (2025) found that explainers rated explainees’ knowledge higher on both Architecture and Relevance items when explainees expressed Relevance interest, suggesting that Relevance demands may lead explainers to overestimate the degree to which Architecture has been understood. If Architecture knowledge is taken for granted, expressed Relevance interest can be interpreted as evidence that some foundational layer is already in place. Architecture thus functions as an assumed baseline not only in the content explainers produce but potentially also in how they model what their explainees already know.

4.2 The Collaborative Criterion

The collaborative criterion requires explanations to be adaptive and responsive to the explainee, rather than transmitting content in the sense of the Nürnberger Trichter. Instead, the Co-Constructive Dual Nature Framework implies that via the processes of monitoring, scaffolding, and grounding, the two perspectives on the dual nature can be integrated and aligned across explainer and explainee. The findings demonstrate that co-constructive processes are empirically active in artifact explanation but reveal certain constraints on their operation.

BB3 directly tested the monitoring component: explainers detected and responded to explainee interest signals, demonstrating that monitoring is active in these interactions. The adaptation was content-specific, with explainers increasing Rele-

vance content from 28% to 40% in the Relevance condition, demonstrating that scaffolding responds to monitored signals rather than operating on a predetermined script. Confirmation comes from Schaffer and Buhl (2025), who employed six retrospective video recall measurement points and found that explainers' perception of explainees' interests differed significantly between Architecture and Relevance conditions, with significant interaction effects between condition and time. Notably, perceived Relevance interest increased over the course of explanations in both conditions, consistent with the progressive shift toward Relevance documented in BB2 and BB3. The finding in BB2 that shifts toward Relevance were often triggered by explainee questions, perceived interests, and expressed confusion provides support for the collaborative criterion's claim that explanatory content is shaped by the dynamics of the interaction itself (Rohlfing et al., 2021). These shifts were not planned by the explainers but emerged from the interaction instead. The video recall data from BB2 further show that explainers actively reason about explainee knowledge and interest states when making content decisions. This is precisely the monitoring process co-construction theory identifies as a condition for adaptive explanation (Reusser and Pauli, 2015; Rohlfing et al., 2021), operating through the real-time adjustments to shared understanding that grounding theory describes (Clark and Brennan, 1991; Wilkes-Gibbs and Clark, 1992). Content selection, in the empirical data, is therefore not only a precondition of some explanation plan of the explainer, but becomes a product of an ongoing negotiation between explainer and explainee.

This responsiveness is also where the "moving target" character of everyday explanation (Rohlfing et al., 2021) becomes empirically visible. What began as an Architecture description of game components evolved into a Relevance focus as explainee questions redirected the explanation's focus towards, for example, strategies or purposes (e.g., in BB2). In BB4, the ratchet dynamic provides a further instantiation of this phenomenon. Soon after Relevance content was introduced, explainees' engagement patterns shifted steadily toward more balanced dual-perspective reasoning, effectively influencing what is to be explained from that point forward. The explanandum was, in that sense, not a fixed object that the explanation worked toward but something that took shape through the interplay of explainer intentions and explainee responses. Telling instances from the empirical data in BB2 and BB4 (A-Condition) illustrate this dynamic quite well. In some dyads, if the explainers devoted an unusually extended stretch to Architecture content, the explainees eventually intervened with a direct question targeting Relevance, for example by asking: "What's all that useful for?" These moments capture the co-constructive mechanism through which the explanandum shifts as the explainee's question did not merely request additional information but *reframed* what counted as an adequate explanation, triggering a stance transition from a perspective on the Architecture toward a teleological stance helpful for potentially integrating Relevance.

However, the adaptation operates within a remarkably stable Architecture-dominant baseline that persists even under systematic influence from confederate explainees in BB3. Explainers maintain at least approximately 60% Architecture content regardless of explainee interest condition. This adaptation pattern complicates a strong reading of co-constructive theory: if monitoring is operating, why does it not produce more substantial content reorganization? The findings suggest that the dynamics of co-construction do not simply override what explainers regard as important to convey during an explanation. This is interpretively connected to the “Inference to Best Explanation” (IBE) and “Explaining for the Best Inference” (EBI) distinction developed by Wilkenfeld and Lombrozo (2015). Explaining for the Best Inference (EBI) shifts the focus from product to process, arguing that engaging in the act of explaining improves epistemic standing regardless of whether the explanation produced is the best available (Wilkenfeld and Lombrozo, 2015). In this light, one interpretation is that explainers in the present studies appear to operate, in early stages of the explanations, according to an IBE logic, maintaining Architecture content they regard as prerequisite knowledge rather than reorganizing it in response to explainee signals. The monitoring does occur, but it is subordinated to a content model the explainer treats as important to convey correctly. Only after some while, the explainers slowly shift away from that plan.

The grounding component of the collaborative criterion is the least empirically addressed of the three processes. The studies document what content gets established but not the micro-process by which shared vocabulary and partial understanding are negotiated and maintained across the interaction. Preliminary analyses of lexical convergence in the BB2 and BB4 suggest that shared vocabulary between explainer and explainee increases gradually over the course of explanations, consistent with grounding theory’s prediction that interlocutors progressively align their referential expressions (Clark and Brennan, 1991) (see Figure E.4 and Figure E.22 in the Appendix). The content-coding approach operationalized in BB1 captures what is said but not the sequential negotiation through which what is said becomes mutually understood. Grounding operates at a level of interactional detail that the current methodology was not designed to capture. However, conversation analytic work by Lohmer and Kern (2024) on the same dataset reveals that participants systematically deploy interactive gestures (e.g., palm-vertical open hand gestures and beats) at transition points between explanation nodes and when introducing new information within a node, suggesting that the structuring work associated with grounding may be partially accomplished through nonverbal channels that the verbal content analysis does not capture either.

4.3 The Cognitive Criterion

The cognitive criterion identified flexible stance adoption as a requirement for navigating the dual nature, with mechanistic and teleological modes of construal aligning with the Architecture and Relevance perspectives respectively. The findings are consistent with this alignment but reveal an asymmetry between explainers and explainees that the criterion, as formulated, did not foresee.

The ratchet dynamic identified in BB4, whereby the introduction of Relevance content initiated a shift of the explainees toward more balanced engagement regardless of how much Architecture accompanies it, is consistent with the cognitive criterion's claim that the teleological stance may represent some cognitive default (Kelemen et al., 2013; Lombrozo and Wilkenfeld, 2019). Once triggered, it is sustained with apparent ease, but the BB4 data show that explainees do not start there on their own. The Architecture-first pattern from BB2 can be read through the cognitive criterion as follows: explainers who are domain experts operate comfortably in the mechanistic stance and treat it as the natural entry point; the shift toward Relevance requires a stance transition that is triggered by co-constructive dynamics (BB3), for example, by the explainer's judgment that the explainee is ready or interested (BB2), rather than being planned from the outset. McCarthy and Keil (2023) found that adults prefer function-before-mechanism orderings in explanations and, more broadly, that explanations address wholes before parts—a hierarchical bias that may underlie the function-before-mechanism preference. This stands in interesting tension with the Architecture-first production pattern (cf. *laying out the tools*) observed in BB2: explainers took the mechanistic stance first, while explainees may have started from a teleological perspective (Kelemen et al., 2013). The integration condition in BB4 may have partly resolved this mismatch. It provided the explicit instruction toward bridging the stances which neither party's default stances would produce on its own in early stages of the explanation.

The cognitive criterion (see Section 2.3) treats mode of construal adoption as relatively complementary: mechanistic and functional modes are both available and flexibly deployed depending on context and (part of the) explanandum the explanation focuses on. The data suggest the distribution is asymmetric in practice. Explainers showed a consistent, autonomous pull toward the mechanistic stance that persists regardless of co-constructive pressure (BB3) and is explicitly rationalized in video recall (BB2). Explainees did not show a corresponding autonomous teleological pull, though. BB4's temporal data reveal, instead, a conditional responsiveness: across conditions, explainees' utterances addressing the dual nature were indistinguishable in the first third of the explanation ($p = .433$), but diverged significantly in the second and third thirds ($p < .001$). Architecture-condition explainees maintained Architecture-dominant engagement while those receiving Relevance content shifted toward more balanced focus. The introduction of Relevance by the explainer is the trigger; without it, explainees remain Architecture-oriented.

The challenge of artifact explanation may be one of overcoming a genuine explainer production default on the one hand, and activating a latent explainee responsiveness to Relevance on the other.

The most theoretically productive opening from the cognitive criterion concerns the function of integration itself. The explainer/explainee default stance asymmetry reframes what integration might do cognitively: it may not simply be that both perspectives are present, but that the integrated condition triggered a stance-bridging neither participant would produce intuitively. Explainers default to the mechanistic stance (BB2–3), thus providing the Architecture foundation not necessarily the integration moves. Explainees, responsive to Relevance once it appears but not actively seeking it in its absence, would remain Architecture-oriented without such trigger. When explainers were instructed to integrate the two sides, the integration might provide the connective move that activates the explainee's responsiveness rather than leaving each party to operate within their respective orientation.

This interpretation would explain why integration specifically benefits enabledness. Agency over the artifact requires precisely the stance coordination that integration provides, while basic comprehension might be achieved through either perspective in isolation. Given the limited complexity of Quarto this might have been an especially viable strategy, where either focused condition was sufficient for knowing-that even without the bridging move. Whether this asymmetry holds in the context of explainers who are not domain experts is an open question, as the mechanistic production default may weaken or disappear when the explainer lacks structural fluency with the artifact.

4.4 Revisiting the Research Questions

The three-criterion analysis addresses the internal architecture of the framework; the four research questions posed in Chapter 1 cut across it. This section returns to each in turn, drawing on the material just developed to state concisely what the Building Blocks establish.

RQ1 – Dual Nature Expression. BB1 establishes that intonation-phrase segmentation, rather than sentence- or turn-based segmentation, makes the two perspectives reliably codable at $\kappa = .80$. Applied across three independent corpora (BB2, BB3, BB4), the Architecture/Relevance distinction maps onto observable explanatory behavior in virtually every explanation. The dual nature is therefore empirically tractable as a content-coding distinction, and the methodological step that makes it tractable is the segmentation choice justified in BB1.

RQ2 – Explanatory Patterns. BB2 documents the *laying out the tools* pattern: an Architecture-dominant baseline of roughly 60–70% concentrated in the early phase of the explanation, with Relevance introduced later. BB3 replicates this baseline under systematic pressure from explainee interest signals, and BB4 shows that even confederate explainers instructed to maintain a focused or integrated condition show similar distributional tendencies over time. Across the studies, explainers tend to address Architecture first and Relevance later, when moving back and forth between the two. Whether this already amounts to meaningful integration is not something BB2 and BB3 were designed to assess. BB4 speaks to it indirectly: its focused conditions already contained some content from the other side, yet this incidental presence did not produce the enabledness advantage that the explicitly integrated AR condition achieved. Deliberate integration of the two perspectives, as BB4 operationalizes it and which led to higher enabledness, may therefore not be a naturalistic default.

RQ3 – Collaborative Navigation. Collaborative processes shape the dual-nature navigation on both sides, but asymmetrically. On the explainer side, BB2's video recalls and BB3's confederate manipulation converge on a partial monitoring response: explainees signalling Relevance interest are recognized and partly accommodated, while Architecture appears to function as an assumed baseline that monitoring does not immediately override. On the explainee side, BB4 documents a ratchet dynamic whereby exposure to Relevance content shifts explainees toward more balanced dual-perspective engagement over time. Collaborative navigation is thus visibly at work and recognizable as explainers adapt within an Architecture baseline and respond to explainee interests.

RQ4 – Effect on Understanding. BB4 provides the experimental evidence on this question and yields an outcome-sensitive answer. Architecture-focused and Relevance-focused conditions produced equivalent comprehension scores; the integrated condition produced significantly higher enabledness. Within the scope of a low-complexity artifact, either perspective in isolation is consistent with knowing-that (comprehension), while integration is required for the form of understanding that enables agency over the artifact (enabledness). The interpretive contingencies surrounding this result, in particular the comprehension null and the artifact's low complexity, are addressed in Section 4.6.

Taken together, the four research questions receive answers that are internally consistent but asymmetric in scope: RQ1–RQ3 characterize the dual nature as it appears in explanatory practice, and RQ4 specifies the outcome domain in which that practice makes a measurable difference.

4.5 Implications

The empirical findings carry consequences beyond the empirical results. Two domains are most directly related to this: computer science education, where the integration finding reframes what effective artifact explanation requires from teachers and curricula; and explainable AI, where the co-constructive dynamics and the enabledness result provide ideas for how synthetic explanation systems could be designed and evaluated.

4.5.1 Implications for Education

The BB4 enabledness finding reframes the question of what integration achieves pedagogically when read against the two dominant traditions in computing education. The notional machine tradition (Du Boulay, 1986; Sorva, 2013) treats understanding as correspondence between a learner's mental model and a correct structural account of the mechanism, and the misconceptions tradition (Qian and Lehman, 2017) frames pedagogy as the identification and correction of deviations from that model. Both traditions share the assumption that the epistemic target is a sufficiently correct Architecture-level representation and that the pedagogical task is to build or repair it. This is roughly what the comprehension dimension of BB4 operationalises, and the data show that all three conditions, including the Relevance-focused one, achieved equivalent comprehension. If the findings generalise, one reading is that Architecture-dominant pedagogy does produce the outcome those traditions value. What it does not produce, and what neither tradition addresses, is enabledness: the capacity to act on and with the artifact in practice. The finding does not discredit those traditions within their scope, but it specifies their limitations. When comprehension is the goal, a focused approach (e.g., Architecture focused) is adequate; where the goal, however, is agency over the artifact, it is insufficient.

What distinguishes the integrated condition from either focused one is not the amount or kind of perspective-specific content but the active, recurring bridging between perspectives. This matters because all three BB4 conditions were co-constructive: every explanation was a face-to-face, dyadic interaction in which both parties participated. Co-construction alone, then, did not produce the enabledness advantage. The EBI reframing (Wilkenfeld and Lombrozo, 2015) argues that engaging in explanatory processes, including imperfect ones, can improve one's epistemic position independently of the quality of the resulting explanation; the present data sharpen this claim by specifying what kind of process matters. It is not the co-constructive activity as such that produces agency-enabling understanding, but a specific operation within it: the recurring work of connecting how the artifact is built to what it is for. The BB3 monitoring dynamics are consistent with

this at the level of interactional practice, showing that explainers adapt content in response to explainee signals, but that this adaptation alone does not guarantee the bridging move that the integrated condition makes explicit.

The Architecture-dominant baseline documented across BB2 and BB3 suggests that even motivated, attentive explainers do not reliably default to the integration of both perspectives. Whether the degree of integration that does occur naturally is sufficient for enabledness was not tested in these studies, but BB4's finding that only deliberately instructed integration produced enabledness gains gives reason to doubt it. Frederik et al. (2011) described that such a difficulty may extend to experienced technology teachers, and the BB4 temporal data add another dimension on the explainee side: explainees do not autonomously gravitate toward Relevance but remain Architecture-oriented unless Relevance is actively introduced, at which point engagement shifts toward a more balanced pattern. Evidence that adults prefer function-before-mechanism orderings when evaluating explanations (McCarthy and Keil, 2023) suggests that learners may find Relevance-first entry points more accessible, even when their engagement tracks whatever perspective the explainer provides. For the lay explainers in the present studies, the most immediate explanation of the Architecture-first default is pragmatic and culturally conventional: explaining a board game means laying out the components and rules before turning to strategy, a pattern the BB2 video recall data confirm explicitly. This pattern finds a structural parallel in formal computing education, which has historically been organised around intellectual traditions that each privilege a different epistemic mode without making integration a shared concern (Tedre and Sutinen, 2008; Winkelkemper et al., 2024). The pedagogical consequence is spelled out directly by Sentance et al. (2018, Chapter 5): teaching computer science by first conveying how it works, before or without addressing what it is good for, fails to capture the role of digital technology in everyday life. Instead, teaching towards integration requires not just introducing some Relevance content to curricula that already teach Architecture, but reconceiving the integrated understanding as a learning objective in its own right (cf. Höper and Schulte, 2023, for teaching material and a framework specifically pursuing this aim).

The integration finding also has implications for scaffolding and assessment. Scaffolding integration, for example by tracing how specific architectural features enable particular Relevance functions or by analysing how Relevance requirements constrain architectural choices, is a direct practical consequence. Assessment implications follow from the comprehension/enabledness distinction. Existing tiered competency models differentiate reproduction from transfer and problem-solving, but they do not target integrated understanding *across* the dual nature specifically. BB4 provides a principled empirical basis for such differentiated assessment design: tasks that require explaining why particular implementations achieve specific goals, or predicting how architectural changes affect Relevance, evaluate precisely the bridging competence that the enabledness finding identifies as consequential

for agency (cf. Buschmeier et al., 2025). Failing to assess this dimension leaves genuine uncertainty about whether learners have developed the agency-enabling understanding that integration is meant to produce.

Finally, the explainer/explainee stance asymmetry has direct implications for teacher training. The BB4 temporal data show that explainees do not autonomously gravitate toward Relevance but remain Architecture-oriented unless Relevance is actively introduced, at which point engagement shifts toward a more balanced pattern. Evidence that adults prefer function-before-mechanism orderings when evaluating explanations (McCarthy and Keil, 2023) suggests that learners may find Relevance-first entry points more accessible. Teachers who default to explaining Architecture, especially when starting new topics, may therefore systematically underserve students whose reception preferences favour Relevance, not through lack of effort but through unreflected Architecture orientation that effective PCK development should address (cf. Coe et al., 2014; Shulman, 1986). The developmental evidence reviewed in Section 2.3, that functional reasoning is the earlier-emerging and less cognitively demanding mode, reinforces this point. Teacher training should therefore not merely inform prospective teachers about the stance asymmetry but cultivate an internalised awareness of it, so that the default toward mechanistic explanation becomes an object of professional reflection rather than an unquestioned habit.

4.5.2 Implications for Explainable AI

The content patterns identified in the present study can be situated within broader process models of explanation that provide useful vocabulary for XAI design. El-Assady et al. (2019) propose a phased model in which explanations consist of self-contained explanation blocks, each realising a particular strategy, that are sequenced into phases and combined with verification blocks. Fisher et al. (2023) apply this notion to human game explanations of Quarto, too, operationalising explanation blocks as explanation nodes whose coverage and ordering can be traced across and within dialogues. The present findings extend this line of work by showing that the temporal ordering of explanation blocks, and the content they foreground, are not independent: both are shaped by whether the explainer orients toward the artifact's Architecture or its Relevance, and this orientation itself shifts in response to co-constructive dynamics. Where El-Assady et al.'s model treats explanation blocks as given units to be arranged, the empirical evidence presented here suggests that the composition of these blocks is itself shaped by the explainer's dual-nature orientation as it develops through interaction with the explainee. A broader imbalance in XAI research is relevant here: much of the field has concentrated on what content to present in an explanation, with less attention to how that content is structured and adapted through interaction (cf. Anjomshoae et al., 2019; Fisher et al., 2023). Process models of explanation that aim

to inform XAI design should therefore account not only for sequencing strategies and verification, but also for the perspective-dependent content decisions that shape which aspects of the dual nature are addressed, and how these decisions emerge through the co-constructive process.

The null comprehension result in BB4 has a direct implication for XAI system evaluation: systems optimising for user comprehension of system behaviour are optimising a target that may be relatively easy to satisfy, but insufficiently sensitive to the kind of understanding that supports genuine user agency. Enabledness, the ability to act with and upon the system, is the more demanding and arguably more relevant target for XAI design (see discussions in Miller, 2019; Rohlfing et al., 2021; Sokol and Flach, 2020).

The ratchet dynamic suggests a more specific design principle that can be grounded in El-Assady et al.'s (2019) strategy taxonomy. Their model distinguishes inductive (bottom-up) strategies, which begin with smaller observable details and build toward complex relations, from deductive (top-down) strategies, which begin with the general idea and add detail progressively. The Architecture-first pattern observed across BB2–BB4 is, in these terms, a consistently inductive strategy: explainers start with parts, components, and mechanisms before addressing purposes and functions. The ratchet dynamic, however, suggests that this inductive default may not be optimal for explainee engagement. Introducing Relevance content early and explicitly, even briefly, may reshape user engagement with subsequent Architecture content. In El-Assady et al.'s (2019) terms, this amounts to a deductive entry, framing the artifact's purpose and significance before detailing its mechanisms, followed by inductive elaboration. Such designs that front-load Relevance framing before architectural detail may be more cognitively effective than either a purely inductive or a uniformly balanced approach. For interactive systems, this could be implemented through tools that allow zooming between high-level functionality and low-level mechanisms, with scaffolding that helps users connect levels and dimensions (in the sense of the block model Schulte, 2008a).

The Architecture-dominant baseline across all conditions constitutes a warning for systems designed by technical practitioners. Miller et al. (2017) argue that XAI risks reproducing what Cooper termed the “inmates running the asylum” problem: AI researchers build explanatory systems oriented toward their own mechanistic understanding rather than toward end users' needs. The present findings suggest this is not merely a design oversight but may reflect a genuine cognitive default that extends beyond technical practitioners to artifact explanation more generally. When explainers with domain expertise explain a technical artifact, they may gravitate toward Architecture content regardless of the co-constructive context, and the same default is likely to shape engineer-authored synthetic explanations unless Relevance integration is built in as a design constraint rather than added as an afterthought. The prevalence of feature-importance and saliency-map ap-

proaches in current XAI practice, which foreground *which* internal components drove a decision rather than *why* that decision matters to the user, can be read as an instantiation of this pattern (cf. Anjomshoae et al., 2019; van der Waa et al., 2021). Register and Ko (2020) demonstrate, in teaching machine learning, that combining details on the Architecture with personally meaningful data, effectively expanding the Relevance dimension, helped learners ground their own interest arguments in the model mechanics. It suggests that the mechanistic default can be productively counteracted when Relevance is deliberately foregrounded in the design. Similarly, Sokol and Flach (2020) argue that personalising the content of explanations matters more than personalising the user interface, a position that aligns with the present finding that content-level dual-nature decisions shape understanding outcomes. The issue, then, is not that technical practitioners are incapable of producing Relevance-integrated explanations but that their default to Architecture will not produce such explanations without deliberate design intervention.

The monitoring finding from BB3 also reframes the role of verification in explanation process models. El-Assady et al. (2019) position verification blocks as optional checkpoints at the end of each explanation phase; the present data suggest that in naturalistic explanation, verification is instead continuous, with explainees signalling understanding and interest throughout and explainers adapting their dual-nature orientation in response. This extends Madumal et al. (2019)'s interaction protocol, which formalises explanation as a structured sequence of locutions (explanation, affirmation, follow-up, return question): the content these moves carry shifts with the explainer's dual-nature orientation as the co-constructive process unfolds.

The present data extend this approach: the dual-nature orientation of the explainer shapes not only which moves are made but also what content those moves carry, and this content shifts as the co-constructive process unfolds. Users who can signal their interest alter the explanation they receive, and this alteration, while constrained, is not cosmetic. Co-constructive XAI systems that model user interest states and adapt content accordingly are better motivated by the present findings than systems that generate a single explanation optimised for an assumed user (see discussions in Rohlfing et al., 2021; Sokol and Flach, 2020). Drawing on the partner model findings of Schaffer and Buhl (2025), these dynamics can be translated into concrete criteria for adaptive explanation systems: an explanation start is marked by high interests and low knowledge, continuation by unexpressed knowledge needs, perspective changes by shifting interests, and explanation endpoints by high knowledge and declining interests. These indicators, structured by the dual nature, may thus provide a first operationalisation of the monitoring process for adaptive explanation systems, complementing the content-level findings of the present dissertation with a process-level model of when and why explanations should shift perspective. Taken together, the findings motivate a shift in XAI

evaluation criteria: from measuring explanation accuracy or user satisfaction in isolation toward assessing the quality of co-constructive interaction and the degree of dual-nature integration, or more generally, agency achieved.

4.6 Limitations

The findings establish the framework’s usefulness within a specific but bounded domain of dyadic, face-to-face explanation of a low-complexity artifact with trained or naturalistic explainers and a single explainee in a controlled laboratory setting. The conclusions should therefore be read against the following limitations.

Artifact complexity and architectural granularity. Quarto is a low-complexity explanandum whose Relevance is bounded by a game domain. A more fundamental scope condition concerns what “Architecture” denotes in this empirical context. Which aspects of the Architecture matter is not a fixed property of the artifact but depends on the goal of the explanation. In Quarto, the Architecture one needs to understand includes the pieces’ attributes, the board layout, the governing rules, and their interrelations. This operates at a shallow level of Architecture, but these properties are directly observable and exhaustively characterize what a player needs to know about the artifact’s Architecture to engage with it. Quarto does, however, possess deeper architectural layers (material composition, manufacturing tolerances, production choices) that simply did not matter for the explanation purpose and therefore did not appear in the data.

The Architecture utterances coded in BB2–BB4 are thus Architecture descriptions shaped by the goal of the explanation: whether the goal is to explain how to *play* or how to *produce* Quarto! may determine which aspects of the Architecture become salient and what counts as having agency. It is the bridging of this goal-specific Architecture layer with Relevance that produced the enabledness advantage in BB4. For digital artifacts, therefore, the question which Architecture layer matters depends on the explanation goal. If the goal is competent use of a word processor’s formatting capabilities, the Architecture that matters may remain at the surface of the artifact (menus, formatting options, style templates), and bridging these with their purposes may already suffice for agency, without any penetration into the software’s internal implementation. If the goal is to improve the word processor’s text-processing algorithms, the Architecture that matters shifts downward into implementation details that are inaccessible to most users. Schulte et al. (2010)’s Block Model captures exactly this granularity, distinguishing text surface, program execution, and functional goals as levels at which Architecture understanding can operate. The Architecture that computing educators discuss when they refer to algorithms, data structures, and computational processes (Winkelkemper et al., 2024) typically occupies deeper levels of that hierarchy than the directly

observable properties that the Quarto studies captured. The present findings establish that the integration imperative holds at the end of this continuum, where directly observable Architecture is bridged with Relevance. Whether the same dynamics obtain when the Architecture in question reaches into genuinely deeper, unobservable layers, whether such depth changes the character of the bridging required, and at what granularity level the integration becomes necessary or sufficient for a given understanding outcome, are open questions the present data do not address. When the implications sections of this dissertation draw consequences for computing education or XAI design, those inferences rest on the principled but empirically untested assumption that the integration imperative generalizes across levels of architectural depth.

Absence of the explanandum. Quarto was not present during the explanations. This was a deliberate choice, intended to create a controlled black-box scenario in which explainers could not offload explanatory work onto the artifact's physical properties, and to ensure that verbalization patterns reflected explanatory reasoning rather than perceptual reference. The trade-off is that the absence of the artifact may have inflated the proportion of verbal Architecture content relative to explanations where pointing, demonstrating, or handling could replace or supplement verbal description. Whether the Architecture-dominant baseline reflects a stable explanatory disposition or is partly an artifact of the absent-explanandum design is a question that studies with the game present could address.

Dyadic structure. The co-constructive dynamics observed here are two-person dynamics involving a single designated explainer and a single explainee. Group explanation, human-AI explanation, and asynchronous explanation contexts introduce different monitoring and scaffolding constraints. The explainer/explainee stance asymmetry identified in the cognitive criterion discussion may manifest differently when multiple explainees with heterogeneous defaults are addressed simultaneously, as the explainer cannot adapt to a single explainee's interest profile.

Confederate design constraints. The confederate designs used in BB3 and BB4 improve experimental control at the cost of ecological validity. In BB3, explainees were instructed to signal either Architecture or Relevance interest, which means the monitoring adaptation documented there reflects explainer responsiveness to relatively clear and consistent interest signals. Naturalistic explainees produce more varied, ambiguous, and dynamic interest displays, and the robustness of the monitoring finding under such conditions remains to be established. In BB4, confederate explainers were trained to condition fidelity, which placed different degrees of constraint on their speech production: the integrated condition was the least prescriptive, allowing confederates more latitude in how they structured

the explanation, while the focused conditions required sustained self-regulation by consciously suppressing one perspective. Any advantage observed in the AR condition should be interpreted with this asymmetry in mind, as it cannot be fully disentangled from the differential naturalness of the confederate's production.

Verbal operationalization. The content layer provides empirical access to the framework's constructs, but the verbal record is an incomplete window onto the co-constructive process. The decision to use intonation-phrase segmentation as the coding unit is theoretically motivated (Chafe, 1994; Park, 2002; Simpson, 2016) and produces strong reliability, but it means that the Architecture/Relevance proportions that figure prominently in the findings are contingent on this specific segmentation choice. A different transcription unit might produce differently sized segments and different proportions, and the content patterns may not transfer directly to multimodal analyses that include gestural and embodied behavior. The cognitive and collaborative layers of the framework remain underdetermined by the current verbal evidence: the inferential step from content shifts to stance transitions and grounding processes rests on theoretical claims that are principled but not independently validated within the present data. This could be improved by richer, qualitative analyses using more sophisticated methods than qualitative content analysis.

Measurement and sample. The understanding instrument developed for BB4, while content-validated across multiple independent samples, rests on a factor structure derived using Pearson rather than tetrachoric correlations, which are more appropriate for binary item responses. The comprehension null result in BB4 could therefore reflect either a genuine theoretical finding (i.e., both perspectives being sufficient for knowing-that) or measurement insensitivity in the comprehension subscale; the present data do not cleanly separate these interpretations. Replication with a revised instrument would be required to resolve this ambiguity. Additionally, the sample, predominantly university students from a single German institution, constrains the generalizability of both the content-pattern findings and the understanding outcomes, as participants likely share similar educational backgrounds and prior knowledge profiles, and culturally shaped explanation norms may differ across institutional or national contexts.

4.7 Closing

The three criteria were designed to guide the theoretical review toward an integrated analytical framework; the empirical findings enrich each of them while identifying specific points of refinement that a purely theoretical treatment could

not have anticipated. The central result is that integration produced a measurable difference in enabledness that neither isolated perspective achieved. It now grounds the framework's core claim in observable explanatory practice. What the findings also establish, however, is a tension that the integration imperative alone does not resolve: the form of explanation that produces agency is not necessarily the form that naturalistic explainers did reliably produce. The following chapter returns to the origins of that tension, draws out its theoretical and practical consequences, and connects it to the broader questions about technological literacy, computer science education, and explainable AI that motivated the investigation.

5 Conclusion

VON Neumann's observation, as relayed by Dijkstra, draws attention to a contrast between simple and complex mechanisms: for the former, it is easier to describe how they work than what they do; for the latter, the reverse tends to hold. The board game Quarto is, by any measure, simple. The explainers in this dissertation's studies did almost exactly what von Neumann predicted: they reached first for how it works, dwelling on pieces, boards, and rules before turning to what the game is for, the strategy, the decisions, the experience of play. The laying out the tools pattern in the explanation data is consistent with von Neumann's observation in its simplest reading.

But the empirical findings do not stop there. The experimental evidence from Building Block 4 establishes something von Neumann's contrast does not capture: that choosing one side, however naturally, may be insufficient. Architecture-focused explanations and Relevance-focused explanations resulted in explainees who could comprehend the game but could not play it well. Only explanations that bridged the two perspectives produced enabledness. The problem with technical artifact explanation, from an epistemic standpoint, is therefore not which description is easier, but that neither description alone is adequate. The task for effective explanation is their integration.

Yet the data reveal something that the integration imperative alone does not predict. The explainers in these studies were neither indifferent nor inattentive; they monitored their explainees, adapted to expressed interests, and rationalized their content choices explicitly. They were, by any reasonable standard, doing their best. And still, integration did not emerge as their default. Instead, they consistently established an Architecture-dominant baseline first, treating it as a necessary scaffold. This pattern persisted even when explainees signalled interest in Relevance (BB3) and even when explainers were instructed to address both sides (BB4). The puzzle is not that explainers neglect one perspective but that connecting both perspectives requires a deliberate move they do not routinely make. The Co-Constructive Dual Nature Framework locates the origins of this puzzle in three places: in the ontological structure of technical artifacts, which makes the two perspectives irreducible to one another; in the co-constructive dynamics, which can prompt but do not guarantee perspective shifts; and in the asymmetric cognitive stances that explainers and explainees bring to the interaction, which might pull in different directions. Documenting that the Architecture-dominant default persists even under co-constructive pressure and demonstrating that deliberately integrated

explanation produces measurably different understanding when it is provided are the contributions of this work. Ensuring that such integration becomes a feature of explanatory practice, becomes a deliberate choice of teachers instructing in classrooms, or is made use of in synthetic explanations of XAI systems is the task it leaves behind.

5.1 Summary of Findings

The dissertation pursued the core argument through four Building Blocks that moved from methodological foundation through naturalistic observation to experimental investigation.

Building Block 1 established that the choice of transcription system is not a neutral preparatory decision but an interpretive one with measurable consequences for what becomes analytically visible. The transition from verbatim to GAT2 transcription, which segmented speech into intonation phrases rather than sentences, resolved an intercoder reliability issue that had no solution at the level of coding manual design, and in doing so grounded the claim that observable content transitions correspond to some cognitive events.

Building Block 2 documented what explainers naturally do without specific guidance: they adopt an Architecture-first strategy in the early stages of an explanation and shift gradually toward Relevance as the interaction develops. This *laying out the tools* pattern was robust across nearly all dyads and was explicitly validated by explainers in video recalls as a deliberate choice. Both sides of the dual nature appeared, albeit in largely different distributions, in virtually all explanations, indicating the framework's prediction that neither perspective entirely displaces the other in naturalistic settings.

Building Block 3 tested the limits of that natural pattern under controlled influence of explainee interests. Explainers partially adjusted to explainee interest signals, increasing Relevance content by approximately twelve percentage points in the Relevance condition. However, they persistently maintained an Architecture-dominant baseline, even when explainees persistently steered toward Relevance. Architecture appears to be treated not as a preference but as a structural prerequisite that effective monitoring does not simply override.

Building Block 4 closed the causal loop. Architecture-focused and Relevance-focused explanations produced identical comprehension outcomes, a null result that is as theoretically informative as the positive finding. Within the constraints of a low-complexity artifact, either perspective appears sufficient for comprehension, for knowing-that. Their integration, however, produced significantly higher enabledness scores. The distinction between comprehension and enabledness, following Buschmeier et al. (2025), proved to be the axis on which the integration

imperative turns: the dual nature framework's central claim holds, but it holds specifically for the form of understanding that enables genuine agency over the artifact.

5.2 Theoretical Contributions

The dissertation makes three contributions that extend beyond the specific findings of the individual Building Blocks.

Empirical grounding of the dual nature theory as an explanatory framework. The dual nature theory of technical artifacts has been developed primarily as a philosophical account of artifact ontology (Kroes, 2010; Vermaas and Houkes, 2006), with little of the experimental grounding that comparable cognitive-science constructs enjoy. This dissertation provides a systematic empirical test of its claims in everyday explanatory settings. The finding that both perspectives are applied in everyday explanations, and that their systematic integration has measurable consequences for understanding outcomes, substantiates the theory's central claims. A second contribution operates at the level of theoretical connection: by mapping Architecture onto the mechanistic mode of construal and Relevance onto the teleological one, the dissertation establishes an interface between dual nature theory and the substantial experimental literature on stances and modes of construal (e.g., Dennett, 1987; Keil, 1994; Kelemen et al., 2013; Lombrozo and Wilkenfeld, 2019). That literature supplies an empirical foundation, including developmental trajectories, default-mode findings under cognitive load, and function-before-mechanism preferences, that dual nature theory on its own does not possess. The mapping thus allows dual nature theory to inherit empirical purchase from a neighbouring tradition while the tradition gains, in return, a theoretically principled account of why the two modes are irreducible when the explanandum is a technical artifact. Together, these moves shift dual nature theory from a philosophical framework toward one useful for empirical analysis of everyday explanatory practice.

The Co-Constructive Dual Nature Framework as an analytical and generative tool. The framework synthesizes dual nature theory, co-constructive explanation research, and cognitive stance theory into an integrated account of technical artifact explanation. Its analytical utility is demonstrated by the Building Blocks' ability to produce cumulative, mutually interpretable findings across different methodological approaches. Its generative utility lies in the constraints it specifies for (synthetic) explanation design. Effective artifact explanation requires *integrating* dual nature content, should support co-constructive processes, and be attentive to the cognitive stance alignment simultaneously. In other words, the framework provides a principled basis for evaluating, and designing explanations in educational and XAI contexts.

A methodological toolkit for temporal content analysis of dialogues. Beyond its theoretical and empirical contributions, the dissertation develops a set of analytical tools that are reusable for studying how dialogues distribute content over time. The temporal normalization approach, introduced in Building Block 2 and refined across subsequent studies, rescales dialogues of varying length to a common timeline and represents coded content as continuous speaker-separated distributions across that timeline. The resulting visualizations proved to be the primary instrument through which the temporal patterns reported in this dissertation became visible: the *laying out the tools* pattern (BB2), the adaptation-within-baseline dynamic (BB3), and the ratchet-like shift in explaineé engagement (BB4) were each identified through visual inspection of normalized timelines before being confirmed statistically. The complementary phase-based analysis, which partitions the normalized timeline into thirds for mixed-ANOVA comparison, converts these visual patterns into testable hypotheses about phase-dependent content dynamics. Taken together with the dual-nature coding manual, whose reliability was established across two independent applications, these tools constitute a portable methodological package for investigating how any binary or categorical content distinction unfolds temporally in dialogic interaction. They are not specific to the dual-nature categories and could, in principle, be applied to other content codings in any setting where time-aligned coded transcripts are available.

The explainer–explaineé stance asymmetry as a theoretical refinement. The cognitive criterion as formulated in the existing literature treats mechanistic and teleological stances as symmetrically available and context-dependently deployed. The combined evidence from Building Blocks 2, 3, and 4 suggests a more specific claim: explainers of Quarto show a consistent pull toward the mechanistic stance, while explaineés, though also gravitating toward Architecture-oriented content, prove specifically responsive to Relevance. Shortly after it is introduced, they are more often taking the teleological stance. It is this responsiveness that reshapes the co-constructive dynamic in ways that Architecture content alone does not. Integration performs the stance-bridging work that neither party’s default would produce on its own. This asymmetry is not captured by the existing framework and represents an empirical refinement with direct implications for how co-construction theory accounts for explanatory practice in technical domains.

5.3 Directions for Future Research

The following future directions are organized by proximity to the present data, moving from re-analyses of existing material toward broader extensions of the research programme.

Further analyses of the present data. The datasets collected across the Building Blocks support several lines of analysis not yet pursued, some of which have been explored in preliminary form. First, the transcripts provide access to the development of shared vocabulary over the course of each explanation. Tracking how artifact-specific terms stabilize and spread across turns would offer a fine-grained window into lexical grounding, revealing which labels for piece attributes, board positions, and rules the explainees adopts, and when. Exploratory analyses of lexical convergence patterns in the present data suggest that this approach is tractable and yields interpretable trajectories, but systematic validation against understanding outcomes remains to be conducted (see Figure E.4 and Figure E.22). Second, the coded content sequences can be visualized as dual-nature flow profiles that make the temporal structure of each explanation inspectable at a glance (see Appendix F). Exploratory visualizations of this kind have proven useful for identifying candidate integration episodes and for locating stretches of extended single-perspective content, providing a descriptive foundation for the more systematic sequential analyses described below (see Section E.4). Third, the integration episodes identified quantitatively invite qualitative treatment: how explainers accomplish perspective shifts—through explicit bridging formulations, example selection, or responsive uptake of explainees questions—is a question the present analysis and methods can locate but not resolve. A deeper qualitative examination of these transition points would specify what, concretely, makes some perspective shifts succeed in integration while others stall.

Artifact complexity and the integration imperative. The most immediate open question is whether the null comprehension result from Building Block 4 reflects the low complexity of Quarto or a more general pattern. For complex digital artifacts such as recommendation algorithms or medical decision-support systems, the Architecture is partially inaccessible to the explainer and the Relevance may be contested across user communities. Under these conditions, integration may require different explanatory moves than the deliberately integrated explanations observed here, and the comprehension/enabledness distinction may become harder to separate empirically because comprehension itself presupposes some grasp of how the two perspectives relate. Extending the research program to such artifacts is the most direct path to testing the generalizability of the framework's claims. Such extensions would also sharpen what the present data already suggest: that the productive question is not whether both perspectives are present, but what kind of integration a given understanding outcome requires. The comprehension/enabledness distinction (Buschmeier et al., 2025) provides the starting point, but its systematic embedding into the Co-Constructive Dual Nature Framework remains a task for future work.

A related strategy for extending the framework to opaque artifacts is to apply the integration imperative not to a specific artifact's actual Architecture, which may be partially inaccessible due to proprietary algorithms, patented components, or sheer complexity, but to *explanation models* (Winkelkemper et al., 2024, cf.)

that approximate what is reliably known about a class of such artifacts. A fitness tracker's exact sensor fusion algorithm may be inaccessible, instead some Architecture insights, that it combines accelerometer and heart-rate data, that it applies smoothing heuristics, that it calibrates against population norms, might hold across different tracking devices and thus may be sufficient to ground meaningful integration with Relevance. The hypothesis would be that integration at the level of such approximate models can produce enabledness even when full architectural transparency is unavailable, offering one solution how integration might still be achievable in black-box scenarios. Whether approximate Architecture content supports the same comprehension-to-enabledness transition observed in BB₄, or whether it produces a qualitatively different form of understanding, is an empirical question the present data cannot answer but that the framework is well positioned to investigate.

Multimodal and process extensions. The verbal content analysis does not address how gestural behavior interacts with dual-nature structure. Work by Lohmer and Kern (2024) and Lohmer et al. (2023) shows that participants deploy interactive gestures systematically at transition points between explanation nodes; Kern et al. (2024) links reinforcing iconic gestures to scaffolding intensity, connecting multimodal behavior directly to the collaborative dynamics the framework theorizes but does not yet address at the gestural level. The intonation phrase segmentation established in Building Block 1 provides a principled alignment unit for integrating verbal and gestural coding, making this extension methodologically tractable. At the process level, the integration episodes identified here serve as focal episodes for conversation-analytic examination of how monitoring, scaffolding, and grounding unfold at the moment of perspective shift. For cognitive alignment, stimulated recall anchored to content shift locations would allow direct examination of the stance transitions the framework predicts; initial steps in this direction have been taken by Lazarov et al. (2026) and Schaffer and Buhl (2025), who demonstrated that retrospective video recall at targeted measurement points yields condition-sensitive and temporally differentiated data within the same empirical paradigm. Another promising way forward is continuous post-hoc annotation of comprehension and confusion by either explainer or explainee while re-watching the recorded explanation, using instruments such as PAGAN (Melhart et al., 2019). Aligned to the intonation-phrase units established in BB₁, such traces would yield time-resolved understanding signals that can be related directly to the Architecture/Relevance content distribution, providing a finer-grained empirical basis for the stance-transition claims the framework makes but the current content coding alone cannot test.

Educational applications. The finding that integration is necessary for enabledness (BB₄), combined with the observation that explainers' default content distribution is Architecture-dominant even under co-constructive pressure (BB₂, BB₃), opens several testable research directions. A concrete starting point is teacher education: the dual-nature distinction can function as a reflective instrument, training

prospective teachers to identify which perspective their explanations foreground and to deliberately construct bridging moves toward the other. In a sense, similar to teachers having to make informed decisions about which tradition (Tedre and Apiola, 2013) their material addresses, the idea would be that such deliberate choice has positive consequences on learning outcomes, yet whether such training measurably improves students' enabledness remains an open empirical question. Assessment tasks can correspondingly be designed to probe not only how well learners understand either side in isolation, but how well they can articulate the relations between them; whether such integrative assessments predict transfer performance better than conventional content-focused measures would test the framework's evaluative claims directly. The Architecture-first pattern also raises a curriculum-sequencing question: if explainers reliably treat structural content as a prerequisite, whether instruction benefits from making that phase explicit and bounded, with a deliberate transition toward Relevance content, can be tested through controlled comparison of sequencing strategies.

XAI and synthetic explanation design. The Architecture-first pattern and the integration imperative raise three open questions for synthetic explanation systems: whether the Architecture-first sequence should be replicated or inverted in line with the function-before-mechanism preference documented by McCarthy and Keil (2023); whether co-constructive perspective-shift triggers analogous to those observed in BB2 and BB3 can be detected from user signals algorithmically; and whether constraining LLM-generated explanations to empirically observed or theoretically idealized dual-perspective flow patterns improves their effectiveness over unconstrained generation. The third question has a concrete methodological path: clustering coded content sequences from a larger sample with accompanying understanding measurements to identify empirical explanation types, which preprocessing could abstract into idealized flow templates for synthetic generation (cf. Wachsmuth et al., 2015, for useful preprocessing ideas). The present data demonstrate the feasibility of flow description at the sequence level (see Figure F.1), but BB4's conditions are too small for the clustering-to-outcome connection this pipeline requires; a dedicated study combining controlled conditions with a sample sized for cluster analysis would be the natural next step. Across all three questions, the BB4 finding that comprehension alone is an insufficient evaluation target applies.

A complementary, more immediately applicable direction concerns the explainee side. For synthetic dialogue-based explanation systems that do not yet integrate both perspectives by design, a pragmatic interim strategy is to provide explainees with structured guidance, for instance in the form of question prompts or interaction scaffolds, that helps them elicit integrated explanations from synthetic dialogue-based explanation systems. Rather than waiting for all synthetic explainers to be redesigned, such scaffolds would leverage the co-constructive capacity that the present data attribute to the explainee role. Whether explainee-initiated prompts produce integration effects comparable to those observed under explainer-

driven conditions in BB4 is an open empirical question, but the co-constructive dynamics provides a principled reason to expect that they can shift the explanatory trajectory.

The present dissertation is situated within a broader collaborative research effort (cf. <https://trr318.upb.de>), and several parallel lines of work by colleagues extend its findings in directions the present data can open but not resolve. On the XAI side, Robrecht and Kopp (2023) and Robrecht-Hilbig (2026) develop sequential, non-stationary decision-process models for adaptive explanation generation that operationalize the rational explainer as a partner-model-driven agent, asking how the co-constructive monitoring dynamics documented here can be formalized and implemented in synthetic systems. Mindlin et al. (2024) and Mindlin and Cimiano (2025) address the complementary evaluation question: how user understanding can be measured within dialogue-based XAI systems across the continuum from static to conversational interfaces. On the content and structure side, Fisher et al. (2023) examine the semantic dialogue patterns of game explanations of the same explanandum, complementing the dual-nature content analysis with a semantic-structural account at the level of nodes and their sequencing. On the multimodal side, Lohmer and Kern (2024) investigates how gestures function in the same explanatory interactions, extending the verbal content account toward the embodied channels the present methodology leaves underspecified. Finally, Schaffer and Buhl (2025) address how explainers' representations of explainee knowledge change over the course of explanations as a function of adaptivity, providing a process-level account of the monitoring dynamics that the content-coding approach of this dissertation captures only in aggregate.

5.4 Broader Implications

The findings of this dissertation have consequences beyond the specific empirical domain of board game explanations. Two areas in which the dual nature tension is most consequential are education and explainable AI.

5.4.1 For Education

The introduction posed a question this dissertation can now answer, at least for everyday explanations of technical artifacts: how do we help people develop meaningful technological literacy when the artifacts they encounter grow increasingly complex and opaque (Pasquale, 2016)? Literacy and competency frameworks in computing and technology education have tended to organize the relevant content around either the structural logic of technology or its situated use, rarely both in relation (see e.g., Clark and Boyle, 2006; Jia et al., 2025; Tedre and Apiola, 2013; Wing, 2008). The findings of this dissertation suggest that this separation

is inadequate for the form of understanding that underpins genuine agency, and that frameworks targeting self-determination as a learning outcome (cf. Höper and Schulte, 2023) point in the more productive direction.

The broader trajectory of technological development underscores the urgency of this point. If the people who design, deploy, and regulate technical systems have been educated through curricula that separate Architecture from Relevance knowledge, they may arrive technically fluent but without the habit of connecting how a system works to what it means for those affected by its operation. The dissertation's findings suggest that this is not a motivational failure but a structural one: even motivated, attentive explainer did not produce integration by default, and there is little reason to assume that curricula which do not explicitly target integration will produce it either. The risk, in other words, is not that future practitioners will lack technical competence but that they will exercise it without the integrative understanding needed to reason about whether and how a system should be built, not only how it can be.

5.4.2 For Explainable AI

The Introduction framed the black-box problem in AI as a dual nature problem: feature-importance maps provide some insight into Architecture with limited Relevance, while decision rationales provide Relevance often without sufficient grounding in Architecture. As neither alone produces enabledness, the absence of integration invites predictable threats, such as, for example, unhealthy trust (cf. Peters and Visser, 2023). The dissertation's empirical findings now substantiate that framing with evidence from human explanatory practice.

The vision of co-constructive explanation systems developed here converges with the broader programme of Social Explainable AI recently articulated by Rohlfing et al. (2026), which defines sXAI as systems that interactively adapt to users in order to co-construct satisfying explanations, organized around three properties of social interaction: that it is patterned, incremental, and multimodal. The empirical findings of this dissertation speak most directly to the first two. The laying out the tools pattern and its co-constructive modulation by explainee interest (BB2, BB3) provide empirical content for what the sXAI programme calls patterned interaction; the ratchet dynamic and the monitoring-driven perspective shifts (BB3, BB4) instantiate what it calls incrementality. What the present work adds is a content-level specificity that the sXAI framework can be extended by: the dual-nature orientation of the explainer determines not only *how* the explanation unfolds interactively but *what* it contains at each stage. Because Architecture and Relevance follow from the ontological structure of technical artifacts as such, these content-level findings are in principle transferable to any technical artifact that needs explaining. The third property, multimodality, falls outside the main scope of the present work and is taken up in complementary analyses of the same dataset

(Lohmer and Kern, 2024). For XAI design, the central risk that follows from the present data is not that synthetic explanations will be too technical or too simple, but that they will default to a single perspective; what comprehension-focused evaluation cannot see, and what enabledness-focused evaluation can, is precisely the gap Pasquale’s black-box concern identifies at a societal scale.

5.5 Closing Remarks

The dissertation started with von Neumann’s contrast, relayed by Dijkstra, between simple and complex mechanisms. At its core, it is a contrast about explanatory access: which side of the dual nature offers the easier entry point depends on how opaque the other side has become. For Quarto, the mechanisms are visible and the rules can be laid out. For a neural network classifying medical images, the mechanisms are largely inaccessible; only the decisions can be described. Both cases represent the same underlying challenge displaced along a scale of complexity.

What this dissertation adds is a finding about what happens at the level of understanding when only one side is addressed. For a simple artifact, explaining how something works or what it is for may be sufficient for a person to demonstrate comprehension. It is not sufficient for enabledness and thus the capacity to act competently with the artifact. Whether this holds equally for complex artifacts, where even comprehension may presuppose some integration of the two sides, is the most pressing question left open. But the gap between knowing and doing, between comprehension and enabledness, is the gap that integration of Architecture and Relevance bridged in the present data, and there is no theoretical reason to expect it to narrow as artifacts become more complex. Whether the artifact is a board game or an algorithm shaping a consequential decision, the question is the same: under what conditions can an explanation help someone not merely comprehend a system but act with it? This dissertation has taken one step toward answering it.

6 Publications

This chapter lists the scholarly publications produced during my doctoral work at the Collaborative Research Centre TRR 318 “Constructing Explainability” and the Computing Education Research group at Paderborn University from 2019 until 2026. The four first-authored papers that constitute the empirical core of this dissertation (BB1–BB4) are marked with an asterisk (*). The remaining entries document related contributions to collaborative projects, and co-authored work.

* L. Terfloth et al. (2023). “Adding Why to What? Analyses of an Everyday Explanation”. en. In: *Explainable Artificial Intelligence*. Ed. by L. Longo. Vol. 1903. Cham: Springer Nature, pp. 256–279. ISBN: 978-3-031-44070-0. DOI: [10.1007/978-3-031-44070-0_13](https://doi.org/10.1007/978-3-031-44070-0_13)

* L. Terfloth et al. (Apr. 2025). “Transcription in Computing Education Research: A Review and Recommendations”. en. In: *Informatics in Education*. ISSN: 1648-5831, 2335-8971. DOI: [10.15388/infedu.2025.09](https://doi.org/10.15388/infedu.2025.09)

* (in press) L. Terfloth et al. (n.d.). “Navigating the Dual Nature: Do Explainers Adapt to Explainee Interests When Explaining Technical Artifacts”. en. In: *International Journal of Technology and Design Education* ()

*(submitted) L. Terfloth et al. (Mar. 2026). *Bridging the Dual Nature: How Integrated Explanations Enhance Understanding of Technical Artifacts*. arXiv:2603.24325 [cs]. DOI: [10.48550/arXiv.2603.24325](https://doi.org/10.48550/arXiv.2603.24325)

H. Buschmeier et al. (Dec. 2025). “Forms of understanding for XAI-Explanations”. In: *Cognitive Systems Research* 94, p. 101419. ISSN: 1389-0417. DOI: [10.1016/j.cogsys.2025.101419](https://doi.org/10.1016/j.cogsys.2025.101419)

L. Fichtel et al. (Aug. 2025). “Investigating Co-Constructive Behavior of Large Language Models in Explanation Dialogues”. In: *Proceedings of the 26th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. Ed. by F. Béchet et al. Avignon, France: Association for Computational Linguistics, pp. 1–20

M. E. Schaffer et al. (Nov. 2024). *Explainers' Mental Representations of Explainees' Needs in Everyday Explanations*. DOI: [10.48550/arXiv.2411.08514](https://doi.org/10.48550/arXiv.2411.08514)

V. Lohmer et al. (2023). "Explaining the Technical Artifact Quarto!: How Gestures are used in Everyday Explanations". English. In: *1st International Multimodal Communication Symposium*. Barcelona Spain, pp. 133–134

H. Beierling et al. (Sept. 2024). "What you need to know about a learning robot: Identifying the enabling architecture of complex systems". In: *Cognitive Systems Research* 88, p. 101286. DOI: [10.1016/j.cogsys.2024.101286](https://doi.org/10.1016/j.cogsys.2024.101286)

L. Terfloth et al. (Nov. 2020). *Combining Ideas and Artifacts: an Interaction-Focused View on Computing Education Using a Cybersecurity Example*. Pages: 5. DOI: [10.1145/3428029.3428052](https://doi.org/10.1145/3428029.3428052)

Bibliography

- Amerongen M., v. (2008). "The interpretation of artifacts : a critique of Dennett's design stance". Phd Thesis 1 (Research TU/e / Graduation TU/e). Eindhoven: Technische Universiteit Eindhoven. ISBN: 978-90-386-1339-0. DOI: [10.6100/IR636331](https://doi.org/10.6100/IR636331) (cit. on pp. 45, 46).
- Anjomshoae, S., A. Najjar, D. Calvaresi, K. Främling (2019). "Explainable Agents and Robots: Results from a Systematic Literature Review". In: *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems. AAMAS '19*. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems, pp. 1078–1088. ISBN: 978-1-4503-6309-9 (cit. on pp. 197, 199).
- El-Assady, M., W. Jentner, R. Kehlbeck, U. Schlegel, R. Sevastjanova, F. Sperrle, T. Spinner, D. Keim (May 2019). "Towards XAI: Structuring the Processes of Explanations". In: (cit. on pp. 19, 197–199).
- Attisano, E., S. E. Nancekivell, S. Denison (2021). "Components and Mechanisms: How Children Talk About Machines in Museum Exhibits". In: *Frontiers in Psychology* 12, p. 1737. ISSN: 1664-1078. DOI: [10.3389/fpsyg.2021.636601](https://doi.org/10.3389/fpsyg.2021.636601) (cit. on pp. 32, 48).
- Baker, L. R. (June 2004). "The ontology of artifacts". In: *Philosophical Explorations* 7.2, pp. 99–111. ISSN: 1386-9795. DOI: [10.1080/13869790410001694462](https://doi.org/10.1080/13869790410001694462) (cit. on p. 24).
- Beierling, H., P. Richter, M. Brandt, L. Terfloth, C. Schulte, H. Wersing, A.-L. Vollmer (Sept. 2024). "What you need to know about a learning robot: Identifying the enabling architecture of complex systems". In: *Cognitive Systems Research* 88, p. 101286. DOI: [10.1016/j.cogsys.2024.101286](https://doi.org/10.1016/j.cogsys.2024.101286) (cit. on p. 216).
- Betz, N., A. McCarthy, F. Keil (Oct. 2021). *Developmental Disconnects: Adults and Teachers Deprioritize Mechanistic Content when Introducing Young Children to STEM*. en-us. DOI: [10.31219/osf.io/5hpzg](https://doi.org/10.31219/osf.io/5hpzg) (cit. on p. 33).
- Boix Mansilla, V., A. Schleicher (2022). *Big picture thinking: How to educate the whole person for an interconnected world* (cit. on p. 20).
- Budde, L. (2021). "Entwicklung und Rekonstruktion einer interaktionsgeprägten Sichtweise auf das komplementäre Mensch-Maschine-Verhältnis". ger. PhD thesis. Paderborn. DOI: [10.17619/UNIPB/1-1119](https://doi.org/10.17619/UNIPB/1-1119) (cit. on pp. 25, 26, 30, 33, 43, 46).
- Buhl, H. M. (Nov. 2001). "Partner Orientation and Speaker's Knowledge as Conflicting Parameters in Language Production". en. In: *Journal of Psycholinguistic Research* 30.6, pp. 549–567. ISSN: 1573-6555. DOI: [10.1023/A:1014217421749](https://doi.org/10.1023/A:1014217421749) (cit. on p. 42).

- Buhl, H. M., J. B. Fisher, K. Rohlfing (2024). "Changes in Partner Models – Effects of Adaptivity in the Course of Explanations". en. In: *Proceedings of the Annual Meeting of the Cognitive Science Society*. Vol. 46. Cognitive Science Society (cit. on pp. 37, 42).
- Buschmeier, H., H. M. Buhl, F. Kern, A. Grimminger, H. Beierling, J. Fisher, A. Groß, I. Horwath, N. Klowait, S. Lazarov, M. Lenke, V. Lohmer, K. Rohlfing, I. Scharlau, A. Singh, L. Terfloth, A.-L. Vollmer, Y. Wang, A. Wilmes, B. Wrede (Dec. 2025). "Forms of understanding for XAI-Explanations". In: *Cognitive Systems Research* 94, p. 101419. ISSN: 1389-0417. DOI: [10.1016/j.cogsys.2025.101419](https://doi.org/10.1016/j.cogsys.2025.101419) (cit. on pp. 19, 23, 50, 151, 153, 188, 189, 197, 206, 209, 215, 265).
- Castelvecchi, D. (Oct. 2016). "Can we open the black box of AI?" en. In: *Nature News* 538.7623. Cg_type: Nature News, p. 20. DOI: [10.1038/538020a](https://doi.org/10.1038/538020a) (cit. on p. 20).
- Chafe, W. (1994). "DISCOURSE, CONSCIOUSNESS, AND TIME". en. In: p. 334 (cit. on pp. 54, 58, 202).
- Chi, M. T. H. (1997). "Quantifying qualitative analyses of verbal data: A practical guide". In: *Journal of the Learning Sciences* 6.3, pp. 271–315. ISSN: 1532-7809. DOI: [10.1207/s15327809jls0603_1](https://doi.org/10.1207/s15327809jls0603_1) (cit. on p. 53).
- (2000). "Self-explaining: The dual processes of generating inference and repairing mental models". In: *Advances in instructional psychology: Educational design and cognitive science*, Vol. 5. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers, pp. 161–238. ISBN: 978-0-8058-2549-7 (cit. on p. 36).
- Chi, M. T. H., S. A. Siler, H. Jeong (2004). "Can tutors monitor students' understanding accurately?" In: *Cognition and Instruction* 22.3, pp. 363–387. ISSN: 1532-690X. DOI: [10.1207/s1532690xc2203_4](https://doi.org/10.1207/s1532690xc2203_4) (cit. on pp. 17, 40, 43).
- Chi, M. T. H., R. Wylie (Oct. 2014). "The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes". In: *Educational Psychologist* 49.4, pp. 219–243. ISSN: 0046-1520. DOI: [10.1080/00461520.2014.965823](https://doi.org/10.1080/00461520.2014.965823) (cit. on p. 41).
- Clark, H. H., S. E. Brennan (1991). "Grounding in communication". In: *Perspectives on socially shared cognition*. Washington, DC, US: American Psychological Association, pp. 127–149. ISBN: 978-1-55798-121-9. DOI: [10.1037/10096-006](https://doi.org/10.1037/10096-006) (cit. on pp. 17, 38, 190, 191).
- Clark, H. H., M. A. Krych (Jan. 2004). "Speaking while monitoring addressees for understanding". en. In: *Journal of Memory and Language* 50.1, pp. 62–81. ISSN: 0749596X. DOI: [10.1016/j.jml.2003.08.004](https://doi.org/10.1016/j.jml.2003.08.004) (cit. on pp. 37, 41).
- Clark, M. A. C., R. D. Boyle (2006). "Computer Science in English High Schools: We Lost the S, Now the C Is Going". In: *Informatics Education – The Bridge between Using and Understanding Computers*. Ed. by D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, R. T. Mittermeir. Vol. 4226. Series Title: Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 83–93. ISBN: 978-3-540-48227-7. DOI: [10.1007/11915355_8](https://doi.org/10.1007/11915355_8) (cit. on p. 212).
- Coe, R., C. Aloisi, S. Higgins, L. E. Major (Oct. 2014). "What makes great teaching?" en-GB. In: (cit. on pp. 40, 197).

- Csibra, G. (Mar. 2003). "Teleological and referential understanding of action in infancy." In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 358.1431, pp. 447–458. ISSN: 0962-8436. DOI: [10.1098/rstb.2002.1235](https://doi.org/10.1098/rstb.2002.1235) (cit. on p. 47).
- Dawkins, R. (1995). *River out of eden : a Darwinian view of life*. eng. New York, NY : Basic Books. ISBN: 978-0-465-01606-8 (cit. on p. 47).
- Dennett, D. C. (1987). *The intentional stance*. en. Cambridge, Mass: MIT Press. ISBN: 978-0-262-04093-8 (cit. on pp. 17, 18, 23, 44, 207).
- Dennett, D. (Jan. 2009). "Intentional Systems Theory". In: *The Oxford Handbook of Philosophy of Mind*. Ed. by A. Beckermann, B. P. McLaughlin, S. Walter. Oxford University Press, pp. 339–350. ISBN: 978-0-19-926261-8. DOI: [10.1093/oxfordhb/9780199262618.003.0020](https://doi.org/10.1093/oxfordhb/9780199262618.003.0020) (cit. on p. 45).
- Dijk, T. A. van, W. Kintsch (1983). *Strategies of discourse comprehension*. New York: Academic Press. ISBN: 978-0-12-712050-8 (cit. on p. 30).
- Dijkstra, E. W. (1982). *Selected Writings on Computing: A personal Perspective*. New York, NY: Springer. ISBN: 978-1-4612-5695-3. DOI: [10.1007/978-1-4612-5695-3](https://doi.org/10.1007/978-1-4612-5695-3) (cit. on p. 15).
- Du Boulay, B. (Feb. 1986). "Some Difficulties of Learning to Program". EN. In: *Journal of Educational Computing Research* 2.1, pp. 57–73. ISSN: 0735-6331. DOI: [10.2190/3LFX-9RRF-67T8-UVK9](https://doi.org/10.2190/3LFX-9RRF-67T8-UVK9) (cit. on p. 195).
- Engbring, D. (2003). "Informatik im Herstellungs- und Nutzungskontext: Ein technikbezogener Zugang zur fachübergreifenden Lehre". PhD thesis. Universität Paderborn (cit. on p. 25).
- Fichtel, L., M. Spliethöver, E. Hüllermeier, P. Jimenez, N. Klowait, S. Kopp, A.-C. Ngonga Ngomo, A. Robrecht, I. Scharlau, L. Terflöth, A.-L. Vollmer, H. Wachsmuth (Aug. 2025). "Investigating Co-Constructive Behavior of Large Language Models in Explanation Dialogues". In: *Proceedings of the 26th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. Ed. by F. Béchet, F. Lefèvre, N. Asher, S. Kim, T. Merlin. Avignon, France: Association for Computational Linguistics, pp. 1–20 (cit. on p. 215).
- Fisher, J., A. Robrecht, S. Kopp, K. Rohlfing (Aug. 2023). *Exploring the Semantic Dialogue Patterns of Explanations - a Case Study of Game Explanations* (cit. on pp. 24, 53, 197, 212).
- Fisher, J. B., A. S. Robrecht, S. Kopp, K. J. Rohlfing (n.d.). "Exploring the Semantic Dialogue Patterns of Explanations – a Case Study of Game Explanations". en. In: () (cit. on p. 294).
- Frederik, I., W. Sonneveld, M. J. De Vries (Aug. 2011). "Teaching and learning the nature of technical artifacts". en. In: *International Journal of Technology and Design Education* 21.3, pp. 277–290. ISSN: 0957-7572, 1573-1804. DOI: [10.1007/s10798-010-9119-3](https://doi.org/10.1007/s10798-010-9119-3) (cit. on pp. 20, 196).
- Gredebäck, G., A. Melinder (Oct. 2011). "Teleological Reasoning in 4-Month-Old Infants: Pupil Dilations and Contextual Constraints". In: *PLoS ONE* 6.10, e26487. ISSN: 1932-6203. DOI: [10.1371/journal.pone.0026487](https://doi.org/10.1371/journal.pone.0026487) (cit. on p. 47).

- Hattie, J. (Mar. 2012). *Visible Learning for Teachers: Maximizing Impact on Learning*. London: Routledge. ISBN: 978-0-203-18152-2. DOI: [10.4324/9780203181522](https://doi.org/10.4324/9780203181522) (cit. on pp. 40, 42, 43).
- Helfferrich, C. (2014). "Leitfaden- und Experteninterviews". de. In: *Handbuch Methoden der empirischen Sozialforschung*. Ed. by N. Baur, J. Blasius. Wiesbaden: Springer Fachmedien, pp. 559–574. ISBN: 978-3-531-18939-0. DOI: [10.1007/978-3-531-18939-0_39](https://doi.org/10.1007/978-3-531-18939-0_39) (cit. on p. 265).
- Hempel, C. G., P. Oppenheim (Apr. 1948). "Studies in the Logic of Explanation". en. In: *Philosophy of Science* 15.2, pp. 135–175. ISSN: 0031-8248, 1539-767X. DOI: [10.1086/286983](https://doi.org/10.1086/286983) (cit. on pp. 23, 35).
- Hernik, M., G. Csibra (Feb. 2015). "Infants learn enduring functions of novel tools from action demonstrations". In: *Journal of experimental child psychology* 130, pp. 176–192. ISSN: 0022-0965. DOI: [10.1016/j.jecp.2014.10.004](https://doi.org/10.1016/j.jecp.2014.10.004) (cit. on p. 47).
- Höper, L., C. Schulte (Dec. 2023). "The data awareness framework as part of data literacies in K-12 education". In: *Information and Learning Sciences* 125.7-8, pp. 491–512. ISSN: 2398-5348. DOI: [10.1108/ILS-06-2023-0075](https://doi.org/10.1108/ILS-06-2023-0075) (cit. on pp. 20, 196, 213).
- Jefferson, G. (Jan. 1984). "Notes on a systematic deployment of the acknowledgement tokens "Yeah"; and "Mm Hm";" in: *Paper in Linguistics* 17.2, pp. 197–216. ISSN: 0031-1251. DOI: [10.1080/08351818409389201](https://doi.org/10.1080/08351818409389201) (cit. on p. 37).
- Jia, K., T. H. M. Leung, N. Y. I. Cheung, Y. Li, J. Yu (June 2025). "Developing a Holistic AI Literacy Framework for Children". In: *ACM Trans. Comput. Educ.* 25.2, 21:1–21:30. DOI: [10.1145/3727986](https://doi.org/10.1145/3727986) (cit. on pp. 20, 212).
- Kallia, M. (Aug. 2023). "The Search for Meaning: Inferential Strategic Reading Comprehension in Programming". en. In: *Proceedings of the 2023 ACM Conference on International Computing Education Research V.1*. Chicago IL USA: ACM, pp. 1–14. ISBN: 978-1-4503-9976-0. DOI: [10.1145/3568813.3600135](https://doi.org/10.1145/3568813.3600135) (cit. on p. 32).
- Keil, F. C. (1994). "The birth and nurturance of concepts by domains: The origins of concepts of living things". In: *Mapping the mind: Domain specificity in cognition and culture*. New York, NY, US: Cambridge University Press, pp. 234–254. DOI: [10.1017/CB09780511752902.010](https://doi.org/10.1017/CB09780511752902.010) (cit. on pp. 17, 44, 207).
- (2006). "Explanation and Understanding". In: *Annual Review of Psychology* 57.1, pp. 227–254. DOI: [10.1146/annurev.psych.57.102904.190100](https://doi.org/10.1146/annurev.psych.57.102904.190100) (cit. on pp. 15, 18, 23, 35, 36, 40, 41, 43, 44, 47).
- Kelemen, D. (Apr. 1999a). "The scope of teleological thinking in preschool children". eng. In: *Cognition* 70.3, pp. 241–272. ISSN: 0010-0277. DOI: [10.1016/S0010-0277\(99\)00010-4](https://doi.org/10.1016/S0010-0277(99)00010-4) (cit. on p. 47).
- (Nov. 1999b). "Why are rocks pointy? Children's preference for teleological explanations of the natural world". eng. In: *Developmental Psychology* 35.6, pp. 1440–1452. ISSN: 0012-1649. DOI: [10.1037//0012-1649.35.6.1440](https://doi.org/10.1037//0012-1649.35.6.1440) (cit. on p. 18).
- Kelemen, D. (Dec. 1999c). "Function, goals and intention: children's teleological reasoning about objects". en. In: *Trends in Cognitive Sciences* 3.12, pp. 461–468. ISSN: 13646613. DOI: [10.1016/S1364-6613\(99\)01402-3](https://doi.org/10.1016/S1364-6613(99)01402-3) (cit. on pp. 35, 44, 45).

- Kelemen, D., E. Rosset (Apr. 2009). "The Human Function Compunction: Teleological explanation in adults". In: *Cognition* 111.1, pp. 138–143. ISSN: 0010-0277. DOI: [10.1016/j.cognition.2009.01.001](https://doi.org/10.1016/j.cognition.2009.01.001) (cit. on p. 47).
- Kelemen, D., J. Rottman, R. Seston (2013). "Professional physical scientists display tenacious teleological tendencies: Purpose-based reasoning as a cognitive default." en. In: *Journal of Experimental Psychology: General* 142.4, pp. 1074–1083. ISSN: 1939-2222, 0096-3445. DOI: [10.1037/a0030399](https://doi.org/10.1037/a0030399) (cit. on pp. 17, 47, 48, 192, 207).
- Kern, F. (July 2011). "Der Erwerb kommunikativer Praktiken und Formen – Am Beispiel des Erzählens und Erklärens". de. In: *Textsorten, Handlungsmuster, Oberflächen*. Ed. by S. Habscheid. DE GRUYTER, pp. 231–254. ISBN: 978-3-11-018902-5. DOI: [10.1515/9783110229301.231](https://doi.org/10.1515/9783110229301.231) (cit. on p. 36).
- Kern, F., U. Boden, A. Nemeth, S. Koutalidis, O. Abramov, S. Kopp, K. J. Rohlfing (2024). "Preschool children's discourse competence in different genres and how it relates to iconic gestures". eng. In: *Journal of Child Language* 51.3, pp. 1–25. ISSN: 0305-0009. DOI: [10.1017/S030500092300065X](https://doi.org/10.1017/S030500092300065X) (cit. on pp. 21, 42, 210).
- Keysar, B., D. J. Barr, W. S. Horton (1998). "The egocentric basis of language use: Insights from a processing approach". In: *Current Directions in Psychological Science* 7.2, pp. 46–50. ISSN: 1467-8721. DOI: [10.1111/1467-8721.ep13175613](https://doi.org/10.1111/1467-8721.ep13175613) (cit. on pp. 40, 43).
- Kintsch, W. (Jan. 1991). "The Role of Knowledge in Discourse Comprehension: A Construction-Integration Model". In: *Advances in Psychology*. Ed. by G. E. Steinhilber, P. A. Vroom. Vol. 79. Text and Text Processing. North-Holland, pp. 107–153. DOI: [10.1016/S0166-4115\(08\)61551-4](https://doi.org/10.1016/S0166-4115(08)61551-4) (cit. on p. 30).
- Klein, J. (July 2008). "Erklären und Argumentieren als interaktive Gesprächsstrukturen". en. In: 2. *Halbband Text- und Gesprächslinguistik 2. Halbband*. De Gruyter Mouton, pp. 1309–1329. ISBN: 978-3-11-019422-7. DOI: [10.1515/9783110169188.2.18.1309](https://doi.org/10.1515/9783110169188.2.18.1309) (cit. on p. 24).
- Knight, E. H. (H. (1882). *Knight's American mechanical dictionary : a description of tools, instruments, machines, processes and engineering, history of inventions, general technological vocabulary ; and digest of mechanical appliances in science and the arts - Volume 2*. eng. Vol. 2. Boston : Houghton, Mifflin and Company (cit. on pp. 28, 29).
- Kotthoff, H. (Jan. 2009). "Erklärende Aktivitätstypen in Alltags-und Unterrichtskontexten Einleitung". In: *Spreckels* 2009 (cit. on p. 24).
- Kroes, P. (1998). "Technological Explanations". In: *Phil & Tech* 3.3, pp. 124–134. ISSN: 1091-8264. DOI: [10.5840/techne19983325](https://doi.org/10.5840/techne19983325) (cit. on pp. 15–18, 23, 26, 33, 91).
- (May 2002). "Design methodology and the nature of technical artefacts". en. In: *Design Studies*. Philosophy of Design 23.3, pp. 287–302. ISSN: 0142-694X. DOI: [10.1016/S0142-694X\(01\)00039-4](https://doi.org/10.1016/S0142-694X(01)00039-4) (cit. on p. 33).
- (Mar. 2006). "Coherence of structural and functional descriptions of technical artefacts". In: *Studies in History and Philosophy of Science Part A*. The dual nature of technical artefacts 37.1, pp. 137–151. ISSN: 0039-3681. DOI: [10.1016/j.shpsa.2005.12.015](https://doi.org/10.1016/j.shpsa.2005.12.015) (cit. on pp. 16–18, 25, 29, 30, 33, 38).

- Kroes, P. (2010). "Engineering and the dual nature of technical artefacts". In: *Cambridge journal of economics* 34.1, pp. 51–62. DOI: [10.1093/cje/bep019](https://doi.org/10.1093/cje/bep019) (cit. on pp. 16, 24, 25, 33, 188, 207).
- Kroes, P., A. Meijers (2002). "The Dual Nature of Technical Artifacts – presentation of a new research programme". en. In: p. 5 (cit. on p. 25).
- Landis, J. R., G. G. Koch (1977). "The Measurement of Observer Agreement for Categorical Data". In: *Biometrics* 33.1, pp. 159–174. ISSN: 0006-341X. DOI: [10.2307/2529310](https://doi.org/10.2307/2529310) (cit. on p. 188).
- Lazarov, S., M. Schaffer, V. Gladow, H. Buschmeier, H. Buhl, A. Grimminger (Mar. 2026). *Retrospective Video Recall for Analyzing Cognitive Processes in Naturalistic Explanations*. DOI: [10.31219/osf.io/u24kz_v2](https://doi.org/10.31219/osf.io/u24kz_v2) (cit. on pp. 38, 210).
- Leung, A., D. Yurovsky, R. D. Hawkins (2025). "Parents spontaneously scaffold the formation of conversational pacts with their children". In: *Child Development* 96.2, pp. 546–561. ISSN: 0009-3920. DOI: [10.1111/cdev.14186](https://doi.org/10.1111/cdev.14186) (cit. on p. 42).
- Liquin, E. G., T. Lombrozo (Dec. 2018). "Structure-function fit underlies the evaluation of teleological explanations". In: *Cognitive Psychology* 107, pp. 22–43. ISSN: 0010-0285. DOI: [10.1016/j.cogpsych.2018.09.001](https://doi.org/10.1016/j.cogpsych.2018.09.001) (cit. on p. 48).
- Lohmer, V., P. D. F. Kern (2024). "The role of interactive gestures in explanatory interactions". English. In: *International Multimodal Communication Symposiums (MMSYM)*. Frankfurt am Main, Germany, pp. 157–158 (cit. on pp. 53, 121, 191, 210, 212, 214).
- Lohmer, V., L. Terfloth, P. D. F. Kern (2023). "Explaining the Technical Artifact Quartet: How Gestures are used in Everyday Explanations". English. In: *1st International Multimodal Communication Symposium*. Barcelona Spain, pp. 133–134 (cit. on pp. 210, 216).
- Lombrozo, T. (Nov. 2007). "Simplicity and probability in causal explanation". English (US). In: *Cognitive Psychology* 55.3, pp. 232–257. ISSN: 0010-0285. DOI: [10.1016/j.cogpsych.2006.09.006](https://doi.org/10.1016/j.cogpsych.2006.09.006) (cit. on p. 36).
- (Feb. 2009). "Explanation and categorization: How "why?" informs "what?"". en. In: *Cognition* 110.2, pp. 248–253. ISSN: 00100277. DOI: [10.1016/j.cognition.2008.10.007](https://doi.org/10.1016/j.cognition.2008.10.007) (cit. on p. 46).
- Lombrozo, T., S. Carey (Mar. 2006). "Functional explanation and the function of explanation". en. In: *Cognition* 99.2, pp. 167–204. ISSN: 0010-0277. DOI: [10.1016/j.cognition.2004.12.009](https://doi.org/10.1016/j.cognition.2004.12.009) (cit. on pp. 17, 44, 46).
- Lombrozo, T., N. Z. Gwynne (2014). "Explanation and inference: mechanistic and functional explanations guide property generalization". In: *Frontiers in Human Neuroscience* 8. ISSN: 1662-5161. DOI: [10.3389/fnhum.2014.00700](https://doi.org/10.3389/fnhum.2014.00700) (cit. on p. 18).
- Lombrozo, T., D. Wilkenfeld (Aug. 2019). "Mechanistic versus Functional Understanding". en. In: *Varieties of Understanding: New Perspectives from Philosophy, Psychology, and Theology*. Ed. by S. R. Grimm. Oxford University Press. ISBN: 978-0-19-086098-1. DOI: [10.1093/oso/9780190860974.003.0011](https://doi.org/10.1093/oso/9780190860974.003.0011) (cit. on pp. 17, 18, 46, 47, 192, 207).

- Lonati, V., A. Brodnik, T. Bell, A. P. Csizmadia, L. De Mol, H. Hickman, T. Keane, C. Mirolo, M. Monga (Dec. 2022). "What We Talk About When We Talk About Programs". en. In: *Proceedings of the 2022 Working Group Reports on Innovation and Technology in Computer Science Education*. Dublin Ireland: ACM, pp. 117–164. ISBN: 979-8-4007-0010-1. DOI: [10.1145/3571785.3574125](https://doi.org/10.1145/3571785.3574125) (cit. on p. 18).
- Madumal, P., T. Miller, L. Sonenberg, F. Vetere (2019). "A Grounded Interaction Protocol for Explainable Artificial Intelligence". In: *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*. AAMAS '19. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems, pp. 1033–1041. ISBN: 978-1-4503-6309-9 (cit. on p. 199).
- McCarthy, A. M., F. C. Keil (Sept. 2023). "A right way to explain? Function, mechanism, and the order of explanations". en. In: *Cognition* 238, p. 105494. ISSN: 00100277. DOI: [10.1016/j.cognition.2023.105494](https://doi.org/10.1016/j.cognition.2023.105494) (cit. on pp. 47, 49, 192, 196, 197, 211).
- Melhart, D., A. Liapis, G. N. Yannakakis (Sept. 2019). "PAGAN: Video Affect Annotation Made Easy". In: *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*. ISSN: 2156-8111, pp. 130–136. DOI: [10.1109/ACII.2019.8925434](https://doi.org/10.1109/ACII.2019.8925434) (cit. on p. 210).
- Mercer, N. (May 2002). *Words and Minds: How We Use Language to Think Together*. London: Routledge. ISBN: 978-0-203-46498-4. DOI: [10.4324/9780203464984](https://doi.org/10.4324/9780203464984) (cit. on pp. 17, 38, 39, 51).
- Mermelshtine, R. (June 2017). "Parent-child learning interactions: A review of the literature on scaffolding". eng. In: *The British Journal of Educational Psychology* 87.2, pp. 241–254. ISSN: 2044-8279. DOI: [10.1111/bjep.12147](https://doi.org/10.1111/bjep.12147) (cit. on pp. 38, 39, 42).
- Miller, T. (Feb. 2019). "Explanation in artificial intelligence: Insights from the social sciences". en. In: *Artificial Intelligence* 267, pp. 1–38. ISSN: 0004-3702. DOI: [10.1016/j.artint.2018.07.007](https://doi.org/10.1016/j.artint.2018.07.007) (cit. on pp. 20, 35–37, 198).
- Miller, T., P. Howe, L. Sonenberg (Dec. 2017). "Explainable AI: Beware of Inmates Running the Asylum Or: How I Learnt to Stop Worrying and Love the Social and Behavioural Sciences". In: DOI: [10.48550/arXiv.1712.00547](https://doi.org/10.48550/arXiv.1712.00547) (cit. on p. 198).
- Mindlin, D., P. Cimiano (Feb. 2025). "Dialogue-Based XAI Approaches: Exploring the Continuum from Static Explanations to Free-Form Chat Interfaces". English. In: IEEE Computer Society, pp. 1–8. ISBN: 979-8-3315-2426-5. DOI: [10.1109/ICSC64641.2025.00015](https://doi.org/10.1109/ICSC64641.2025.00015) (cit. on p. 212).
- Mindlin, D., A. S. Robrecht, M. Morasch, P. Cimiano (2024). "Measuring User Understanding in Dialogue-based XAI Systems". In: Version Number: 2. arXiv. DOI: [10.48550/ARXIV.2408.06960](https://doi.org/10.48550/ARXIV.2408.06960) (cit. on p. 212).
- Miyake, N. (Apr. 1986). "Constructive Interaction and the Iterative Process of Understanding". en. In: *Cognitive Science* 10.2, pp. 151–177. ISSN: 0364-0213, 1551-6709. DOI: [10.1207/s15516709cog1002_2](https://doi.org/10.1207/s15516709cog1002_2) (cit. on pp. 18, 31, 33, 34).

- Nickerson, R. S. (1999). "How we know—and sometimes misjudge—what others know: Imputing one's own knowledge to others". In: *Psychological Bulletin* 125.6, pp. 737–759. ISSN: 1939-1455. DOI: [10.1037/0033-2909.125.6.737](https://doi.org/10.1037/0033-2909.125.6.737) (cit. on pp. 41, 43).
- Nückles, M., J. Wittwer, A. Renkl (Jan. 2006). "Information About a Layperson's Knowledge Supports Experts in Giving Effective and Efficient Online Advice to Laypersons." In: *Journal of experimental psychology. Applied* 11, pp. 219–36. DOI: [10.1037/1076-898X.11.4.219](https://doi.org/10.1037/1076-898X.11.4.219) (cit. on p. 42).
- Park, J. S.-Y. (Jan. 2002). "Cognitive and interactional motivations for the intonation unit". en. In: *Studies in Language. International Journal sponsored by the Foundation "Foundations of Language"* 26.3, pp. 637–680. ISSN: 0378-4177, 1569-9978. DOI: [10.1075/sl.26.3.07par](https://doi.org/10.1075/sl.26.3.07par) (cit. on pp. 54, 58, 202).
- Pasquale, F. (2016). *The black box society: the secret algorithms that control money and information*. en. First Harvard University Press paperback edition. Cambridge, Massachusetts London, England: Harvard University Press. ISBN: 978-0-674-97084-7 (cit. on pp. 19, 212).
- Pennington, N. (1987). "Comprehension strategies in programming". In: *Empirical studies of programmers: second workshop*. Ed. by G. M. Olson, S. Sheppard, E. Soloway. Ablex Publishing Corp., pp. 100–113. ISBN: 0-89391-461-4. DOI: [10.5555/54968.54975](https://doi.org/10.5555/54968.54975) (cit. on p. 31).
- Perlman, M. (2004). "The Modern Philosophical Resurrection of Teleology". In: *The Monist* 87.1, pp. 3–51. ISSN: 0026-9662. DOI: [10.5840/monist20048711](https://doi.org/10.5840/monist20048711) (cit. on p. 29).
- Peters, T. M., R. W. Visser (2023). "The Importance of Distrust in AI". en. In: *Explainable Artificial Intelligence*. Ed. by L. Longo. Cham: Springer Nature Switzerland, pp. 301–317. ISBN: 978-3-031-44070-0. DOI: [10.1007/978-3-031-44070-0_15](https://doi.org/10.1007/978-3-031-44070-0_15) (cit. on pp. 19, 20, 213).
- Pol, J. E. van de (2012). *Scaffolding in teacher-student interaction: exploring, measuring, promoting and evaluating scaffolding*. en. ISBN: 978-94-6191-399-9 (cit. on pp. 41, 43).
- Popper, K. (1935). *Logik der Forschung - ZUR ERKENNTNISTHEORIE DER MODERNEN NATURWISSENSCHAFT*. de. Vienna: Springer Vienna. ISBN: 978-3-7091-4177-9. DOI: <https://doi.org/10.1007/978-3-7091-4177-9> (cit. on p. 35).
- Qian, Y., J. Lehman (2017). "Students' Misconceptions and Other Difficulties in Introductory Programming: A Literature Review". In: *ACM Trans. Comput. Educ.* 18.1, 1:1–1:24. DOI: [10.1145/3077618](https://doi.org/10.1145/3077618) (cit. on p. 195).
- Quasthoff, U., V. Heller, M. Morek (2017). "On the sequential organization and genre-orientation of discourse units in interaction: An analytic framework". In: 19.1, pp. 84–110. ISSN: 14614456. DOI: [10.1177/1461445616683596](https://doi.org/10.1177/1461445616683596) (cit. on p. 18).
- Register, Y., A. J. Ko (Aug. 2020). "Learning Machine Learning with Personal Data Helps Stakeholders Ground Advocacy Arguments in Model Mechanics". en. In: *Proceedings of the 2020 ACM Conference on International Computing Education Research*. Virtual Event New Zealand: ACM, pp. 67–78. ISBN: 978-1-4503-7092-9. DOI: [10.1145/3372782.3406252](https://doi.org/10.1145/3372782.3406252) (cit. on pp. 49, 199).

- Reusser, K., C. Pauli (Jan. 2015). "Co-constructivism in Educational Theory and Practice". In: *International Encyclopedia of the Social & Behavioral Sciences (Second Edition)*. Ed. by J. D. Wright. Oxford: Elsevier, pp. 913–917. ISBN: 978-0-08-097087-5. DOI: [10.1016/B978-0-08-097086-8.92026-9](https://doi.org/10.1016/B978-0-08-097086-8.92026-9) (cit. on pp. 17, 36, 37, 40, 41, 43, 190).
- Rey, G. (Aug. 2025). *Philosophy of mind - Thoughts, Attitudes, Consciousness | Britannica*. en (cit. on p. 16).
- Ridder, J. de (Mar. 2006). "Mechanistic artefact explanation". en. In: *Studies in History and Philosophy of Science Part A. The dual nature of technical artefacts* 37.1, pp. 81–96. ISSN: 0039-3681. DOI: [10.1016/j.shpsa.2005.12.009](https://doi.org/10.1016/j.shpsa.2005.12.009) (cit. on pp. 16, 17, 33).
- (2007). *Reconstructing design, explaining artifacts: philosophical reflections on the design and explanation of technical artifacts*. ISBN: 978-90-90-21903-5 (cit. on pp. 16, 33).
- Robrecht, A. S., S. Kopp (2023). "SNAPE: A Sequential Non-Stationary Decision Process Model for Adaptive Explanation Generation." In: *Proceedings of the 15th International Conference on Agents and Artificial Intelligence (ICAART 2023)*. Lissabon, Portugal: SCITEPRESS – Science and Technology Publications, pp. 48–58. DOI: [10.5220/0011671300003393](https://doi.org/10.5220/0011671300003393) (cit. on p. 212).
- Robrecht-Hilbig, A. S. (2026). "Adaptive Explanations as Co-Constructed Processes. Modeling a Rational Explainer Through the Interaction of Dynamic Partner Models and Non-Stationary Decision Making". eng. In: (cit. on p. 212).
- Rohlfing, K. J., P. Cimiano, I. Scharlau, T. Matzner, H. M. Buhl, H. Buschmeier, E. Esposito, A. Grimminger, B. Hammer, R. Hab-Umbach, I. Horwath, E. Hullermeier, F. Kern, S. Kopp, K. Thommes, A.-C. Ngonga Ngomo, C. Schulte, H. Wachsmuth, P. Wagner, B. Wrede (Sept. 2021). "Explanation as a Social Practice: Toward a Conceptual Framework for the Social Design of AI Systems". en. In: *IEEE Transactions on Cognitive and Developmental Systems* 13.3, pp. 717–728. ISSN: 2379-8920, 2379-8939. DOI: [10.1109/TCDS.2020.3044366](https://doi.org/10.1109/TCDS.2020.3044366) (cit. on pp. 15, 17, 18, 20, 25, 36, 37, 39, 42, 190, 198, 199).
- Rohlfing, K. J., K. Främling, B. Lim, S. Alpsancar, K. Thommes, eds. (2026). *Social Explainable AI: Communications of NII Shonan Meetings*. en. rohlfling_social_2026. Singapore: Springer Nature. ISBN: 978-981-96-5289-7. DOI: [10.1007/978-981-96-5290-7](https://doi.org/10.1007/978-981-96-5290-7) (cit. on p. 213).
- Roscoe, R. D. (May 2014). "Self-monitoring and knowledge-building in learning by teaching". en. In: *Instructional Science* 42.3, pp. 327–351. ISSN: 0020-4277, 1573-1952. DOI: [10.1007/s11251-013-9283-4](https://doi.org/10.1007/s11251-013-9283-4) (cit. on p. 37).
- Rozenblit, L., F. Keil (Sept. 2002). "The misunderstood limits of folk science: an illusion of explanatory depth". en. In: *Cognitive Science* 26.5, pp. 521–562. ISSN: 0364-0213, 1551-6709. DOI: [10.1207/s15516709cog2605_1](https://doi.org/10.1207/s15516709cog2605_1) (cit. on p. 15).
- Salmon, W. C. (1990). *Four Decades of Scientific Explanation*. University of Pittsburgh Press. ISBN: 978-0-8229-5926-7. DOI: [10.2307/j.ctt5vkdm7](https://doi.org/10.2307/j.ctt5vkdm7) (cit. on p. 35).

- Schaffer, M. E., H. M. Buhl (May 2025). *Do Explainers Perceive the Explainees' Specific Interests and Knowledge?* DOI: [10.31219/osf.io/tph6d_v1](https://doi.org/10.31219/osf.io/tph6d_v1) (cit. on pp. 189, 190, 199, 210, 212).
- Schaffer, M. E., L. Terfloth, C. Schulte, H. M. Buhl (Nov. 2024). *Explainers' Mental Representations of Explainees' Needs in Everyday Explanations*. DOI: [10.48550/arXiv.2411.08514](https://doi.org/10.48550/arXiv.2411.08514) (cit. on pp. 37, 38, 53, 120, 216).
- Schegloff, E. A. (2007). *Sequence Organization in Interaction: A Primer in Conversation Analysis*. Vol. 1. Cambridge: Cambridge University Press. ISBN: 978-0-521-82572-6. DOI: [10.1017/CB09780511791208](https://doi.org/10.1017/CB09780511791208) (cit. on p. 37).
- Schulte, C. (2008a). "Block Model - an Educational Model of Program Comprehension as a Tool for a Scholarly Approach to Teaching". In: *Proceeding of the Fourth international Workshop on Computing Education Research*. ICER '08. Sydney, Australia: ACM, pp. 149–160. ISBN: 978-1-60558-216-0. DOI: [10.1145/1404520.1404535](https://doi.org/10.1145/1404520.1404535) (cit. on pp. 31, 198).
- (2008b). "Duality Reconstruction - Teaching digital artifacts from a socio-technical perspective". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 5090 LNCS, pp. 110–121. ISSN: 03029743. DOI: [10.1007/978-3-540-69924-8_10](https://doi.org/10.1007/978-3-540-69924-8_10) (cit. on pp. 15, 16).
- Schulte, C., L. Budde (Nov. 2018). "A Framework for Computing Education: Hybrid Interaction System: The need for a bigger picture in computing education." In: *18th Koli Calling International Conference on Computing Education Research (Koli Calling '18)*. Vol. 18. Koli, Finland: ACM, p. 10. DOI: [10.1145/3279720.3279733](https://doi.org/10.1145/3279720.3279733) (cit. on pp. 26, 33, 43).
- Schulte, C., T. Clear, A. Taherkhani, T. Busjahn, J. H. Paterson (June 2010). "An introduction to program comprehension for computer science educators". In: *Proceedings of the 2010 ITiCSE working group reports*. ITiCSE-WGR '10. New York, NY, USA: Association for Computing Machinery, pp. 65–86. ISBN: 978-1-4503-0677-5. DOI: [10.1145/1971681.1971687](https://doi.org/10.1145/1971681.1971687) (cit. on pp. 31, 33, 200).
- Selting, M., P. Auer, D. Barth-Weingarten, J. Bergmann, P. Bergmann, K. Birkner, E. Couper-Kuhlen, A. Deppermann, P. Gilles, S. Günthner, M. Hartung, F. Kern, C. Mertzluft, C. Meyer, M. Morek, F. Oberzaucher, J. Peters, U. Quasthoff, W. Schütte, S. Uhmann (Jan. 2009). "Gesprächsanalytisches Transkriptionssystem (GAT 2)". In: *Gesprächsforsch* 10, pp. 152–183 (cit. on p. 57).
- Sentance, S., E. Barendsen, C. Schulte (2018). *Computer Science Education: Perspectives on Teaching and Learning in School*. Englisch. London ; New York: BLOOMSBURY ACADEMIC. ISBN: 978-1-350-05710-4 (cit. on p. 196).
- Sentance, S., J. Waite (Aug. 2021). "Teachers' Perspectives on Talk in the Programming Classroom : Language as a Mediator". en. In: *Proceedings of the 17th ACM Conference on International Computing Education Research*. Virtual Event USA: ACM, pp. 266–280. ISBN: 978-1-4503-8326-4. DOI: [10.1145/3446871.3469751](https://doi.org/10.1145/3446871.3469751) (cit. on p. 39).

- Shulman, L. (Feb. 1986). "Those Who Understand: Knowledge Growth in Teaching". EN. In: *Educational Researcher* 15.2, pp. 4–14. ISSN: 0013-189X. DOI: [10.3102/0013189X015002004](https://doi.org/10.3102/0013189X015002004) (cit. on pp. 40, 197).
- (Apr. 1987). "Knowledge and Teaching: Foundations of the New Reform". en. In: *Harvard Educational Review* 57.1, pp. 1–23. ISSN: 0017-8055, 1943-5045. DOI: [10.17763/haer.57.1.j463w79r56455411](https://doi.org/10.17763/haer.57.1.j463w79r56455411) (cit. on p. 40).
- Simpson, H. E. (2016). "The role of intonation units in memory for spoken English". English. Dissertation. Santa Barbara, Calif.: University of California, Santa Barbara. ISBN: 978-1-369-14651-6 (cit. on pp. 54, 58, 202).
- Sokol, K., P. Flach (June 2020). "One Explanation Does Not Fit All". en. In: *KI - Künstliche Intelligenz* 34.2, pp. 235–250. ISSN: 1610-1987. DOI: [10.1007/s13218-020-00637-y](https://doi.org/10.1007/s13218-020-00637-y) (cit. on pp. 19, 20, 198, 199).
- Soloway, E. (Sept. 1986). "Learning to program = learning to construct mechanisms and explanations". In: *Communications of the ACM* 29.9, pp. 850–858. ISSN: 0001-0782. DOI: [10.1145/6592.6594](https://doi.org/10.1145/6592.6594) (cit. on p. 30).
- Sorva, J. (June 2013). "Notional machines and introductory programming education". en. In: *ACM Transactions on Computing Education* 13.2, pp. 1–31. ISSN: 19466226. DOI: [10.1145/2483710.2483713](https://doi.org/10.1145/2483710.2483713) (cit. on p. 195).
- Spohrer, J. C., E. Soloway, E. Pope (June 1985). "A Goal/Plan Analysis of Buggy Pascal Programs". en. In: *Human-Computer Interaction* 1.2, pp. 163–207. ISSN: 0737-0024, 1532-7051. DOI: [10.1207/s15327051hci0102_4](https://doi.org/10.1207/s15327051hci0102_4) (cit. on pp. 30, 50).
- Tedre, M., M. Apiola (2013). "Three computing traditions in school computing education". In: *Improving computer science education*. Ed. by D. M. Kadijevich, C. Angeli, C. Schulte. 00013. Routledge New York, NY and London, pp. 100–116 (cit. on pp. 211, 212).
- Tedre, M., E. Sutinen (2008). "Three traditions of computing: what educators should know". In: *Computer Science Education* 18.3. 00044 tex.ids: threetradi-tionTedre, pp. 153–170. DOI: [10.1080/08993400802332332](https://doi.org/10.1080/08993400802332332) (cit. on p. 196).
- Terfloth, L., L. Budde, C. Schulte (Nov. 2020). *Combining Ideas and Artifacts: an Interaction-Focused View on Computing Education Using a Cybersecurity Example*. Pages: 5. DOI: [10.1145/3428029.3428052](https://doi.org/10.1145/3428029.3428052) (cit. on p. 216).
- Terfloth, L., H. M. Buhl, V. Lohmer, F. Kern, M. Schaffer, C. Schulte (n.d.). "Navigating the Dual Nature: Do Explainers Adapt to Examinee Interests When Explaining Technical Artifacts". en. In: *International Journal of Technology and Design Education* () (cit. on pp. 119, 215).
- Terfloth, L., H. M. Buhl, V. Lohmer, M. Schaffer, F. Kern, C. Schulte (Mar. 2026). *Bridging the Dual Nature: How Integrated Explanations Enhance Understanding of Technical Artifacts*. arXiv:2603.24325 [cs]. DOI: [10.48550/arXiv.2603.24325](https://doi.org/10.48550/arXiv.2603.24325) (cit. on pp. 151, 215).
- Terfloth, L., V. Lohmer, F. Kern, C. Schulte (Apr. 2025). "Transcription in Computing Education Research: A Review and Recommendations". en. In: *Informatics in Education*. ISSN: 1648-5831, 2335-8971. DOI: [10.15388/infedu.2025.09](https://doi.org/10.15388/infedu.2025.09) (cit. on pp. 57, 215).

- Terfloeth, L., M. Schaffer (Mar. 2025). “Quarto Understanding”. en-US. In: doi: [0.17605/OSF.IO/W39DC](https://doi.org/10.17605/OSF.IO/W39DC) (cit. on p. 265).
- Terfloeth, L., M. Schaffer, H. M. Buhl, C. Schulte (2023). “Adding Why to What? Analyses of an Everyday Explanation”. en. In: *Explainable Artificial Intelligence*. Ed. by L. Longo. Vol. 1903. Cham: Springer Nature, pp. 256–279. ISBN: 978-3-031-44070-0. DOI: [10.1007/978-3-031-44070-0_13](https://doi.org/10.1007/978-3-031-44070-0_13) (cit. on pp. 91, 215).
- Vaesen, K. (Mar. 2011). “The functional bias of the dual nature of technical artefacts program”. en. In: *Studies in History and Philosophy of Science Part A* 42.1, pp. 190–197. ISSN: 00393681. DOI: [10.1016/j.shpsa.2010.11.001](https://doi.org/10.1016/j.shpsa.2010.11.001) (cit. on p. 25).
- Vasilyeva, N., D. Wilkenfeld, T. Lombrozo (2015). “Goals affect the perceived quality of explanations”. In: *Proceedings of the annual meeting of the cognitive science society*. Vol. 37 (cit. on p. 46).
- (Oct. 2017). “Contextual utility affects the perceived quality of explanations”. en. In: *Psychonomic Bulletin & Review* 24.5, pp. 1436–1450. ISSN: 1531-5320. DOI: [10.3758/s13423-017-1275-y](https://doi.org/10.3758/s13423-017-1275-y) (cit. on p. 48).
- Vermaas, P., P. Kroes, I. van de Poel, M. Franssen, W. Houkes (Jan. 2011). “A Philosophy of Technology: From Technical Artefacts to Sociotechnical Systems”. en. In: *Synthesis Lectures on Engineers, Technology and Society* 6.1, pp. 1–134. ISSN: 1933-3633, 1933-3641. DOI: [10.2200/S00321ED1V01Y201012ETS014](https://doi.org/10.2200/S00321ED1V01Y201012ETS014) (cit. on p. 18).
- Vermaas, P. E., W. Houkes (Mar. 2006). “Technical functions: a drawbridge between the intentional and structural natures of technical artefacts”. en. In: *Studies in History and Philosophy of Science Part A. The dual nature of technical artefacts* 37.1, pp. 5–18. ISSN: 0039-3681. DOI: [10.1016/j.shpsa.2005.12.002](https://doi.org/10.1016/j.shpsa.2005.12.002) (cit. on pp. 16, 188, 207).
- Versteeg, M., F. M. van Blankenstein, H. Putter, P. Steendijk (2019). “Peer instruction improves comprehension and transfer of physiological concepts: a randomized comparison with self-explanation”. In: *Advances in Health Sciences Education* 24.1, pp. 151–165. ISSN: 1382-4996. DOI: [10.1007/s10459-018-9858-6](https://doi.org/10.1007/s10459-018-9858-6) (cit. on p. 41).
- Vygotsky, L. S. (1978). *Mind in Society: Development of Higher Psychological Processes*. Ed. by M. Cole, V. Jolm-Steiner, S. Scribner, E. Souberman. Harvard University Press. ISBN: 978-0-674-57628-5. DOI: [10.2307/j.ctvjf9vz4](https://doi.org/10.2307/j.ctvjf9vz4) (cit. on pp. 38, 39).
- Waa, J. van der, E. Nieuwburg, A. Cremers, M. Neerincx (Feb. 2021). “Evaluating XAI: A comparison of rule-based and example-based explanations”. en. In: *Artificial Intelligence* 291, p. 103404. ISSN: 0004-3702. DOI: [10.1016/j.artint.2020.103404](https://doi.org/10.1016/j.artint.2020.103404) (cit. on p. 199).
- Wachsmuth, H., J. Kiesel, B. Stein (Sept. 2015). “Sentiment Flow - A General Model of Web Review Argumentation”. en. In: *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Ed. by L. Màrquez, C. Callison-Burch, J. Su. Lisbon, Portugal: Association for Computational Linguistics, pp. 601–611. DOI: [10.18653/v1/D15-1072](https://doi.org/10.18653/v1/D15-1072) (cit. on p. 211).
- Wilkenfeld, D. A., T. Lombrozo (Nov. 2015). “Inference to the Best Explanation (IBE) Versus Explaining for the Best Inference (EBI)”. en. In: *Science & Education* 24.9, pp. 1059–1077. ISSN: 1573-1901. DOI: [10.1007/s11191-015-9784-4](https://doi.org/10.1007/s11191-015-9784-4) (cit. on pp. 41, 42, 191, 195).

- Wilkes-Gibbs, D., H. H. Clark (Apr. 1992). "Coordinating beliefs in conversation". In: *Journal of Memory and Language* 31.2, pp. 183–194. ISSN: 0749-596X. DOI: [10.1016/0749-596X\(92\)90010-U](https://doi.org/10.1016/0749-596X(92)90010-U) (cit. on pp. 38, 190).
- Wing, J. M. (Oct. 2008). "Computational thinking and thinking about computing". In: *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences* 366.1881, pp. 3717–3725. ISSN: 1364-503X. DOI: [10.1098/rsta.2008.0118](https://doi.org/10.1098/rsta.2008.0118) (cit. on pp. 20, 212).
- Winkelkemper, F., L. Höper, C. Schulte (June 2024). "ARIadne – An Explanation Model for Digital Artefacts". en. In: *Informatics in Education* 23.2, pp. 479–505. ISSN: 1648-5831, 2335-8971. DOI: [10.15388/infedu.2024.09](https://doi.org/10.15388/infedu.2024.09) (cit. on pp. 15, 16, 19, 25–27, 46, 196, 200, 209).
- Wood, D., J. S. Bruner, G. Ross (Apr. 1976). "The Role of Tutoring in Problem Solving". en. In: *Journal of Child Psychology and Psychiatry* 17.2, pp. 89–100. ISSN: 0021-9630, 1469-7610. DOI: [10.1111/j.1469-7610.1976.tb00381.x](https://doi.org/10.1111/j.1469-7610.1976.tb00381.x) (cit. on pp. 17, 38).

All links were last followed on April 20th, 2026.

Appendix

A Data and Code Availability

All analysis code and visualization scripts are available as Jupyter notebooks in the OSF repository.¹ The repository includes statistical analyses for BB2, BB3, and BB4 as well as the scripts used to produce the figures reported in the individual papers. Each appendix for the building blocks may contain a link to additional OSF repositories.

Three sample transcripts are presented here in print. The whole corpus of coded transcripts underlying BB2–BB4 are stored on the central data infrastructure of the TRR 318 and are available to qualified researchers upon reasonable request, subject to the conditions of the original ethics approval granted by the Paderborn University Ethics Board. Requests should be directed to the author or to the TRR 318 data management office.² Raw audio and video recordings cannot be shared, as participants consented to their use for research purposes within the project but not all to unrestricted distribution. Figure A.1 provides a reference for what the video material looked like across studies.

¹https://osf.io/4nsj8/overview?view_only=e3bbccba0def42939209da4bd926b4cd

²<https://trr318.uni-paderborn.de/>



Figure A.1: Anonymized Screenshot of WP4-P2-VP73 - Full consent was granted for printing this screenshot.

B Coding Manual for Content Coding

The dual-nature content coding manual used across BB2–BB4 can be accessed at https://osf.io/29pzu/overview?view_only=3d3ee33fb3af409c82d8b744138c5efa. It documents the category definitions, coding decision rules, and coding examples for the Architecture and Relevance dimensions as well as the segmentation procedure described in Chapter 3.

C Transcript Examples

The following transcripts illustrate the coding scheme and the interactional dynamics analysed across the building blocks. Colour highlighting indicates the dual-nature coding: **Architecture**, **Relevance**, and **Both**.

Note: The transcripts presented in this appendix were exported from time-based annotations created in ELAN. The character-level alignment of annotation labels (e.g., ARCHITECTURE, RELEVANCE) was derived from these time-based intervals and may therefore not map with perfect accuracy to the corresponding text segments, particularly in passages containing overlapping speech.

C.1 Naturalistic Setting — BB2 (WP2-P2-VP05)

1 ---
2 EX ehm Okay
3 wir beginnen mit dem SPIELfeld
4 wir ham ein SPIELfeld
5 mit sechzehn feldERN
6 EE mhm
7 vier mal VIER
8 EX GENau
9 vier mal VIER
10 EHM
11 die sind zunächst LEER
12 es gibt sechzehn SPIELfiguren
13 die ham verschiedene FORMen
14 EE mhm
15 EX EHM
16 VIERecke
17 KREIse
18 MANCHE kreise sind
19 eh (0.6)
20 ehM (2.1)
21 ja es ist AUSgeschnitten
22 sozuSA[gen]
23 EE [mHm] —
24 EX WEIßte was ich meine
25 also LEER
26 AUSgehöhlt
27 ehm
28 geNAU
29 und DANN äh
30 beginnt das siel SO
31 das ICH
32 eine figur DIR gebe
33 und du DAS

1/6



34 mit der FIGUR sozusagen
35 das erste mal ins (1.4) —
36 EE auf das SPIELfeld
37 EX geNAU
38 auf das <:-> SPIELfeld packen
39 "
40 oh GOTT
41 ((lacht))
42 äh geNAU
43 UND äh
44 DU gibst mir dann
45 die NÄCHste figur
46 das HEISST
47 ICH darf mir selber
48 NICHT die figur aussuchen
49 ICH suche !DIR! die figur aus
50 die DU dann (.) ins feld packst
51 und WICHTig ist dabei
52 dass du eine ehm VIERERlinie
53 bildest
54 MIT der (.)
55 selben (.) FORM
56 das heißt (.) wenn du alles
57 ne VIERER reihe in
58 KREISform sozusagen hast
59 dann hast du gewONNen
60 ist theoretisch son bisschen wie
61 vier geWINNT
62 nur das der gegner halt dir die
63 figur halt aussUCHT
64 ehm (0.9)
65 geNAU
66 aber WICHTig ist dabei einfach

2/6



63 dass du äh strateGisch vorgehst
64 und ahaust einFACH (0.4)
65 ..EX_Relevanz ÄHM (0.4) ja (0.5)
66 SIND da gerade
67 !ZWEI! kreise hintereinander ||
68 EE [mhm] _
69 EX [ähm] soll ich ihr JETZT einen
70 eine weitere FIGUR geben
71 wo ein KREIS ist
72 dann sind natürlich DREI drinne
73 ..EX_Relevanz und WIE hoch ist die
wahrscheinlichkeit
74 dass mein !GEG!ner mir dann AUCH
einen kreis gibt
75 oder EIN
76 EIN ähm ((räuspern))
77 ein VIEReck sozusagen ||
78 EE also ist es WICHTig
79 welche FORM die haben
80 also es ist NICHT
81 ..EE_Relevanz es IST eh
82 also DA
83 das die LÄNGlich sind
84 so
85 und SO (0.7) ||
86 EX geNAU
87 DAS ist d[as is]
88 ..EE_Architektur EE [ABER] es muss die
gleiche FORM haben
89 die figuren MÜSSen die gleichen
formen hab[en] _
90 EX
[JA] genau _

3/6



91 EE und !DU! gibst mir die FORM ||
92 EX geNAU
93 ..EX_Architektur und DU gibst !MIR! die form (1.
5) ||
94 EE okAY (0.5) _
95 EX GENau
96 also DA musst du auch
97 so n bisschen strateGisch
vorgehn ||
98 EE O[KAY] _
99 EX [welche] form GEB ich ihr denn
jetzt (1.1) ||
100 das äh ICH dann sozusagen
101 NICHT die chance bekomme
102 eine VIERer reihe zu bilden
103 tatsächlich kannst_e aber auch
SO vorgehen
104 dass DU
105 (.) MIR die figuren so gibst
106 dass ich sie (.) DIR sch schon
ne
107 DREIer (.) reihe bilde
108 und DU nur noch eine einzige
109 (.) FIGUR (0.3) äh
110 beNötigst
111 um diese VIERer reihe (0.3)
112 zu ähm (0.8)
113 komPLETT zu be[kommen] ||
114 EE [ACH so]
115 also SPIELen wir
116 die STEINE die gelegt sind
117 mit denen SPIELen wir beide ||
118 EX geN[AU] _

4/6



```

119 EE [also] es gibt nicht dass du
    BLAU hast
120 =und ich GRÜN [oder so] ||
121 EX [nein ]
122 das gibt es NICHT
123 es geht wircklich NUR
124 [um die] formEN
125 EE [ah: ] _
126 EX und um um mit DIESEN formen
127 eine VIERer bilde z?
128 VIERer (.)
129 REIhe zu bilden (1.1) ||
130 EE okAY
131 EX hast du_s verSTAN[Den]
132 EE [ich ]hab_s
    verSTANDen (1.0) _
133 EX OKE ((lacht))
134 sehr GUT
135 ((lachen) (1.0)
136 GENau(1.8) _
137 EE okEY (1.4) _
138 EX mHM (0.9) _
139 EE ihr könnt KOMMn
140 <<:-)> wenn ihr uns HÖRT>
141 ey das ist ja RIChtig komisch
142 [dass DU mir d]ie steine gibst _
143 EX [oder HAST du ]
144 mh_H
145 du gibst mir AUCH die steine
146 hast du noch FRAGEN
147 EE ((verneinend) MH_MH)
148 EX oKE
149 gut_

```

5/6



```

150 EE ((lacht))
151 EX _hab ich das gut be? äh erKLÄRT
152 EE _JA
153 SCHON
154 EX _okay sehr GUT
155 EE ((gähnt))
156 <<flüstert> richtig CRINGE
    grade>
157 _ (6.7) _
158 ((lacht)) (3.0) _
159 <<flüstert> warum KOMMT keiner>
    "" _
160 (3.9)
161 ((lachen))
162 (3.2) _
163 ((schnalzen)) was hast du
    ANGEKREUZT
164 _ ((Tür öffnet ))

```

6/6



C.2 A-Condition — BB₄ (WP₄-P₂-VP₃₁)

Architecture

Relevance

Both

[00:03.420–00:05.940] EX Also ich soll dir ein Spiel erklären.

[00:07.080–00:07.980] EX Ohne, dass es vor mir ist.

Start of standardisation

00:10.240–00:12.320

[00:10.240–00:12.320] EX Und genau, also das Spiel heißt Quarto.

[00:13.070–00:14.200] EX Sagt dir das vielleicht was?

[00:14.400–00:15.340] EE Nee absolut nicht.

[00:15.360–00:16.220] EX Also ich kannte es vorher auch nicht

[00:16.220–00:16.660] EX von dem her.

[00:18.208–00:18.448] EX Genau

[00:18.550–00:19.850] EX also Quarto funktioniert halt

[00:19.890–00:21.250] EX also Ziel des Spiels ist halt

[00:21.370–00:24.220] EX dass man vier Steine in eine Reihe kriegt

[00:24.220–00:26.590] EX die eine gemeinsame Eigenschaft haben.

[00:26.768–00:27.128] EE Okay

[00:27.188–00:27.388] EX ja.

[00:27.600–00:31.010] EX Und das Ganze findet auf nem 4x4 -Brett statt

[00:31.010–00:35.170] EX und wir haben 16 Spielsteine

[00:35.170–00:37.810] EX also so dass das Feld komplett ausgefüllt werden könnte.

[00:38.840–00:43.570] EX Und es gibt vier verschiedene Sets an Eigenschaften

[00:43.570–00:46.250] EX die man sozusagen dann in eine Reihe bringen kann.

[00:47.430–00:48.820] EX Das heißt wir haben einmal groß und klein

[00:49.030–00:50.340] EX hell und dunkel

[00:50.440–00:52.000] EX rund und eckig

[00:51.996–00:52.336] EX und

[00:52.010–00:54.440] EX mit Loch und ohne Loch.

[00:54.690–00:57.700] EX Also das heißt auch glatt ist wirklich eine Eigenschaft.

[00:58.031–00:58.153] EE Ja.

[01:00.230–01:02.220] EX Genau und man muss halt versuchen

[01:02.360–01:05.260] EX dass man in einer Reihe zum Beispiel alle großen kriegt.

[01:05.300–01:06.940] EX Also vier Stück die groß sind.

[01:07.980–01:09.660] EX Oder alle die dunkel sind.

[01:10.250-01:10.580] EE Okay, ja.

[01:10.656-01:11.036] EX Genau

[01:11.180-01:13.100] EX das ist so das grobe Ziel.

[01:13.096-01:13.696] EX Und

[01:13.110-01:18.440] EX jetzt... also normalerweise spielt man ja so:

[01:18.630-01:20.460] EX Du setzt einen Stein

[01:20.530-01:21.200] EX ich setze einen Stein.

[01:21.280-01:22.100] EX So schön abwechselnd.

[01:22.150-01:23.560] EX Jetzt funktioniert das aber so

[01:23.650-01:25.340] EX dass wir nicht getrennte Steine haben

[01:25.630-01:28.240] EX sondern wir haben gemeinsame Steine.

[01:28.850-01:30.920] EX Und ich suche für dich den Stein aus

[01:30.950-01:32.320] EX du setzt den aufs Spielbrett.

[01:32.470-01:35.020] EX Und dann suchst du den Stein für mich aus

[01:35.020-01:36.590] EX und ich setze den aufs

[01:36.124-01:36.604] EX Spielbrett.

[01:36.660-01:38.560] EE Aber ich darf selber entscheiden wo ich den hinsetze ne?

[01:38.600-01:39.060] EX Genau ja.

[01:39.070-01:41.160] EX Du darfst nur nicht aussuchen welchen Stein du setzt.

[01:41.164-01:41.624] EE Okay.

[01:42.200-01:43.780] EX Genau und also...

[01:44.960-01:46.720] EX man muss halt selber die Reihe beenden

[01:46.910-01:51.360] EX also nur der der die Reihe beendet gewinnt nachher auch.

[01:51.520-01:53.600] EX Also wenn hier drei dunkle stehen

[01:53.660-01:55.580] EX es kann sein dass du den gesetzt hast den gesetzt hast

[01:55.630-01:56.560] EX ich habe den gesetzt

[01:56.570-01:58.300] EX und du setzt aber den letzten

[01:58.440-01:59.960] EE dann hab ich die komplette Reihe.

[02:00.030-02:00.740] EX Genau ja.

[02:00.772-02:00.861] EE Okay.

[02:00.880-02:02.520] EX Und wenn ich den setzen würde

[02:02.620-02:03.460] EX dann hätte ich gewonnen.

[02:04.330-02:06.160] EX Und sobald eine Reihe halt voll ist

[02:06.230-02:07.140] EX ist das Spiel beendet.

[02:07.600-02:10.060] EX Und dann muss die Person die halt den letzten Stein gesetzt hat

[02:10.100-02:11.260] EX der halt passend ist...

[02:12.200-02:12.760] EX Quarto sagen.

[02:12.989-02:13.178] EE Okay.

[02:13.390-02:15.090] EX wenn die Person das aber nicht macht

[02:15.400-02:16.900] EX dann geht es an die andere Person

[02:16.910-02:18.340] EX und wenn die Person das gesehen hat

[02:18.470-02:20.440] EX also wenn du jetzt vergisst Quarto zu sagen

[02:20.440-02:23.630] EX und ich würde sehen oh da sind aber viele da hast du aber theoretisch gewonnen

[02:23.630-02:25.500] EX und du gibst mir einen Stein

[02:25.500-02:26.140] EX und ich habe den Stein

[02:26.240-02:27.120] EX dann kann ich sagen Quarto

[02:27.120-02:29.540] EX und dann habe ich sozusagen deinen Sieg geklaut.

[02:29.630-02:30.340] EE Okay ja.

[02:30.540-02:30.940] EX Genau.

[02:33.789-02:33.978] EX joa

[02:34.760-02:36.820] EE Ich habe noch eine Frage zu den Eigenschaften.

[02:36.820-02:37.890] EE Es gab vier Eigenschaften oder

[02:37.890-02:38.980] EE also einmal glatt.

[02:39.650-02:41.350] EE ist damit die Oberfläche gemeint?

[02:41.450-02:45.340] EX Genau die [sind wie] so Holzklötze

[02:41.930-02:43.050] EE [also wie der Stein...]

[02:45.340-02:48.310] EX und die sind ich weiß nicht wie groß

[02:48.460-02:48.770] EE Ja. []

[02:48.510-02:49.970] EX aber [so] relativ klein

[02:49.970-02:55.870] EX und dann hast du wirklich wie so eine glatte also wie so die typischen Spielsteine eigentlich.

[02:55.951-02:56.055] EE Okay.

[02:56.110-02:57.870] EX Also wirklich einfach so ein Holzklotz.

[02:58.010-02:58.910] EE Und die anderen sind dann?

[03:00.090-03:01.030] EX Die anderen haben wie so...

[03:01.030-03:04.070] EX Das ist wie so reingefräst einfach so von oben.

[03:04.093-03:04.172] EE Okay.

[03:04.180-03:05.930] EX Also einfach nur so man erkennt es

[03:05.980-03:07.550] EX also es ist auf jeden Fall so eine Kuhle dann da.

[03:08.980-03:11.260] EX Also groß und klein ist glaube ich relativ selbsterklärend []

[03:10.960-03:11.840] EE klar.] [Ja

[03:11.260-03:12.700] EX also das ist dann einfach so in der Höhe

[03:12.736-03:14.596] EX genau

[03:14.780-03:17.200] EX und rund ist halt wie so zylindermäßig

[03:17.200-03:21.820] EX und eckig ist halt so klotzartig

[03:21.820-03:25.960] EX und was vergess ich denn jetzt?

[03:26.310-03:26.740] EE hell und dunkel

[03:26.836-03:26.976] EE also

[03:26.850-03:27.200] EE dunkel.

[03:27.430-03:28.330] EX Ja ja genau ja

[03:28.416-03:28.796] EX also

[03:28.430-03:29.740] EX auch richtig

[03:29.756-03:30.216] EX genau

[03:30.730-03:32.240] EX einfach unterschiedliches Holz.

[03:33.515-03:34.469] EX Joa. ach ja und

[03:34.608-03:35.912] EX oh genau

[03:35.948-03:36.727] EX Da fällt mir gerade noch ein

[03:36.881-03:38.629] EX wir haben ja dieses 4x4 Brett

[03:38.629-03:45.851] EX und du kannst sowohl horizontal als auch vertikal als auch diagonal eine 4er -Reihe bilden

[03:45.856-03:46.701] EX und gewinnen.

[03:46.791-03:47.671] EX Also ja

[03:48.272-03:49.766] EX vielleicht noch ganz wichtig zu wissen

[03:49.823-03:51.152] EX sonst versucht man nur...

[03:51.184-03:53.006] EX weil ich so horizontal gezeigt habe

[03:54.348-03:56.266] EX Klang gut.

[04:00.557-04:01.247] EX Hast du noch Fragen?

[04:01.816-04:02.646] EE Ich überlege gerade

[04:06.709-04:08.487] EE also kann man das nur zu zweit spielen? oder?

[04:08.584-04:08.784] EX Ja.

[04:08.824-04:09.164] EE Okay.

[04:09.481-04:10.646] EX Also soweit ich weiß.

[04:12.191-04:12.397] EE Okay.

[04:15.335-04:16.842] EE Ansonsten habe ich keine Frage mehr.

[04:17.646-04:19.601] EX Ich überlege auch gerade ob irgendwie so

[04:19.608-04:21.386] EX ich habe dir die Eigenschaften gezeigt

[04:21.424-04:22.525] EX ich habe dir gesagt wie man gewinnt

[04:22.525-04:24.684] EX und wie das Spielbrett aussieht.

[04:26.038-04:27.924] EX Mehr ist halt bei dem Spiel irgendwie nicht.

[04:27.964-04:28.144] EE Ja.

[04:29.551-04:30.430] EE Okay cool.

[04:35.519-04:37.260] EX Zum Glück ein recht einfach zu erklärendes Spiel.

End of standardisation

04:35.519-04:37.260

[04:38.268-04:38.395] EE Ja.

C.3 R-Condition — BB₄ (WP₄-P₂-VP₆₄)

Architecture

Relevance

Both

[00:02.890-00:05.400] EX Du darfst jetzt da sitzen und dir anhören

[00:05.396-00:05.516] EX wie

Start of standardisation

00:05.410-00:08.750

[00:05.410-00:08.750] EX ich dir versuche Strategien zu einem Spiel zu erklären.

[00:10.940-00:13.750] EX Also es ist ein Strategiespiel.

[00:15.310-00:19.170] EX Und... Ich weiß nicht, spielst du oft Spiele oder so?

[00:19.306-00:19.486] EE Also

[00:19.320-00:23.350] EE ich spiele so normale Gesellschaftsspiele

[00:23.440-00:25.460] EE wie so UNO Skippo also sowas.

[00:25.730-00:27.650] EE Also nicht sowas wie Siedler von Katan

[00:27.650-00:28.230] EE oder wie das heißt.

[00:28.510-00:28.730] EX Okay, ja.

[00:29.760-00:33.080] EX Ja, aber z.B. bei Skippo gibt es ja gewisse Strategien

[00:33.120-00:34.540] EX die man anwenden kann um zu

[00:34.240-00:34.560] EX gewinnen

[00:34.540-00:37.960] EX und hier bei dem Spiel ist es so

[00:38.210-00:41.080] EX dass es ganz wichtig ist

[00:41.170-00:43.090] EX den Überblick zu behalten

[00:43.090-00:44.700] EX und drei Schritte voranzudenken

[00:44.700-00:51.560] EX und das Spiel eröffnet eigentlich sehr viele Möglichkeiten

[00:51.710-00:53.020] EX also je weiter das Spiel voranschreitet

[00:53.120-00:54.900] EX desto mehr Möglichkeiten gibt es zum Gewinnen

[00:55.060-00:55.940] EX aber auch zum Verlieren

[00:55.940-01:00.540] EX und wir haben verschiedene Steine

[01:00.920-01:04.500] EX die wir in verschiedenen Gewinnkonditionen an sich setzen können

[01:04.500-01:09.520] EX und das Problem bei diesen Gewinnkonditionen ist aber

[01:09.640-01:11.740] EX dass man nicht nur selber mit denen gewinnen kann

[01:11.900-01:12.760] EX sondern auch der Gegner

[01:14.790-01:16.480] EX und zwar wir haben ein Brett

[01:16.480-01:20.000] EX und da drauf stehen dann die Figuren die wir setzen

[01:20.000-01:26.090] EX und ja vielleicht ist das erste worauf man achten muss

[01:26.090-01:27.480] EX beim Gegner ist

[01:27.750-01:31.010] EX dass man nicht selber den Stein auswählt den man setzen möchte

[01:31.220-01:32.850] EX sondern man sucht den für den Gegner aus.

[01:33.360-01:35.990] EX Das bedeutet hier muss man auch schon einen Schritt vorwärts denken

[01:35.990-01:39.110] EX und sagen welchen Stein gefällt mir gut

[01:39.220-01:39.970] EX welchen möchte ich haben

[01:40.030-01:41.350] EX den soll ich natürlich nicht abgeben.

[01:42.180-01:46.420] EX Und stattdessen überlegen welcher Stein nützt dem Gegner

[01:46.152-01:46.432] EX nichts

[01:46.420-01:48.050] EX oder schadet ihm vielleicht sogar.

[01:50.410-01:51.910] EX Genau, also das ist so das erste

[01:52.040-01:53.570] EX wo man schon einen Schritt voraus denken muss.

[01:53.572-01:53.752] EE Ist

[01:53.600-01:56.860] EE das so ein Spielbrett wie... ein Schachspiel oder...

[01:56.880-01:58.400] EX Ist so ein 4x4 -Brett

[01:58.396-01:58.636] EX ähnlich

[01:58.410-01:59.280] EX wie Schach ja.

[01:59.941-02:00.109] EX Genau.

[02:00.890-02:04.240] EX Und jetzt wenn ich jetzt überlegt

[02:04.320-02:05.340] EX du hast mir einen Stein gegeben

[02:05.580-02:07.040] EX dann setze ich den irgendwo aufs Brett

[02:07.200-02:09.160] EX das darf ich frei auswählen.

[02:10.020-02:13.760] EX Natürlich kann es sein dass es gewisse Stellen gibt die besser sind als andere.

[02:14.020-02:16.340] EX Am Anfang macht das aber tatsächlich keinen Unterschied.

[02:17.220-02:20.300] EX Und ich setze jetzt zum Beispiel hier einen Stein in die Ecke einfach rein.

[02:23.500-02:25.390] EX das bedeutet [genau]

[02:24.120-02:25.000] EE [Den Stein habe ich für dich ausgesucht.]

[02:25.540-02:27.000] EX den hast du

[02:26.412-02:27.012] EX ausgesucht

[02:27.000-02:27.730] EX und hast gesagt so

[02:28.310-02:30.550] EX der Stein mit den Eigenschaften sieht toll aus

[02:30.560-02:31.510] EX und du darfst ihn jetzt setzen.

[02:31.570-02:32.430] EX Jetzt habe ich den gesetzt

[02:32.550-02:33.430] EX jetzt gebe ich dir einen

[02:33.500-02:34.590] EX den setzt du vielleicht hier hin

[02:34.870-02:35.810] EX dann gibst du mir wieder einen

[02:36.120-02:37.370] EX den setze ich dann hier hin.

[02:37.590-02:39.450] EX Und jetzt haben wir drei Steine in einer Reihe.

[02:39.720-02:43.330] EX Und für diese Gewinnkondition die können zwar unterschiedlich aussehen

[02:43.430-02:44.550] EX aber was bei allen gleich ist

[02:44.630-02:46.590] EX es müssen vier Steine in einer Reihe sein.

[02:49.270-02:52.020] EX und das heißt sobald eine Reihe drei Steine hat

[02:52.170-02:53.040] EX wird es gefährlich

[02:53.430-02:54.280] EX weil es fehlt nur noch...

[02:54.560-02:56.320] EE Der Nächste der dann den Stein setzt

[02:56.560-02:57.280] EE der hat gewonnen?

[02:57.416-02:57.516] EX Genau.

[02:57.520-02:58.780] EE Oder hast du dann gewonnen

[02:58.930-03:02.380] EE weil du ja meinen Stein ausgesucht hast?

[03:02.500-03:03.710] EX Nee die Person die den setzt.

[03:03.996-03:04.256] EE Okay.

[03:04.496-03:04.756] EX Genau.

[03:05.770-03:08.040] EX Und also sobald drei Steine da sind

[03:08.360-03:09.550] EX ist gefährlich potenziell.

[03:10.450-03:14.110] EX Jetzt ist es so diese Gewinnkonditionen können so aussehen

[03:14.180-03:16.320] EX dass die zum Beispiel alle gleich groß sind.

[03:17.650-03:19.090] EX Sagen wir das sind alles große Steine.

[03:19.530-03:24.070] EX Das heißt wenn ich jetzt den Stein für dich auswählen müsste

[03:24.160-03:26.410] EX weil den habe ich gesetzt den hast du gesetzt den habe ich gesetzt

[03:27.300-03:29.410] EX dann sollte ich dir natürlich keinen großen Stein geben.

[03:29.940-03:32.990] EX Und wir haben hier so einen hier oder hier darf man sich dann aussuchen

[03:33.420-03:34.850] EX so einen Pool an Steinen

[03:35.290-03:37.910] EX aus dem man auswählen kann.

[03:38.528-03:42.840] EX Und das heißt natürlich da stehen jetzt große wie kleine Steine

[03:43.070-03:45.230] EX dann würde ich dir natürlich einen kleinen geben.

[03:45.410-03:46.960] EE Aber der letzte Stein den man setzt

[03:47.170-03:50.050] EE der bestimmt wie viel man gewonnen hat sozusagen.

[03:50.180-03:52.350] EX Genau also ob man überhaupt gewinnt.

[03:52.510-03:55.590] EX Also wenn du jetzt sagst okay ich gebe dir jetzt den kleinen

[03:56.520-03:57.810] EX sagst naja da bringt der mir nichts

[03:58.400-03:59.410] EX dann setzt ihn vielleicht hier hin.

[04:00.160-04:04.940] EX Also es kommt immer darauf an was man persönlich als Bestes erachtet

[04:04.940-04:09.200] EX und man sollte halt versuchen. irgendwie eine Gewinnkondition zu erfüllen.

[04:10.620-04:12.380] EX und jetzt ist es halt so

[04:12.500-04:14.480] EX dass man halt darauf achten muss was auf diesem Brett ist.

[04:14.560-04:16.480] EX Also habe ich irgendwo schon eine Gefahr?

[04:16.696-04:16.796] EX Also

[04:16.710-04:17.560] EX drei Steine.

[04:17.776-04:18.216] EX Das

[04:17.790-04:21.220] EX kann auch diagonal oder vertikal auch sein.

[04:23.696-04:24.734] EX Oder

[04:23.710-04:26.240] EX ist da alles in Ordnung?

[04:26.420-04:28.720] EX Und es sind überall nur zwei Maximal in einer Reihe?

[04:28.780-04:30.880] EX Oder vielleicht ist auch erst gerade drei Steine gesetzt

[04:30.940-04:32.580] EX aber die stehen komplett unterschiedlich.

[04:32.580-04:36.880] EX Das zweite worauf man achten muss ist...

[04:37.780-04:39.680] EX dass man nicht nur das Feld im Auge behält

[04:39.970-04:41.420] EX sondern auch den Pool an Steinen.

[04:42.030-04:43.740] EX Weil wir haben nicht unendlich viele Steine
 [04:43.840-04:45.720] EX sondern die sind halt auch limitiert.
 [04:46.150-04:47.900] EX Das heißt es kann irgendwann sein
 [04:48.110-04:52.320] EX dass du mir natürlich keinen großen geben möchtest.
 [04:52.897-04:54.240] EX Es sind aber nur noch große da.
 [04:54.990-04:56.040] EX Dann kannst du mir den geben.
 [04:57.350-04:59.620] EX Und deine Hoffnung ist dann dass ich das einfach
 übersehe
 [04:59.620-05:00.960] EX und den woanders hinstelle.
 [05:02.360-05:04.200] EX Dann geht das Spiel halt einfach weiter
 [05:04.360-05:05.980] EX weil ich die Gewinnkondition nicht erfülle.
 [05:06.890-05:09.740] EX Wenn du Pech hast stelle ich ihn natürlich dahin wo er
 passt
 [05:09.900-05:10.680] EX weil ich es gesehen habe
 [05:10.680-05:11.500] EX und dann gewinne ich.
 [05:11.890-05:13.500] EE Kannst du nochmal sagen wie die Gewinnkondition
 [05:13.640-05:16.040] EE also wenn wir jetzt drei in eine Reihe gelegt haben
 [05:16.940-05:18.940] EE dann der größte Stein in der Reihe
 [05:19.090-05:20.120] EE der hat dann gewonnen?
 [05:20.200-05:20.860] EE Oder wie ist das?
 [05:20.870-05:24.100] EX Nee also die Steine haben alle unterschiedliche
 Eigenschaften
 [05:24.100-05:26.420] EX und die kommen immer in so Pärchen
 [05:26.530-05:27.580] EX also groß und klein
 [05:27.850-05:28.700] EX hell und dunkel
 [05:28.940-05:29.820] EX rund und eckig
 [05:29.820-05:31.940] EX und mit Loch ohne Loch.
 [05:33.070-05:37.650] EX Und es muss nur eine dieser Eigenschaften gemeinsam
 sein in allen vier Steinen.
 [05:37.920-05:43.910] EE Also müssen von den vier Steinen müssen zwei sozusagen
 das Gegenteil voneinander sein
 [05:44.090-05:45.830] EE also zwei zusammenhörige Paare?
 [05:46.008-05:46.188] EX Ne
 [05:46.390-05:48.760] EX es müssen alle vier diese Eigenschaft haben.
 [05:49.400-05:50.970] EE Okay, also es müssen alle dann
 [05:51.180-05:52.710] EE wenn du jetzt helle und dunkel
 [05:52.710-05:53.490] EE hast du ja gerade gesagt.
 [05:54.160-05:56.190] EE Dann müssen alle in der Reihe hell
 [05:56.188-05:56.531] EX Genau.

[05:56.530-05:57.810] **EE** oder alle dunkel sein?

[05:57.808-05:57.988] **EX** Genau.

[05:58.765-05:58.973] **EE** Okay.

[05:59.590-06:01.320] **EX** Und also die Steine haben mehrere Eigenschaften.

[06:01.540-06:02.920] **EX** Das macht es halt auch noch mal schwieriger.

[06:03.210-06:05.670] **EX** Also ein Stein kann nicht nur hell sein

[06:05.780-06:07.400] **EX** sondern der kann auch gleichzeitig groß sein.

[06:07.940-06:09.740] **EX** Oder und dann noch rund.

[06:11.160-06:11.368] **EE** Also

[06:11.180-06:12.500] **EE** da muss man abwägen

[06:12.680-06:15.460] **EE** welche Eigenschaften jetzt schon in der Reihe erfüllt sind

[06:15.670-06:16.920] **EE** welchen ich dir dann gebe.

[06:17.800-06:19.260] **EE** Okay, so langsam klärt sich das ein bisschen.

[06:19.261-06:19.400] **EX** Genau.

[06:19.870-06:22.180] **EX** Also es ist halt eben dieses vorausschauende Planen.

[06:22.340-06:23.360] **EX** Eben was sehe ich?

[06:23.480-06:24.690] **EX** Was sieht vielleicht der Gegner?

[06:24.940-06:26.200] **EX** Also eben dieses Abwägen.

[06:26.320-06:27.060] **EX** Gibst du mir jetzt den?

[06:27.110-06:29.240] **EX** Also du hast gesehen du kannst eigentlich nicht mehr gewinnen.

[06:29.960-06:32.520] **EX** Aber du kannst mir trotzdem noch mit einem Poker-Face versuchen zu geben

[06:32.640-06:34.180] **EX** den Stein versuchen zu geben

[06:34.400-06:35.980] **EX** weil vielleicht übersehe ich es ja tatsächlich.

[06:38.870-06:41.880] **EX** Genau, also weil man diesen Pool auch im Auge behalten muss

[06:42.240-06:44.960] **EX** muss man halt noch einen Schritt vorausdenken.

[06:45.180-06:46.500] **EX** Also das heißt zum Beispiel wenn ich sehe

[06:46.980-06:50.000] **EX** okay ich gebe dir jetzt einen großen

[06:50.320-06:52.060] **EX** dann könntest du mir noch einen großen geben.

[06:52.700-06:54.090] **EX** Und dann sind aber nur noch kleine da.

[06:54.670-06:57.090] **EX** Das heißt da muss man dann auch schon ein bisschen im Voraus planen

[06:57.280-07:01.860] **EX** wo setze ich meinen Stein hin dass ich nicht plötzlich nur kleine Möglichkeiten offen habe.

[07:05.658-07:05.860] **EX** Genau.

[07:05.860-07:08.780] **EX** was auch noch ist mit diesen

[07:07.192-07:08.792] **EX** " Pärcheneigenschaften

[07:08.980-07:10.330] EX also dieses hell -dunkel und so.

[07:11.250-07:13.810] EX Ein Stein kann halt nicht gleichzeitig hell und dunkel sein.

[07:13.810-07:17.830] EX So und das bedeutet halt auch

[07:21.360-07:21.860] EX Ja, wie soll ich das erklären?

[07:24.680-07:31.700] EX Also wenn ein dunkler und ein heller Stein in einer Reihe sind

[07:31.700-07:35.500] EX und auch alle anderen Eigenschaften immer in diesen Pärchen -Eigenschaften sind

[07:36.250-07:38.760] EX also sozusagen die genauen Gegensätze voneinander sind

[07:38.950-07:41.380] EX dann kann man in der Reihe auch keine Gewinnkondition bilden

[07:41.540-07:43.730] EX weil sobald einer ist

[07:43.770-07:45.280] EX der hier ist groß der ist klein.

[07:45.840-07:47.360] EX Der hier ist dunkel der hier ist hell

[07:47.500-07:48.900] EX der hier hat ein Loch der hier hat keines

[07:48.980-07:50.420] EX der hier ist rund und der hier ist eckig.

[07:50.760-07:53.320] EX Das heißt die sind in keiner Eigenschaft gemeinsam

[07:53.320-07:56.160] EX und das blockiert halt an sich auch die Reihe.

[07:56.230-07:57.440] EE Okay, dann ist die Reihe also verloren

[07:57.440-07:59.720] EE und man kann dann sowieso nicht mehr gewinnen.

[07:59.914-08:00.097] EX Genau.

[08:00.780-08:04.260] EX Also da ist es auch egal ob da jetzt zwei drei oder vier Steine drinstehen.

[08:05.720-08:08.200] EX Also es kann auch passieren dass eben vier in einer Reihe sind

[08:08.460-08:09.920] EX aber halt nichts erfüllt ist

[08:09.920-08:12.060] EX und dann geht das Spiel einfach ganz normal weiter.

[08:12.720-08:14.540] EX das kann zum Vorteil sein

[08:14.610-08:16.300] EX wenn man diese Reihen eben so zumacht

[08:16.860-08:19.900] EX weil man halt sagt oh hier sind drei mit Loch

[08:20.550-08:22.510] EX Dann stell ich mal lieber eine ohne Loch hier

[08:22.224-08:22.524] EX " hin

[08:23.320-08:25.700] EX um die Reihe zu schließen weil dann passiert damit nichts mehr.

[08:26.450-08:27.420] EX Nachteil ist aber

[08:27.660-08:31.800] EX man verbaut sich halt Gewinnmöglichkeiten

[08:32.020-08:35.740] EX aber manchmal kann es sein dass so viele Möglichkeiten offen sind

[08:35.840-08:37.040] EX dass man die zumachen muss

[08:37.260-08:41.370] EX weil man sonst aufgrund des übrigen Pools sonst verlieren

[08:41.124-08:41.384] EX würde.

[08:42.992-08:43.212] EX Genau.

[08:43.980-08:46.000] EE Und wenn jetzt eine Reihe voll ist

[08:46.000-08:48.940] EE und die in keiner Eigenschaft übereinstimmt

[08:49.060-08:50.220] EE also man die gar nicht gewinnen kann

[08:50.560-08:51.440] EE dann bleibt die stehen.

[08:51.624-08:51.804] EX Genau.

[08:52.180-08:54.830] EE Und wenn jetzt aber eine Reihe gewonnen worden ist

[08:54.950-08:58.230] EE weil zum Beispiel alle hell

[08:57.864-08:58.244] EE sind

[08:58.230-08:59.420] EE in der Reihe z.B.

[08:59.680-09:01.750] EE würde die Reihe dann

[09:01.144-09:01.764] EE verschwinden?

[09:02.190-09:03.520] EX Nee dann ist das Spiel Ende.

[09:03.720-09:04.810] EE Ach dann hat einer schon

[09:04.744-09:04.824] EE gewinnen?

[09:04.824-09:05.024] EX Genau.

[09:05.130-09:06.620] EE Also man muss nur eine Reihe gewinnen?

[09:06.704-09:06.904] EX Genau.

[09:07.140-09:09.000] EX Und dann würde man einfach wieder von vorne anfangen.

[09:09.224-09:09.484] EE Okay.

[09:10.024-09:10.164] EX Ja.

[09:11.370-09:12.230] EX Also an sich...

[09:13.490-09:14.070] EX also es ist

[09:14.200-09:17.610] EX ich finde es ist ein Spiel was an sich relativ simpel ist

[09:17.690-09:20.510] EX weil man nur diese vier gleich in eine Reihe bringen muss.

[09:21.040-09:23.550] EX Aber es ist halt so eine Komplexität die dahinter steckt

[09:23.650-09:28.330] EX dass so eine Runde eigentlich manchmal kann die richtig schnell gehen

[09:28.430-09:29.770] EX weil jemand halt schnell was übersieht

[09:29.770-09:30.930] EX oder man einfach schnell spielt

[09:31.070-09:33.430] EX kann sich aber auch total in die Länge ziehen

[09:34.280-09:36.650] EX einfach weil man so viele Möglichkeiten hat

[09:36.740-09:37.870] EX die man überdenken muss.

[09:38.960-09:39.980] EX Und es macht es halt schwieriger

[09:40.000-09:41.560] EX wenn man auch diagonal gucken muss.

[09:41.820-09:43.980] EX Häufig guckst du so horizontal vertikal

[09:43.980-09:50.320] EX und dann vergisst man leider dass man diagonal auch noch eine Möglichkeit gehabt hätte die da übrig war.

[09:51.680-09:52.100] EE Okay, also

[09:52.100-09:55.080] EE es gibt diesen Pool an Steinen den wir haben

[09:55.180-09:56.360] EE mit verschiedenen Eigenschaften.

[09:56.430-09:59.520] EE Und jeder Stein hat diese vier oder fünf Eigenschaften

[09:59.560-10:01.630] EE die jeweils unterschiedlich sind.

[10:01.800-10:02.840] EE Also hell dunkel

[10:04.150-10:04.760] EE eckig, rund

[10:04.960-10:05.980] EE mit Loch ohne Loch.

[10:06.210-10:06.780] EE Was gab es noch?

[10:07.080-10:09.520] EX Ehh groß und klein hast du glaube ich noch nicht gesagt.

[10:09.780-10:10.940] EE Also vier Eigenschaften

[10:10.936-10:11.164] EX Genau.

[10:11.160-10:12.180] EE sind im Spiel sozusagen.

[10:12.458-10:12.546] EX Ja.

[10:13.110-10:15.680] EE Okay und wir suchen uns immer gegenseitig Steine aus.

[10:15.910-10:17.840] EE Das heißt wenn du jetzt anfängst

[10:17.850-10:19.330] EE würde ich dir als erstes einen Stein

[10:19.004-10:19.344] EE geben.

[10:19.561-10:19.753] EX Genau.

[10:20.256-10:20.487] EE Und

[10:20.270-10:23.320] EE den würdest du dann hier reinsetzen irgendwo.

[10:24.180-10:25.900] EE dann würdest du mir einen Stein geben

[10:26.264-10:26.684] EE den

[10:26.320-10:27.740] EE würde ich dann dazusetzen.

[10:27.860-10:32.260] EE Und man muss versuchen in einer Reihe dass alle in irgendeiner Eigenschaft gleich sind.

[10:32.464-10:32.664] EX Genau.

[10:33.464-10:33.764] EE Okay.

[10:34.164-10:34.304] EX Ja.

[10:35.268-10:35.368] EX dann

[10:35.320-10:37.510] EX muss man halt nur sehr viel vorausdenken

[10:37.510-10:40.540] EX und hoffen dass der Gegner Sachen nicht so schnell sieht wie man selber sieht.

[10:40.790-10:43.210] EE Okay ist ja ein bisschen wie Vier-Gewinnt.

[10:43.348-10:43.568] EX Ja

[10:43.670-10:44.050] EE nur komplexer.

[10:44.610-10:45.450] EX Ja, genau. Hat was.

[10:45.468–10:45.668] **EX** Ist

[10:45.480–10:46.870] **EX** ja auch Vier in eine Reihe bringen.

[10:47.590–10:49.600] **EX** Ja ich hatte auch so ein bisschen an Tic -Tac -Toe gedacht

[10:49.750–10:50.820] **EX** aber das ist ja nur mit drei

[10:50.820–10:54.310] **EX** und sehr einfach würde ich jetzt mal behaupten.

[10:55.140–10:56.750] **EE** Bei Vier-Gewinnt hast du halt so ein größeres Feld.

[10:58.607–10:58.787] **EX** Ja.

[11:01.690–11:03.040] **EX** Hab ich irgendwas vergessen

[11:03.120–11:04.470] **EX** Oder hast du noch Fragen?

[11:06.170–11:06.380] **EE** Ich überlege...

[11:07.980–11:10.290] **EE** aber ich glaube ich habe es verstanden.

[11:11.590–11:12.490] **EE** Erstmal ohne Spiel.

[11:12.912–11:13.052] **EX** Ja.

[11:14.830–11:17.750] **EX** Also wie gesagt man kann halt so strategiemäßig an sich blocken

[11:17.750–11:19.360] **EX** und sagen ich möchte alles zu

[11:19.112–11:19.372] **EX** " haben

[11:20.250–11:21.730] **EX** dass ich halt nichts übersehe

[11:21.730–11:23.140] **EX** oder man macht halt viel

[11:22.732–11:23.152] **EX** auf

[11:23.140–11:25.430] **EX** in der Hoffnung dass der Gegner Zeug übersieht aber

[11:27.370–11:29.620] **EX** ist halt auch nicht immer. Ziel führend

[11:29.680–11:32.380] **EX** weil man selber halt dann vielleicht doch auch selber was übersieht

[11:32.380–11:33.960] **EX** und dann hat man halt direkt verloren.

[11:36.250–11:37.940] **EE** Hast du das Spiel auch gerade schon gespielt?

[11:38.410–11:39.730] **EX** Ich musste mir das ausleihen

[11:39.730–11:43.280] **EX** und dann durfte ich über das Wochenende ausprobieren ja.

[11:43.680–11:46.240] **EX** Also ich habe schon ein paar Runden gespielt.

[11:48.536–11:49.796] **EE** Bestimmt müssen wir das gleich noch spielen.

[11:49.916–11:53.676] **EX** " Ja vermutlich. Ich hoffe " mal sonst wurde es ja nicht so gut vorbereitet.

[11:55.456–11:56.336] **EE** Ich " glaube dann wären wir fertig.

[11:56.476–11:57.436] **EX** " Ja ich denke auch.

[11:58.453–11:59.280] **EE** Wir wären fertig.

[12:00.175–12:00.820] **EX** Ahh, sie hört sogar.

[12:01.060–12:02.900] **EE** " Ja das ist ein bisschen gruselig.

[12:06.390–12:07.140] **EX** Alles mitgehört.

[12:07.590-12:10.160] **EE** Ich kann mir die Steine noch nicht gut vorstellen wie die aussehen

[12:10.940-12:12.300] **EE** wenn die alle Eigenschaften haben.

[12:15.540-12:16.060] **EX** Spannend.

[12:17.385-12:17.512] **EX** Es

[12:17.400-12:19.480] **EX** sind wie so Holzfigürchen an sich.

[12:24.400-12:25.800] **EX** Wie kann ich das am besten beschreiben?

[12:29.040-12:30.050] **EX** ja, wie so Türmchen

[12:30.740-12:32.290] **EX** aber dann ist es halt so quadratisch

[12:32.820-12:34.270] **EX** hat so eine Einkerbung obendrinne

[12:34.780-12:36.500] **EX** ja gut und hell und dunkel ist ja auch irgendwie klar.

[12:37.830-12:40.690] **EX** und das geht halt mit allen Eigenschaften so.

[12:41.030-12:41.350] **EE** Okay.

[12:41.810-12:41.930] **EX** Ja.

[12:48.870-12:49.230] **EE** (unverständlich)

[12:54.041-12:55.930] **EE** Das " heißt du warst dann auch schon eher " hier oder?

[12:58.230-13:00.170] **EX** " Ja Anfang letzter Woche.

[13:02.130-13:05.610] **EX** Da haben sie mich hier irgendwo in einen der Räume zitiert.

[13:06.756-13:08.090] **EX** Da wurde mir das Spiel überreicht.

[13:09.158-13:09.387] **EE** Okay.

[13:09.519-13:09.622] **EX** Ja.

[13:11.650-13:14.770] **EE** Es klingt irgendwie auch als wäre dass ein Spiel das extra dafür entwickelt worden ist.

[13:15.800-13:20.320] **EX** Ja, also ich finde das ist halt sehr... ...mathematisch so

[13:20.540-13:23.080] **EX** also sehr strukturiert irgendwie

End of standardisation

13:20.540-13:23.080

[13:23.280-13:25.190] **EX** also kann schon gut sein

[13:25.190-13:28.080] **EX** also wenn dann ich fand es auf jeden Fall sehr cool also...

[13:29.640-13:29.960] **EE** bin gespannt

[13:30.240-13:31.580] **EE** ob ich das hinkriege.

[13:33.100-13:33.630] **EX** wird schon. wird schon.

[13:33.924-13:35.404] **EE** " Ja hängt ja nichts dran.

[13:38.061-13:39.224] **EX** Es ist der große Lotto -Preis.

[13:40.464-13:40.704] **EE** danach.

[13:42.864-13:45.364] **EX** bist du auch über die E -Mail da an die Studie?

[13:45.544-13:46.524] **EE** ich glaube es gab so eine Rundmail.

[13:46.584-13:47.044] **EX** " Ja genau.

[13:47.592-13:49.208] **EE** da habe ich " gedacht das nutze ich jetzt noch.

[13:50.400-13:51.048] **EE** Was studierst du?

[13:51.648-13:54.328] **EX** Englisch -Geschichte und Kunst auf Lehramt. Was machst du?

[13:54.968-13:55.328] **EE** Grundschullehramt

[13:55.548-13:55.808] **EX** Ahh ja.

[14:00.763-14:02.308] **EX** Bist du schon im Master oder noch im Bachelor?

[14:03.426-14:05.508] **EE** Ja, also ich fange jetzt bald mit dem Ref an im Mai.

[14:06.148-14:06.448] **EX** " ja gut.

[14:06.468-14:11.348] **EE** Also heute ist der letzte " Tag wo ich noch Studentin bin.
[] Aber ich muss eh heute " arbeiten dann kann ich das jetzt noch mitmachen.

[14:08.608-14:08.768] **EX** Wahr.]

[14:11.988-14:12.468] **EX** Glückwunsch.

[14:12.628-14:13.168] **EE** " Ja danke.

[14:14.334-14:15.028] **EX** Ich habe noch ein bisschen.

[14:16.014-14:16.636] **EE** Du bist im Bachelor?

[14:16.956-14:21.276] **EX** " Nee " Master das ist mein zweites " Semester aber ich habe das Praxismester geschoben.

[14:21.596-14:22.396] **EE** Ins vierte dann?

[14:22.576-14:23.116] **EX** " Ja genau.

[14:23.912-14:26.336] **EX** weil drei Fächer und dann noch Praxismester.

[14:26.556-14:28.596] **EE** Aber du machst dann halt ein " Ergänzungsfach oder?

[14:28.716-14:28.996] **EX** Genau.

[14:31.156-14:37.156] **EE** habe ich im Nachhinein auch " gedacht eigentlich wäre es " cool noch irgendwie so ein anderes " Fach also bei mir wäre nur noch Sport in Frage gekommen.

[14:37.536-14:37.776] **EX** Ja.

[14:39.096-14:40.556] **EE** Jetzt ist zu " spät jetzt mache ich es auch nicht mehr.

[14:41.196-14:57.876] **EX** ansonsten kann man auch immer noch zurückgehen. Also ich habe auch eine Kommilitonin " gehabt die ist jetzt schon Lehrerin so schon seit längerem und die hat dann noch Kunst mit dazu genommen jetzt am Ende und hat gut " funktioniert also sie ist da gut durchgekommen.

[14:58.056-14:59.616] **EE** " Ja ist halt die Frage wie es zeitlich " ist nur wenn man dann einmal...

[14:59.616-15:01.696] **EX** Kommt aufs Fach an.

C.4 AR-Condition — BB4 (WP4-P2-VP79)

Architecture

Relevance

Both

[00:02.730-00:03.700] EX Ich soll dir das Spiel erklären.

[00:05.282-00:05.440] EE Okay.

[00:06.650-00:08.060] EX Ich muss kurz überlegen

[00:08.150-00:11.210] EX weil ich dachte ich hätte das Spiel quasi

[00:10.482-00:11.222] EX daneben

[00:11.210-00:12.300] EX zum erklären.

[00:13.030-00:14.780] EE Weißt du was meine Aufgabe dann ist?

[00:15.860-00:17.660] EX Nee also ich soll dir einfach das Spiel erklären.

[00:17.862-00:18.162] EE Okay.

[00:18.640-00:19.940] EE Mir wurde gar nichts gesagt.

[00:20.010-00:20.480] EX Ja okay.

[00:21.200-00:23.020] EE Okay dann hör [ich dir jetzt erstmal einfach] zu.

[00:21.710-00:22.520] EX [(Unverständlich)]

[00:23.650-00:26.760] EX Ja, ich glaube... das war ja auch Explaineer und Explainer

[00:26.762-00:27.022] EE Ja.

[00:27.150-00:28.740] EX Ich glaube du musst nur zuhören.

[00:33.060-00:35.470] EX Warte, ich versuche es zu ordnen

Start of standardisation

00:38.820-00:40.730

[00:38.820-00:40.730] EX Also du hast also es heißt Quatro

[00:41.730-00:44.210] EX also der Name leitet das ja schon so ein bisschen ab

[00:44.268-00:44.328] EX du

[00:44.280-00:47.030] EX musst eine Viererreihe bilden

[00:47.790-00:48.750] EX ist ein bisschen wie Vier-Gewinnt.

[00:49.050-00:50.570] EX wenn du eine Viererreihe gebildet hast

[00:50.570-00:51.210] EX dann hast du gewonnen

[00:51.210-00:54.390] EX und wir haben ein gemeinsames Spielfeld

[00:54.520-00:55.420] EX so ein 4x4 -Feld

[00:55.420-00:58.170] EX und da legen wir dann so Spielfiguren drauf.

[00:58.750-01:00.990] EX Und wenn diese Spielfiguren eben diese Viererreihe bilden

[01:01.630-01:02.910] EX dann haben wir das Ziel erreicht

[01:02.910-01:08.130] EX und die Spielfiguren haben verschiedene Eigenschaften

[01:09.050-01:13.770] EX also die sind entweder nach Größe nach Farbe nach Form

[01:13.770-01:17.070] EX und ob die so eine Kuhle oben haben oder nicht.

[01:18.622-01:18.851] EE Okay.

[01:19.400-01:20.020] EX Also zum Beispiel...

[01:20.010-01:22.050] EE Kann ich mir jetzt nicht so merken aber ja.

[01:22.050-01:24.090] EX Ja also... Du hast zum Beispiel so eine Figur

[01:24.090-01:25.920] EX und dann ist sie zum Beispiel klein

[01:27.228-01:27.491] EX hell,

[01:28.580-01:29.310] EX hat diese Kuhle

[01:29.310-01:31.780] EX und ist rund.

[01:32.240-01:33.880] EX Also sie können rund oder vierkig sein.

[01:35.150-01:37.050] EX Und wenn du die dann hinstellen würdest

[01:37.050-01:38.600] EX und diese Viererreihen versuchst

[01:38.660-01:41.640] EX dann muss halt in dieser Viererreihe eine Eigenschaft stimmen.

[01:42.010-01:43.820] EX Also zum Beispiel alle Figuren müssen rund sein

[01:43.820-01:46.040] EX oder alle Figuren müssen klein sein oder groß.

[01:47.010-01:47.013] EX (*empty*)

[01:47.013-01:49.165] EX Und zum Beispiel wenn du [die dann da hinstellen genau.]

[01:47.778-01:49.095] EE [Oder die gleiche Farbe haben.]

[01:49.677-01:51.095] EX Und wenn du die dann da hinstellen würdest

[01:51.190-01:56.418] EX dann könnte in der Reihe zum Beispiel ja nur noch klein mit Kuhle hell

[01:56.418-01:57.316] EX und ja genau.

[02:01.829-02:03.930] EX Und der Spielverlauf ist so

[02:04.323-02:06.930] EX dass quasi wir teilen uns die Spielfiguren

[02:07.190-02:07.987] EX das sind 16 Stück.

[02:08.867-02:12.380] EX Es gibt quasi von jeder Eigenschaft Kombination eine

[02:12.380-02:16.905] EX und dann auch immer so ein Pendant quasi.

[02:17.715-02:19.987] EX Also eins ist dann anstatt groß klein

[02:20.260-02:21.070] EX anstatt hell dunkel

[02:21.380-02:22.399] EX anstatt rund eckig

[02:22.399-02:23.829] EX und hat eine Kuhle oder keine Kuhle.

[02:27.335-02:28.772] EE die vier Eigenschaften dann halt.

[02:28.880-02:30.006] EX Genau genau das Gegenteil.

[02:30.820-02:30.982] **EX** Ja.

[02:33.703-02:35.981] **EX** Und wir tun die quasi immer aufs Spielfeld.

[02:36.120-02:37.506] **EX** Und das Besondere ist

[02:37.627-02:41.038] **EX** dass du mir eine Figur aussuchen musst

[02:41.038-02:42.101] **EX** und ich muss sie dann stellen.

[02:42.329-02:43.747] **EX** Und dann suche ich dir eine Figur aus

[02:43.804-02:44.582] **EX** dann musst du die stellen.

[02:44.984-02:45.284] **EE** Okay.

[02:48.006-02:50.203] **EE** Du kannst dann quasi mit der Figur

[02:50.222-02:51.544] **EE** die ich dir aussuche gewinnen.

[02:52.323-02:53.127] **EX** Genau. genau.

[02:53.481-02:54.684] **EX** Aber wir teilen uns ja alles.

[02:54.684-02:59.411] **EX** Deswegen das... kommt jetzt auch so ein bisschen zu so Strategien.

[02:59.620-03:01.570] **EX** Es gibt so defensives oder offensives Spielen.

[03:03.747-03:05.171] **EX** wenn du so sehr defensiv spielst

[03:05.190-03:08.190] **EX** dann versuchst du einfach die ganze Zeit diese ganzen Möglichkeiten so zu blocken.

[03:09.044-03:09.975] **EX** Das kannst du halt zum Beispiel gut

[03:09.994-03:12.753] **EX** wenn du diese Pendant -Figur einfach in so eine Reihe reinstellst

[03:12.810-03:14.329] **EX** weil dann ist die ja quasi zu...

[03:14.785-03:15.234] **EX** oder tot

[03:15.519-03:18.190] **EX** weil dann kannst du das ja gar nicht mehr schaffen die Reihe zu...

[03:18.557-03:20.411] **EE** so ein bisschen das Prinzip von Tic -Tac -Toe

[03:20.411-03:22.051] **EE** dann quasi in die Reihen machen.

[03:22.228-03:22.829] **EX** Ja genau.

[03:23.203-03:23.633] **EX** Genau ja.

[03:25.297-03:28.057] **EX** Und dann weil das Schwierigste bei dem Spiel ist

[03:28.228-03:29.354] **EX** die Übersicht zu behalten.

[03:29.709-03:31.158] **EX** Weil so viele Eigenschaften sind

[03:31.158-03:33.076] **EX** und dann irgendwann ist das Spielfeld ja voll

[03:33.076-03:36.487] **EX** und dann siehst du nicht dass sich da gerade eine kleine Reihe gebildet hat

[03:36.487-03:37.519] **EX** oder dass die eine Kuhle haben

[03:37.595-03:39.835] **EX** weil du dann nur noch auf zwei Eigenschaften achtest.

[03:40.449-03:41.956] **EX** Und deswegen wenn du das so defensiv machst

[03:42.076-03:44.975] EX dann kannst du das halt so ein bisschen minimieren das Risiko.

[03:48.297-03:50.633] EX Aber gleichzeitig kannst du dann halt auch sehr offensiv spielen

[03:50.715-03:53.696] EX indem du einfach so viele Möglichkeiten wie möglich so erstellst.

[03:54.025-03:56.329] EX Und einfach hoffst dass die andere Person es übersieht

[03:56.329-03:59.354] EX und du halt dann in dem Moment in der Möglichkeit bist

[03:59.627-04:01.013] EX die Reihe zu schließen.

[04:03.323-04:06.373] EX Und du kannst dann auch dadurch quasi so Fallen stellen.

[04:07.487-04:09.171] EX Weil du darfst ja die Figur aussuchen.

[04:10.234-04:13.652] EX Und du kannst dann so probieren dass zum Beispiel

[04:13.652-04:16.949] EX ab einem bestimmten Punkt einfach die Figuren nicht mehr da sind.

[04:19.019-04:19.209] EX Genau.

[04:19.222-04:19.348] EE Ja.

[04:20.867-04:21.854] EX Habe ich sonst noch was vergessen?

[04:23.747-04:25.405] EX Du darfst die Figur nicht umstellen

[04:26.222-04:27.487] EE Ja macht auch Sinn ja. []

[04:26.525-04:27.184] EX macht [Sinn?]

[04:28.260-04:28.519] EX Ja?

[04:29.146-04:30.462] EX Ich dachte man darf die umstellen.

[04:33.342-04:35.025] EE könnte auch sein wie bei Schach oder sowas.

[04:35.165-04:35.804] EX Ja genau.

[04:36.399-04:37.703] EX Weil es ist auch so von den Farben

[04:37.741-04:38.905] EX es hat so Holzfarben

[04:38.981-04:40.203] EX es hat so helles Holz und dunkles Holz

[04:40.253-04:41.304] EX sieht so ein bisschen aus wie Schach.

[04:42.677-04:43.880] EX Es ist ja auch so ein bisschen komplex

[04:43.880-04:46.247] EX von der Komplexität wie Schach.

[04:49.108-04:50.285] EX Was kann man sonst noch sagen?

[04:53.842-04:54.595] EX Es ist immer voll gut

[04:54.690-04:57.861] EX wenn du die Figuren so an den Rand stellst und in die Mitte.

[04:58.589-05:00.956] EX Wenn du halt dieses Offensive machst

[05:01.044-05:02.760] EX weil du dann so viele Möglichkeiten hast.

[05:04.468-05:06.095] EE Wird das egal

[05:06.095-05:08.880] EE ob halt quer längs diagonal...

[05:08.873-05:12.057] EX Genau die vierer -Reihen sind diagonal waagrecht senkrecht.

[05:13.000-05:13.703] EX Dann hast du gewonnen.

[05:14.563-05:16.095] EX Deswegen muss man dann auch noch mehr achten.

[05:16.399-05:18.715] EX Jetzt kannst du ja sogar noch eine Diagonale bilden.

[05:19.506-05:21.146] EE Deswegen ist die Mitte dann halt gut

[05:21.323-05:21.981] EE aber irgendwie auch schlecht

[05:22.076-05:23.063] EE weil man es sehr schnell verpasst,

[05:23.165-05:23.399] EX Ja genau.

[05:23.867-05:24.563] EX Oder halt die Ecken

[05:24.747-05:25.601] EX das ist ja das Gleiche dann.

[05:29.260-05:30.316] EX Ja, oder du bist halt gemein

[05:30.316-05:34.842] EX und wenn du siehst dass in der Mitte so bestimmte Figurenkombinationen stehen

[05:35.335-05:38.367] EX dann kannst du ja auch versuchen da so jetzt eine hinzustellen

[05:38.462-05:39.405] EX die das so ein bisschen kaputt macht.

[05:39.844-05:40.084] EE Ja.

[05:41.044-05:43.886] EE Wobei ich bedenken muss dass du die hinstellst

[05:43.975-05:44.943] EE die ich dir aussuche.

[05:45.911-05:46.241] EX Ja, genau.

[05:46.241-05:48.532] EX Also das ist ja auch nochmal... das Alles was du kaputt machst

[05:48.608-05:49.810] EX machst du dir auch selber kaputt.

[05:50.532-05:51.677] EX Und was du dir aufbaust

[05:51.728-05:52.734] EX baust du auch mir mit auf.

[05:54.760-05:58.329] EE Man muss ja dann quasi immer zwei Züge so vorausdenken.

[05:59.310-06:03.810] EX und selbst wenn du dir zum Beispiel eine Dreierreihe bilden würdest

[06:04.373-06:06.114] EX du musst mir ja immer noch die Figur geben

[06:06.165-06:07.171] EX damit ich sie beenden kann.

[06:07.172-06:07.732] EE Ja

[06:07.975-06:12.272] EE aber du gewinnst dann quasi wenn...

[06:12.297-06:13.367] EE Ich die setze

[06:13.367-06:15.146] EE oder wenn ich dir die Richtige gebe?

[06:16.418-06:17.146] EX Ah okay

[06:17.443-06:18.146] EX das hab ich auch vergessen.

[06:18.981-06:19.981] EX Also wenn ich die bekomme

[06:19.981-06:20.981] EX und dann gesetzt habe

[06:21.386-06:22.804] EX dann muss ich das quasi erkennen

[06:23.367-06:24.722] EX dass diese Viererreihe sich gebildet hat

[06:24.772-06:26.943] EX weil es kann ja auch sein dass ich das in dem
Kuddelmuddel nicht sehe.

[06:28.038-06:29.082] EX Und das ist wie UNO

[06:29.304-06:31.367] EX also du musst dann einmal so sagen quattro.

[06:31.367-06:32.601] EX Und wenn du es nicht gesagt hast

[06:32.703-06:34.272] EX dann ist das auch nicht dein Siegpunkt.

[06:34.525-06:34.905] EE Ahh, okay.

[06:35.411-06:36.519] EX Also dann könntest du zum Beispiel

[06:36.595-06:37.519] EX wenn du das jetzt gesehen hast

[06:37.557-06:38.582] EX dass du die Viererreihe gebildet hast

[06:38.677-06:39.722] EX dann wäre es wenn du dran wärst

[06:39.722-06:40.519] EX und das Wort sagen würdest

[06:40.601-06:41.418] EX dann wäre es dein Sieg.

[06:41.703-06:43.873] EE Okay also es ist egal wer die gesetzt hat.

[06:43.899-06:45.481] EX Genau also ob [du es siehst die Viererreihe]

[06:44.329-06:45.962] EE [(unverständlich).]

[06:45.960-06:46.120] EX ja.

[06:46.848-06:48.734] EE Also kann ich dir auch quasi die richtige geben

[06:48.930-06:50.158] EE hoffen du setzt sie da hin []

[06:50.089-06:50.354] EX [Ja.]

[06:50.158-06:51.823] EE und schneller einfach rufen.

[06:52.140-06:52.320] EX Ja.

[06:53.057-06:55.120] EX Nee also wenn ich noch dran bin

[06:55.380-06:56.943] EX wäre das quasi wenn ich das dann erkenne

[06:57.133-06:57.639] EX wäre es dann mein.

[06:57.715-06:59.120] EX Aber wenn du dann dran kommst

[06:59.120-07:00.658] EX und ich habe es nicht erkannt

[07:00.899-07:01.918] EX dann wäre es dein Sieg.

[07:02.340-07:02.660] EE Okay.

[07:03.020-07:03.200] EX Ja.

[07:04.342-07:06.690] EE Also hast du schon den Vorteil in dem Moment wo du
setzt?

[07:07.304-07:08.310] EX Mhm. Ja.

[07:11.544-07:12.633] EX Kann man sonst noch...

[07:15.734-07:18.532] EX Ja, es ist halt so ist schon komplex

[07:19.342-07:21.114] EX weil du halt auf die ganzen Sachen achten musst.

[07:21.760-07:23.892] EX Aber gleichzeitig ist es so simpel

[07:24.424-07:24.930] EX so ein Mix.

[07:30.753-07:31.854] EX Habe ich irgendwas vergessen?

[07:33.551-07:34.930] EX Du weißt die Eigenschaften

[07:36.627-07:37.690] EX Du weißt den Aufbau

[07:37.785-07:39.468] EX du weißt wie das Spiel verläuft.

[07:41.690-07:42.373] EX Hast du noch Fragen?

[07:42.867-07:43.652] EE Ich glaube gerade nicht.

[07:44.000-07:44.690] EE Ich finde halt oft,

[07:44.690-07:46.753] EE dass generell ist das so bei Spielen,

[07:46.753-07:49.013] EE dass mans erst beim Spielen dann selbst merkt.

[07:49.342-07:49.810] EX Das stimmt.

[07:50.399-07:51.013] EX Ja das stimmt.

[07:53.816-07:57.892] EE Das ist eine Mischung aus Vier -Gewinnt, Tic -Tac -Toe und keine Ahnung was.

[07:58.405-07:59.291] EX So ein bisschen Mühle

[07:59.411-08:01.671] EX weil du kannst ja auch so Fallen und so einsetzen.

[08:04.965-08:05.065] EX Ja,

[08:05.025-08:06.209] EX weil so zum Beispiel wenn du

[08:07.114-08:07.911] EX Du kannst auch so machen

[08:08.006-08:12.272] EX dass du die ganze Zeit versuchst quasi so eine Eigenschaft abzuziehen

[08:12.778-08:14.304] EX also dass die dann irgendwann leer ist

[08:14.304-08:18.570] EX und dass dann quasi

[08:19.006-08:22.614] EX dass du dann mit der anderen Eigenschaft diese Dreierreihe bildest

[08:22.614-08:25.165] EX und dass die andere Person ja nur noch diese Figur geben kann.

[08:25.672-08:25.912] EE Ja.

[08:26.994-08:29.734] EX Aber da musst du schon sehr... [Ja]

[08:28.013-08:29.475] EE viel [nachdenken.]

[08:31.215-08:32.032] EE Ich habe es verstanden

[08:32.044-08:34.842] EE aber ich... weiß trotzdem gerade nicht so wie ichs spielen würde,

[08:34.943-08:36.943] EE aber ja

[08:37.519-08:40.127] EE Ich glaube da hilft auch keine weitere Erklärung.

[08:40.418-08:42.304] EE Ich glaube ich würde makgucken wie sich das ergibt.

[08:42.984-08:43.164] EX Ja.

[08:43.892-08:45.386] **EE** Ich weiß auch nicht wirklich was schlauer ist.

[08:45.399-08:48.791] **EE** eher defensiv oder... [offensiv.]

[08:48.076-08:53.462] **EX** Ich [glaube das ist so] von der Art
und Weise wie aufmerksam du bist.

[08:53.944-08:54.224] **EE** Ja.

[08:55.019-08:57.684] **EX** Ich glaube wenn du so sehr auf alles achten kannst

[08:57.804-08:58.823] **EX** dann würde ich eher offensiv spielen

[08:58.943-09:01.228] **EX** aber wenn du... Angst hast dass du irgendwas übersiehst

[09:01.228-09:02.177] **EX** und wenn du ehh immer (unverständlich)

[09:02.234-09:03.633] **EX** dann würde ich das eher definitiv spielen.

[09:04.096-09:04.336] **EE** Ja.

[09:06.177-09:07.196] **EX** Aber es ist halt wirklich so

[09:07.266-09:08.975] **EX** weil es ist einfach du gibst mir den Karte

End of standardisation

09:07.266-09:08.975

[09:09.177-09:11.057] **EX** Ich stelle. Stellen stellen stellen

[09:11.399-09:14.038] **EX** irgendwann... glaube ja.

[09:17.000-09:18.614] **EE** Ich glaube ich habe es schon verstanden.

[09:18.736-09:19.016] **EX** Okay.

[09:21.658-09:22.614] **EX** Weißt du was wir jetzt machen sollen?

[09:23.008-09:24.916] **EE** Ich " glaube wir sollen ihr " sagen dass wir fertig sind.

[09:26.696-09:27.056] **EX** Wie?

[09:27.876-09:28.776] **EE** so. Hallo.

[09:31.085-09:32.300] **EE** Hat sie gut erklärt.

[09:35.360-09:36.620] **EX** Ich hoffe - Ich habe nichts vergessen.

[09:36.937-09:39.440] **EE** Ich weiß gar " nicht ob sie uns auch hört. Ich " glaube sie
sieht uns nur.

[09:39.920-09:40.040] **EX** Ah.

[09:41.020-09:43.700] **EE** - Aber die anderen Kameras zeichnen auch den Ton auf.

[09:45.740-09:47.260] **EX** Macht aber auch Sinn bei Erklärungen.

[09:47.380-09:48.100] **EE** Ja, schon.

D Understanding Instrument (Building Block 4)

This chapter provides an overview of the understanding instrument used in BB4 and developed in BB2 and BB3. The detailed development documentation can be found at https://osf.io/w39dc/overview?view_only=c076824465c9479481a6f5ee5ea9d65c (Terfloth and Schaffer, 2025).

The instrument assesses two dimensions of understanding following Buschmeier et al. (2025): *comprehension* (knowing-that) and *enabledness* (knowing-how). Development proceeded in four stages: an initial structured interview guide used to generate and validate item content (2022), followed by three iterative questionnaire versions administered across independent validation samples before the final version was deployed in the BB4 experiment.

D.1 Structured Interview Guide (Pilot, 2022)

The structured interview (*Leitfadeninterview*), developed in early 2022, served as the conceptual basis for item generation. Semi-structured interviews were conducted with participants after an explanation of the game *Quarto!* (pre-test) and again after a play session (post-test). Item IDs follow the pattern **A** (Architecture), **R** (Relevance), **S** (opening question), and **Z** (Miscellaneous questions). Horizontal rules indicate thematic blocks within each section. The guide follows the *Leitfaden-schema* described in Helfferich (2014).

Interview metadata recorded per session: Interview-Nr., Ort, Datum, Beginn, Ende, Interviewer*in, notes on pre-recording conversation.

Pre-test (after explanation, before gameplay)

Table D.1: Structured interview guide — pre-test section.

ID	Question	Content aspects / keywords	Follow-up questions
So1	Wie ging es dir mit der Erklärung, was hast du daraus mitgenommen?		
Ao1	Welche Spielkomponenten sind vorhanden und wie würdest du die beschreiben?	Spielbrett, Formen der Spielsteine, Farbe der Spielsteine	Welche Formen haben die Spielsteine? Wie ist das Spielbrett aufgebaut? Welche Farben haben die Spielsteine?
Ao3	Wie dürfen die Komponenten im Spiel eingesetzt werden?	Setzen von Steinen	Wie werden Spielsteine eingesetzt?
Ao4	Wie wird entschieden, wer das Spiel beginnt?		
Ro1	Warum haben die Komponenten bestimmte Eigenschaften?		
Ro2	Welchen Zweck hat es, dass die Komponenten aus einem bestimmten Material sind?		
Ro3	Welchen Zweck haben die verschiedenen Komponenten?	Sinn/Zweck der Farben, Sinn/Zweck der Formen	Ist es wichtig, dass die Spielsteine die Farben haben? Ist es wichtig, dass die Spielsteine die Formen haben?
Ro4	Was ist der Grund dafür, dass die Regel zur Entscheidung über den Spielbeginn angewendet wird?		
Ao5	Welche Regeln gibt es in dem Spiel?	Zugreihenfolge, Einsatz von Figuren	Wie können Spielfiguren eingesetzt werden? Wie ist der Ablauf einer Runde im Spiel?
Ao6	Welche Spielzüge sind möglich?	Spielerwechsel	Wenn man an der Reihe ist: Wie läuft dann eine Spielrunde ab?

(Table D.1 continued)

ID	Question	Content aspects / Follow-up questions keywords
A07	Wann hat ein Spieler gewonnen?	
Ro5	Inwiefern sind die einzelnen Regeln des Spiels bedeutsam bzw. welchen Sinn haben sie?	
Ro6	Inwiefern sind die unterschiedlichen Spielzüge bedeutsam?	
Ro7	Worauf kann zu Beginn des Spiels geachtet werden?	
A09	Welche Spielzüge könnten einen Sieg ermöglichen?	
A10	Wie würdest du das Regelwerk als Gesamtpaket beschreiben?	
A11	Wie würdest du das gesamte Spiel beschreiben?	
Ro8	Wie würdest du spielen, wenn du gewinnen möchtest?	
Ro9	Welchen Sinn erkennst du hinter dem Regelwerk?	
R10	Wie würdest du das Spiel charakterisieren?	
R11	Welche Gemeinsamkeiten mit anderen Spielen erkennst du?	
R12	Welche Spielsituationen erhöhen die Gewinnchance?	
Z1	Nenne Fehler, die ein Anfänger begehen könnte.	
Z2	Welche Analogien könnten dir helfen das Spiel zu erklären?	

(Table D.1 continued)

ID	Question	Content aspects / keywords	Follow-up questions
Z3	Was ist für dich unklar geblieben? Gab es Unstimmigkeiten?		
Z4	Wie könnte das Spiel besser gemacht werden?		

The pre-test also included three image-based situation tasks presented on paper:

1. Identify which of the presented game board scenarios constitute a Quarto (multiple correct answers possible; verbal justification required).
2. Select one piece from a set that would complete a Quarto with three shown pieces; justify the choice.
3. Select a piece to give to the opponent such that they *cannot* form a Quarto; justify the choice.

Post-test (after gameplay)

The post-test followed the same thematic structure as the pre-test and used the same item IDs. Item Ao1 (component description) was omitted on the grounds that participants had since handled the game physically. All remaining items (Ao3–A11, Ro1–R12, Z1–Z4) were repeated verbatim, along with the three image-based situation tasks.

D.2 Questionnaire Version 3 — Final Version Used in Study

This is the version deployed in the BB4 experiment. Three changes were made relative to Version 2: (1) the response format for both subscales was changed from a 5-point Likert scale to a binary *Stimme zu* / *Stimme nicht zu* format, motivated by the results of an exploratory factor analysis showing that dichotomous scoring improved subscale structure; (2) the Architecture subscale was expanded to 13 items and the Relevance subscale was also set to 13 items through targeted addition and revision; (3) the Quarto-identification task in the enabledness section was replaced by an attention check item, and the set of board-position options in the placement task was reduced from nine to five.

Willkommen bei unserer Umfrage! Dieser Fragebogen enthält einige Fragen, die Sie in weniger als 10 Minuten beantworten können. Für die Antworten brauchen Sie keinerlei Vorkenntnisse, deshalb beantworten Sie bitte alle Fragen spontan und offen. Die im Fragebogen gewonnenen Daten werden nach den gesetzlichen Datenschutzbestimmungen erfasst und absolut vertraulich behandelt. Einzeldaten werden nur in statistisch zusammengefasster Form dargestellt.

Wir danken vielmals für Ihre Teilnahme!

Teilnahmeinformationen (Participation Information)

1. Welche Versuchsnummer hat der Versuch? (*Short text, max. 3 characters; e.g., 1, 15, or 123.*)
2. Welche Rolle haben Sie? (*Single choice.*)
 - Explainer – Ich kenne und erkläre gleich das Spiel
 - Explaineer – Ich bekomme das Spiel gleich erklärt

Teil 1 — Architektur (Architecture Subscale)

Im folgenden Abschnitt des Fragebogens geht es inhaltlich um Ihre Beschreibung und Bewertung des Spiels, der Komponenten und Regeln. Hierfür werden verschiedene Aussagen präsentiert, die Sie bewerten sollen. Falls sich Aussagen stark ähneln, lassen Sie sich hiervon nicht verwirren. Beantworten Sie jede Aussage offen und spontan.

Bewerten Sie die folgenden Aussagen.

Scale: Stimme zu / Stimme nicht zu

Statement	Stimme zu	Stimme nicht zu
Man muss sich um die eigenen Reihen, die man angefangen hat, kümmern.		
Man darf eine Reihe, die die andere Person angefangen hat, vervollständigen und so gewinnen.		
Es ist erlaubt eine Reihe, die die andere Person angefangen hat, zu verbauen und so eine Niederlage zu vermeiden.		
Man sucht sich die Figur, die man setzt, selbst aus.		
Man darf Figuren beliebig auf freien Feldern des Bretts positionieren.		

Statement	Stimme zu	Stimme nicht zu
Man darf auch diagonal ein Quarto bilden.		
Alle Figuren haben verschiedene Größen.		
Alle Figuren haben verschiedene Farben.		
Alle Figuren haben verschiedene Formen.		
Das Spielbrett hat eine gerade Anzahl an Feldern.		
Insgesamt gibt es zehn Möglichkeiten, um ein Quarto auf dem Spielbrett zu bilden.		
Eine Person spielt die hellen und eine Person die dunklen Figuren.		
Man kann nur mit seinen eigenen Reihen ein Quarto bilden.		

Teil 2 — Relevanz (Relevance Subscale)

Im folgenden Abschnitt des Fragebogens geht es inhaltlich um Ihre Interpretation der Hintergründe des Spiels, Besonderheiten der Komponenten und subjektiver Einschätzung des Spiels. Hierfür werden verschiedene Aussagen präsentiert, die Sie bewerten sollen. Falls sich Aussagen stark ähneln, lassen Sie sich hiervon nicht verwirren. Beantworten Sie jede Aussage offen und spontan.

Bewerten Sie die folgenden Aussagen.

Scale: Stimme zu / Stimme nicht zu

Statement	Stimme zu	Stimme nicht zu
Da die Figuren sich sehr ähnlichsehen, ist es schwierig, den Überblick zu behalten.		
Dass man die Figur für die andere Person wählt, sorgt für das gewisse Etwas des Spiels.		
Wenn man gewinnen möchte, sollte man unbedingt beobachten, was die andere Person vor hat.		
Das eigene Spiel wird von den Entscheidungen der anderen Person stark beeinflusst.		
Quarto ist ein recht strategisches Spiel.		
Wer nicht weit genug vorausdenkt, wird Quarto nicht gut spielen können.		
Man sollte sich um seine eigenen Reihen kümmern.		

Statement	Stimme zu	Stimme nicht zu
Meist ist es sinnvoll die Reihe, die die andere Person angefangen hat, zu verbauen.		
Wenn die Möglichkeit besteht, sollte man Quarto bilden.		
Wenn man die zu setzende Figur selbst wählen dürfte, wäre Quarto weniger spannend.		
Stehen viele Figuren auf dem Feld, ist es schwerer den Überblick zu behalten.		
Besonders schwierig ist es, die richtige Figur für die andere Person zu wählen.		
Als Profi kann es sinnvoll sein, Reihen nicht direkt zu verbauen, um Fehler zu provozieren.		

Teil 3 — Enabledness (Enabledness Tasks)

Im folgenden Abschnitt des Fragebogens geht es inhaltlich um Ihre Spielfähigkeit. Hierfür werden verschiedene Spielsituationen präsentiert. Beantworten Sie jede Aussage offen und spontan.

1. **Figurenauswahl / Piece completion A.** Wählen Sie die Figur aus, die ein Quarto mit den drei abgebildeten Figuren bilden würde.

Abgebildete Figuren:



Antwortoptionen:



(A)

(B)

(C)

2. **Figurenauswahl / Piece completion B.** Wählen Sie die Figur aus, die ein Quarto mit den drei abgebildeten Figuren bilden würde.

Abgebildete Figuren:



Antwortoptionen:



(A)

(B)

(C)

3. **Strategische Auswahl / Strategic piece selection A.** Wählen Sie eine Figur aus, welche Sie der anderen Person für den nächsten Zug geben würden.

Spielsituation:



Antwortoptionen:



(A)

(B)

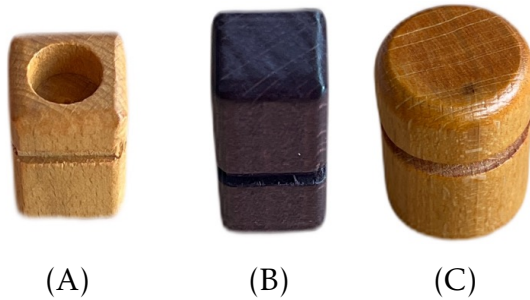
(C)

4. **Strategische Auswahl / Strategic piece selection B.** Wählen Sie eine Figur aus, welche Sie der anderen Person für den nächsten Zug geben würden.

Spielsituation:

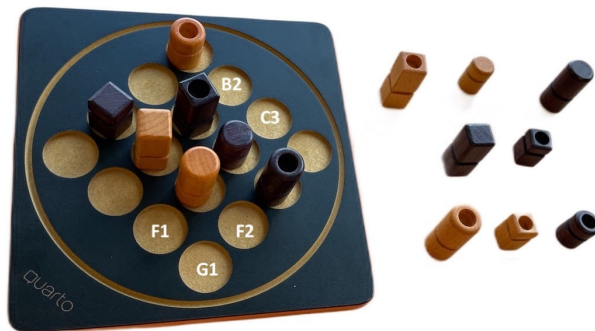


Antwortoptionen:



5. **Platzierungsentscheidung / Placement decision.** Gegeben ist folgende Spielsituation. Sie bekommen von der anderen Person die Figur  (klein, eckig, hell, ohne Loch) gegeben. Wohin würden Sie die Figur setzen?

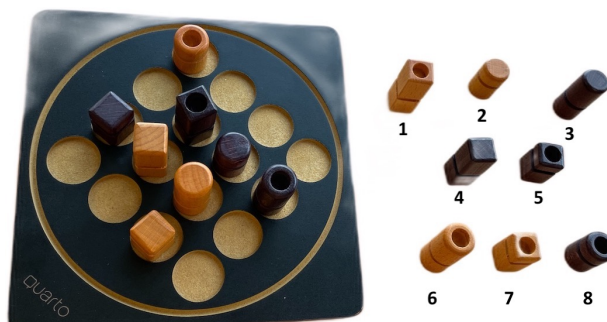
Spielsituation:



Antwortoptionen (Koordinaten): B2, C3, F1, F2, G1.

6. **Figur-für-andere-Person / Piece-to-opponent decision.** Gegeben ist folgende Spielsituation. Welche Figur würden Sie der anderen Person geben?

Spielsituation:



Antwortoptionen: 1, 2, 3, 4, 5, 6, 7, 8.

7. **Aufmerksamkeitscheck / Attention check.** Um zu überprüfen, ob der Browser richtig funktioniert und ob wir alle Ihre Antworten erfassen, wählen Sie bitte die Antwortoption "Helle Figur" aus der folgenden Liste aus.

Antwortoptionen: Helle Figur Dunkler Figur Eckige Figur Runde Figur.

E Supplementary Material by Building Block

The following sections provide corpus overviews, content structure visualisations, and exploratory analyses for each building block. All figures were generated from the Jupyter notebooks archived in the OSF repository referenced in Appendix A.

E.1 BB1 — Transcription in Computing Education Research

The systematic literature review materials are archived at https://osf.io/8ejcu/overview?view_only=b0c48ee77ce24010be6048c4a198a9c9.

E.2 BB2 — Adding Why to What

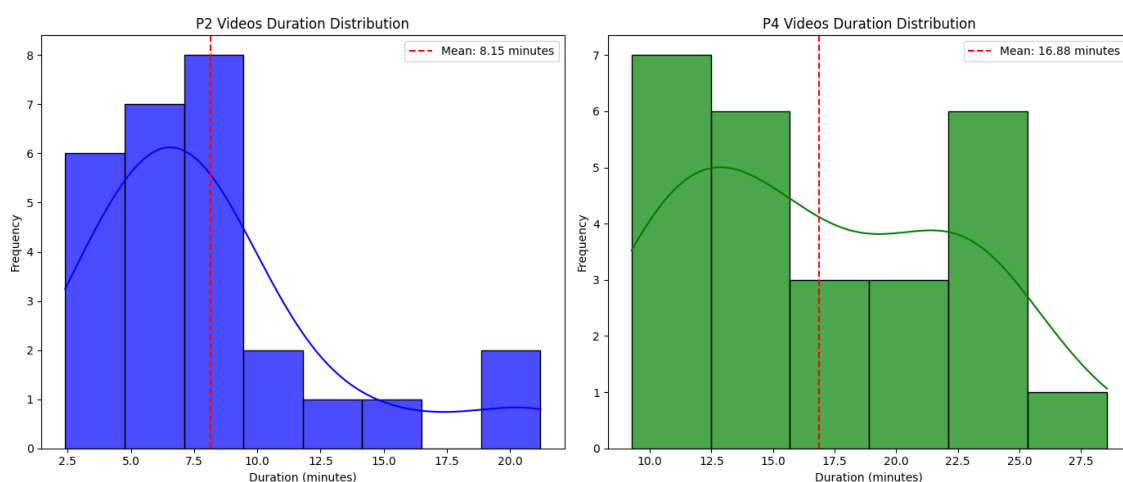


Figure E.1: BB2 corpus duration overview. Only P2 data was analysed in this dissertation as specified in Chapter 3.



Figure E.2: BB2 content structure overview of all experiments used in the paper.

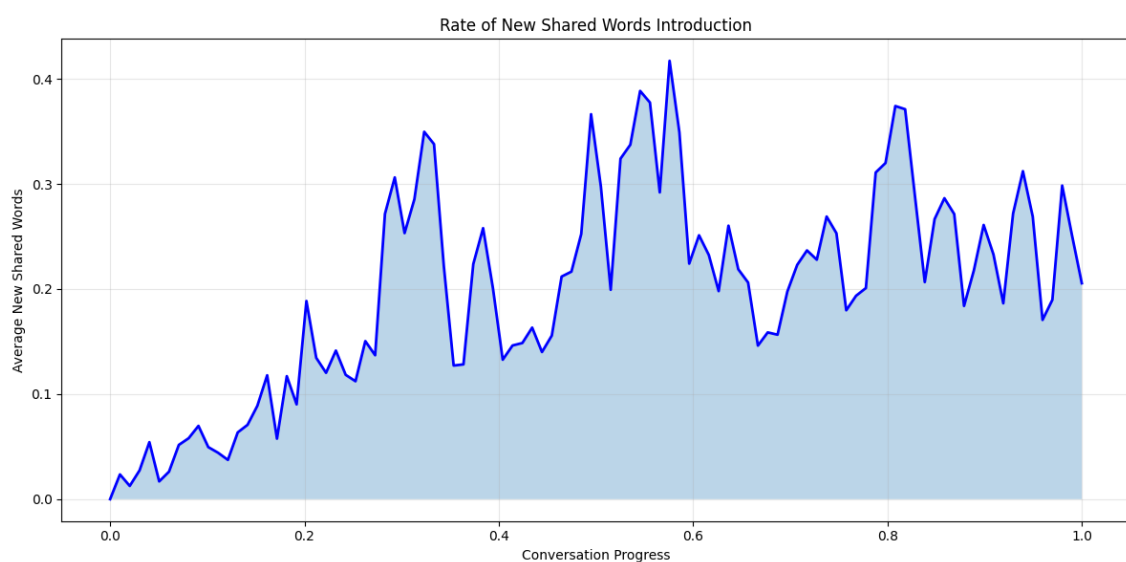


Figure E.3: Exploratory analysis. New word introduction rate (combined)

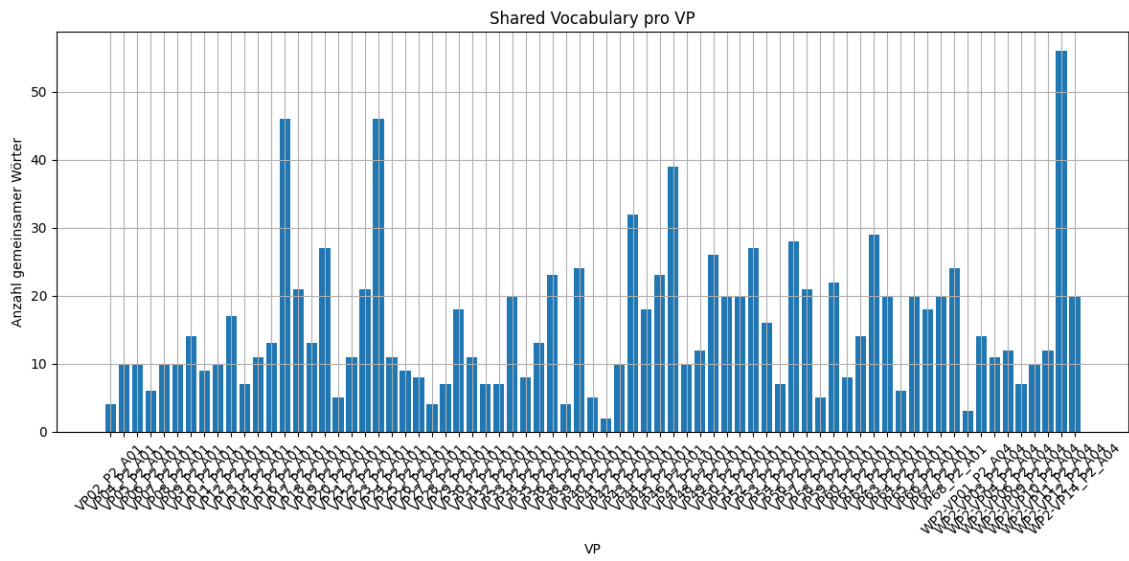
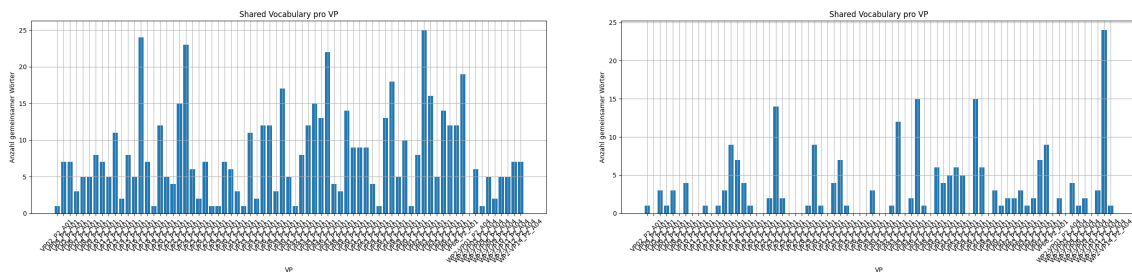


Figure E.4: Shared vocabulary (combined)



(a) Shared vocabulary – A

(b) Shared vocabulary – R

Figure E.5: Shared vocabulary by group

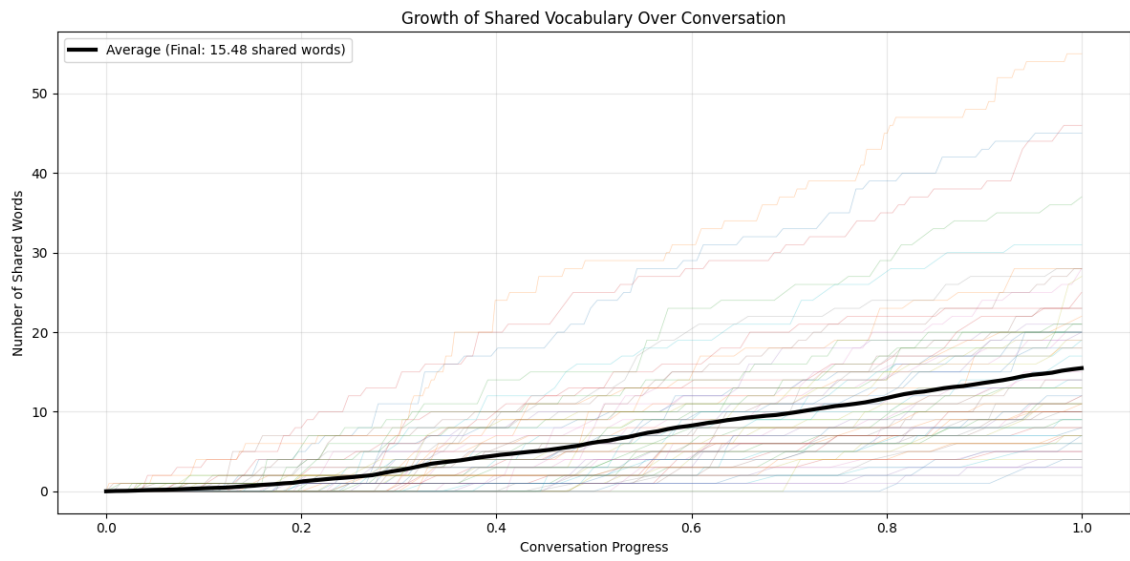
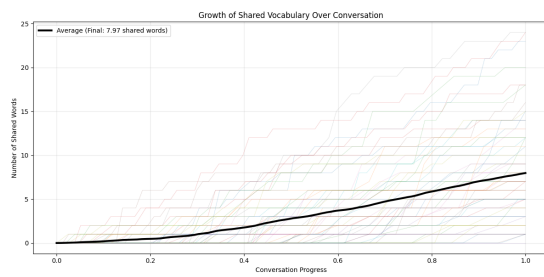
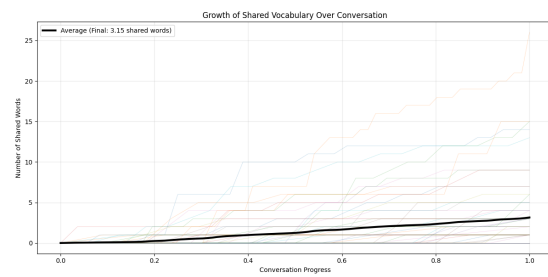


Figure E.6: Shared vocabulary evolution (combined)



(a) Shared vocabulary evolution – A



(b) Shared vocabulary evolution – R

Figure E.7: Shared vocabulary evolution by group

E.3 BB₃ — Navigating the Dual Nature

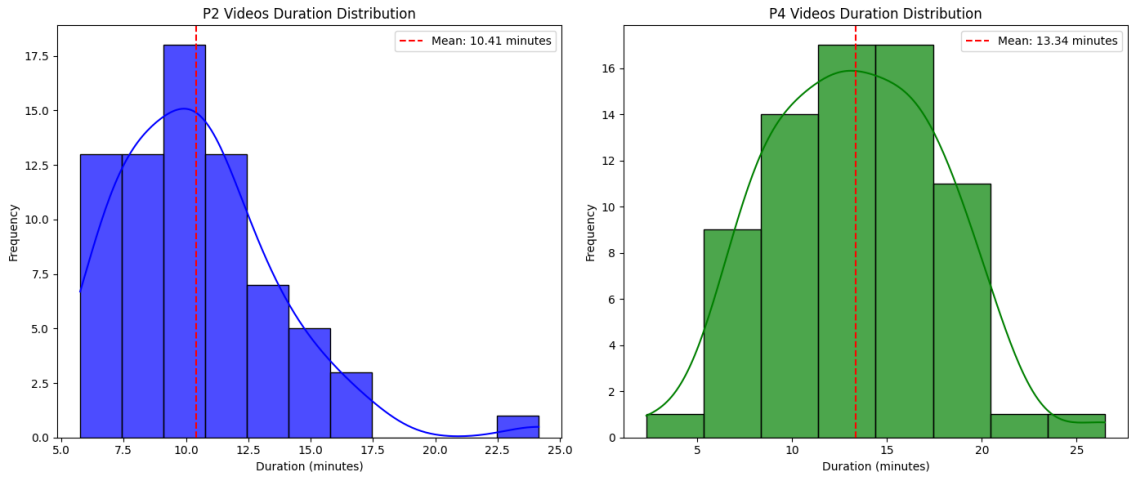


Figure E.8: BB₃ corpus overview

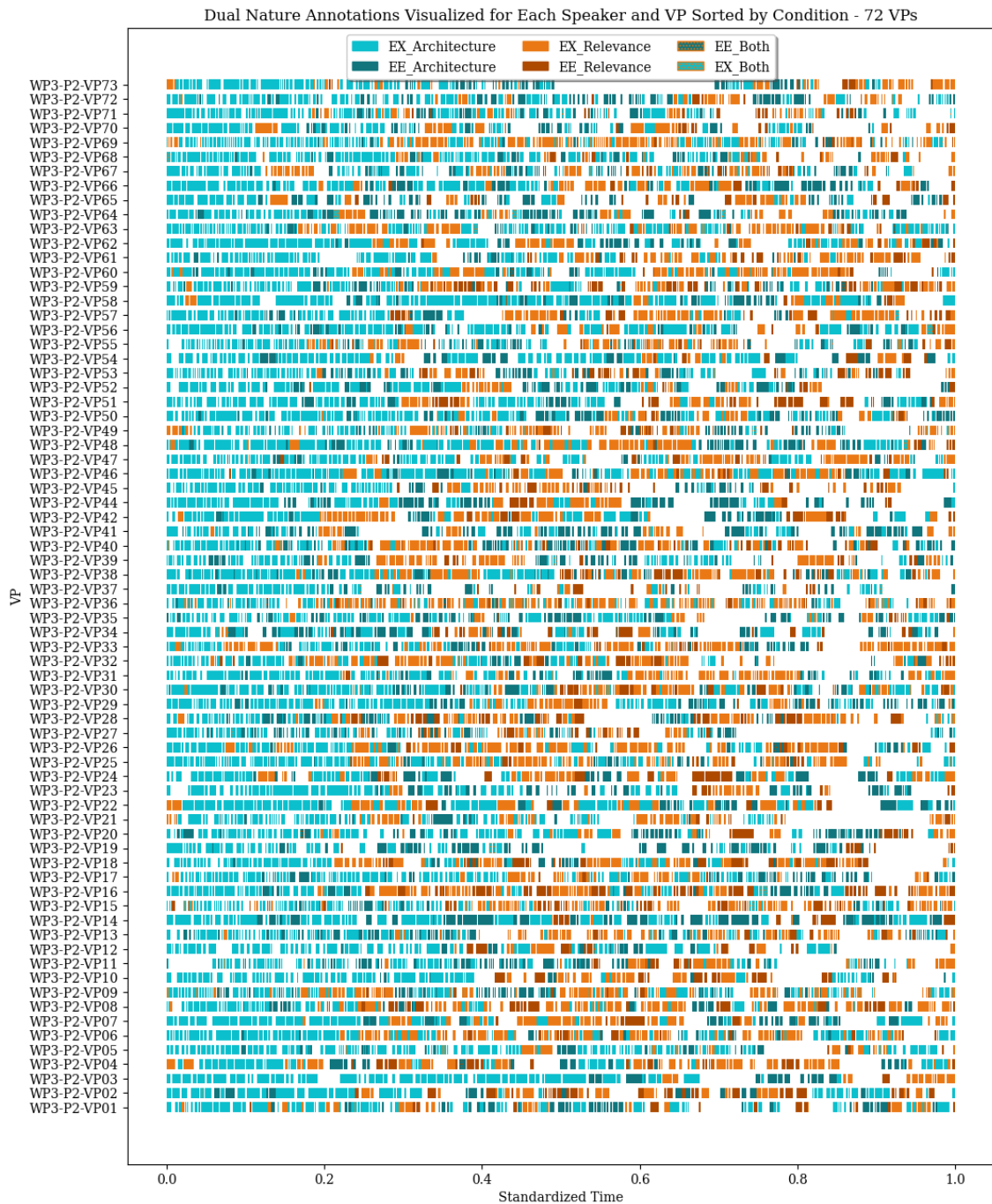


Figure E.9: BB₃ all VPs

E.4 BB4 — Bridging the Dual Nature

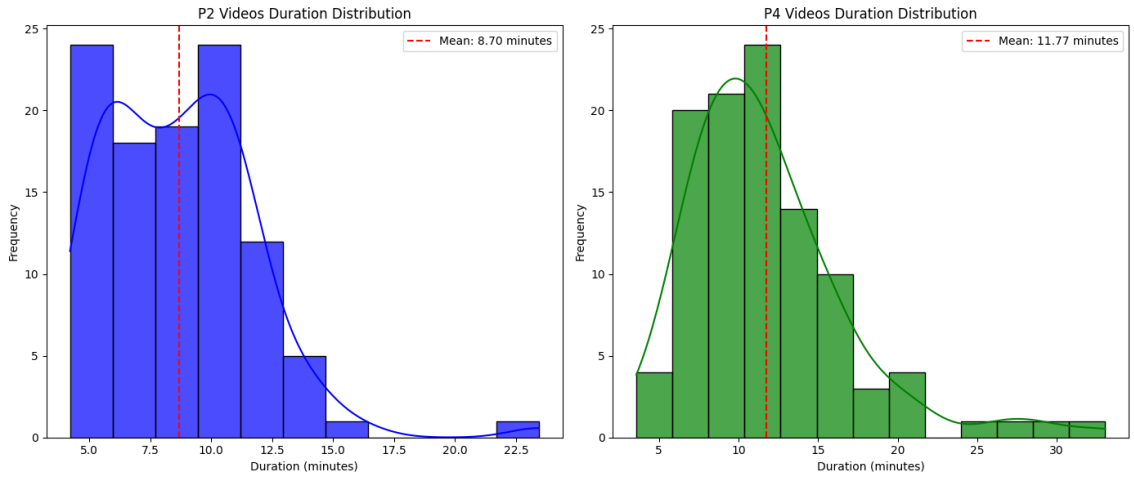


Figure E.10: BB4 corpus overview (notebook cell oo6)

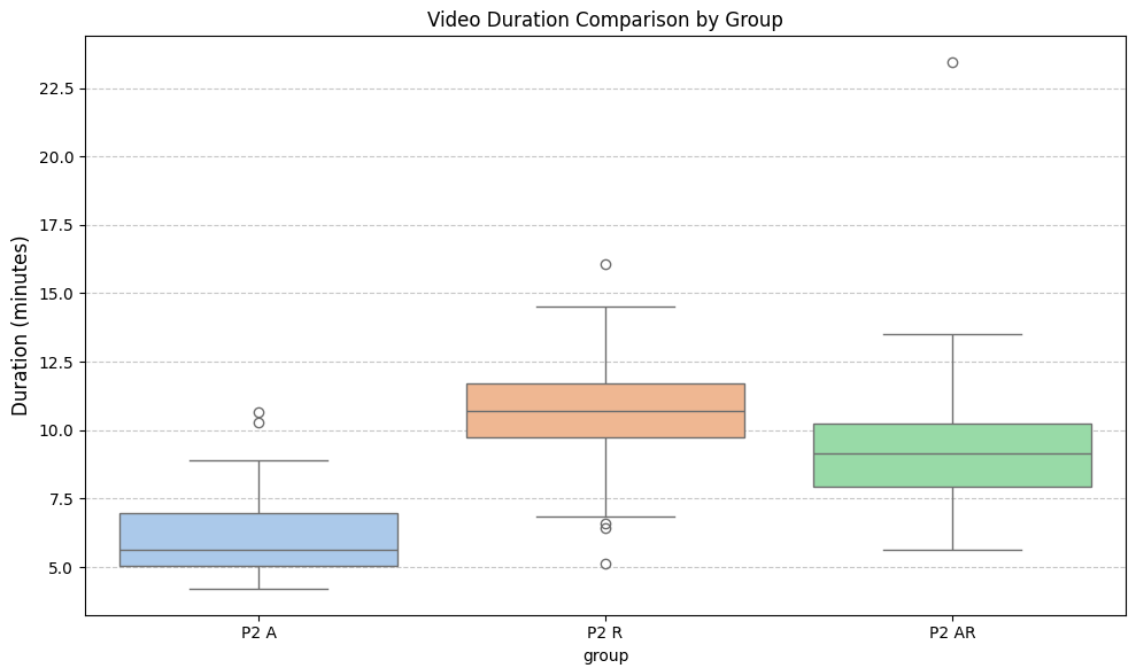


Figure E.11: BB4 corpus overview (notebook cell oo8)

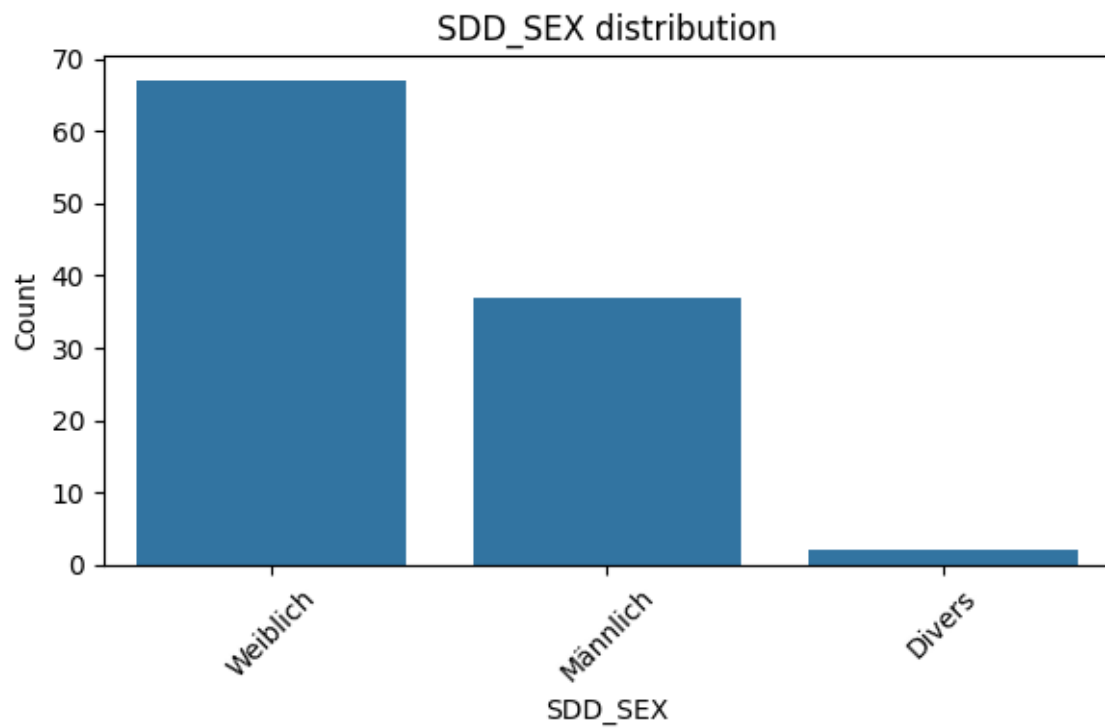


Figure E.12: BB₄ data exploration (notebook cell 002, out 01)

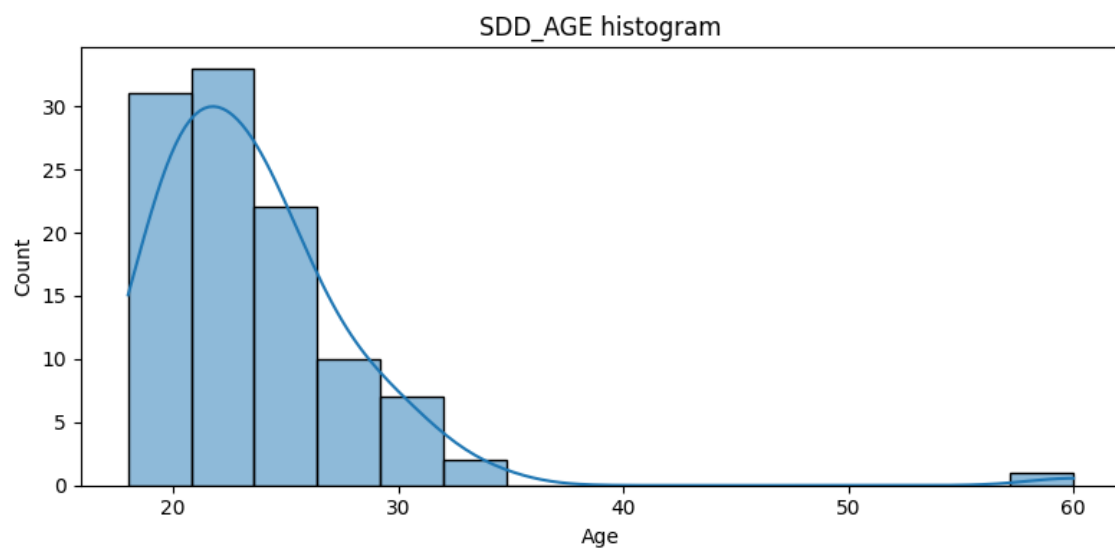


Figure E.13: BB₄ data exploration (notebook cell 002, out 03)

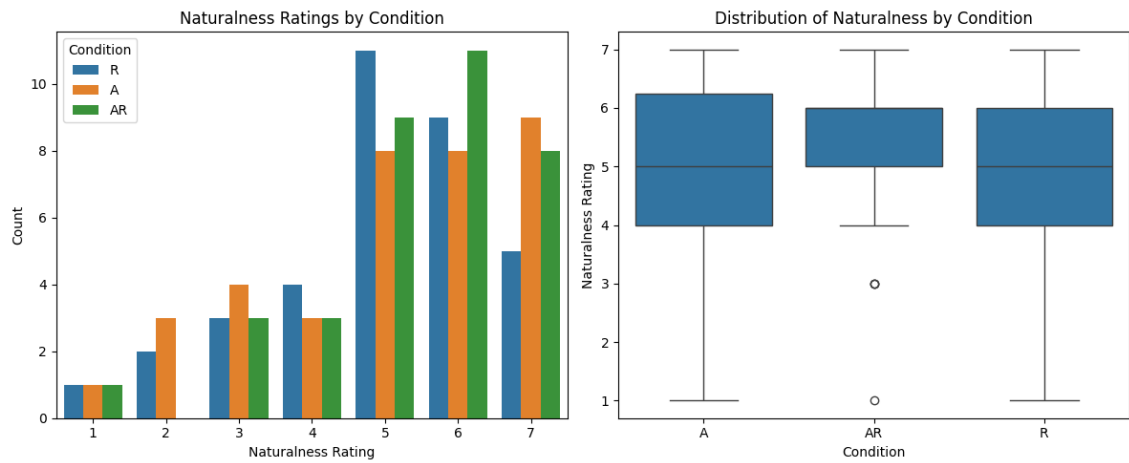


Figure E.14: BB₄ data exploration (notebook cell 005)

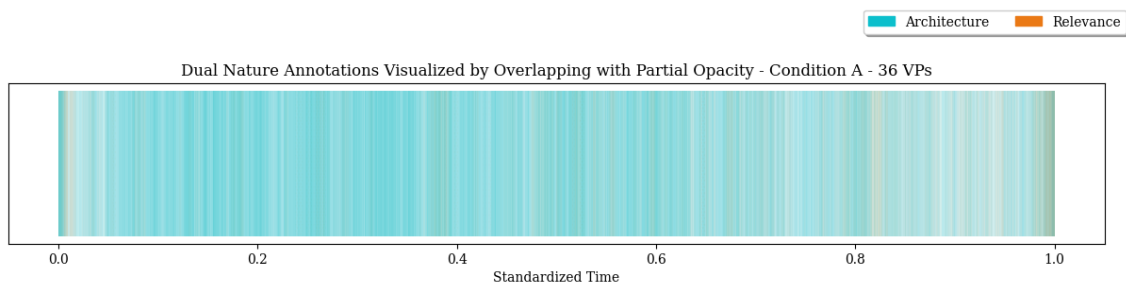


Figure E.15: BB₄ visualisation – A (all)

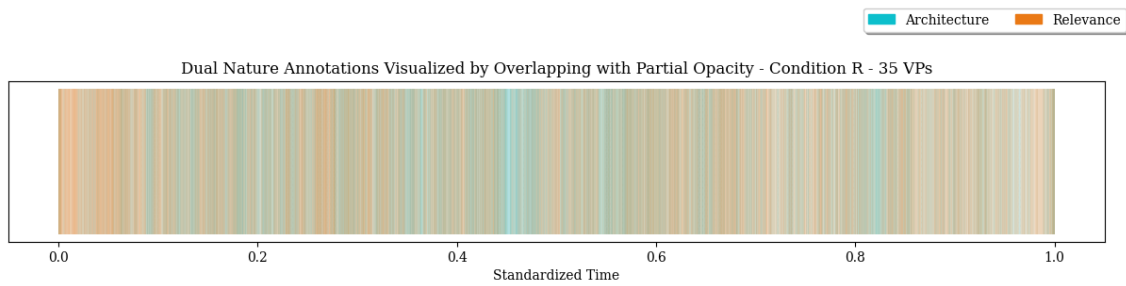


Figure E.16: BB₄ visualisation – R (all)

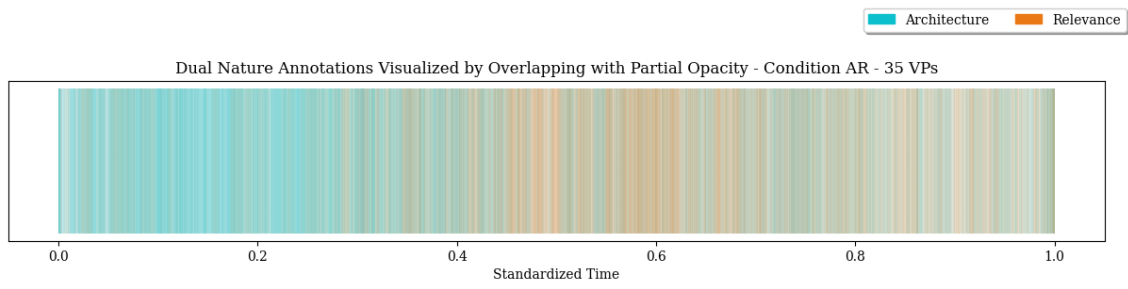


Figure E.17: BB₄ visualisation – AR (all)

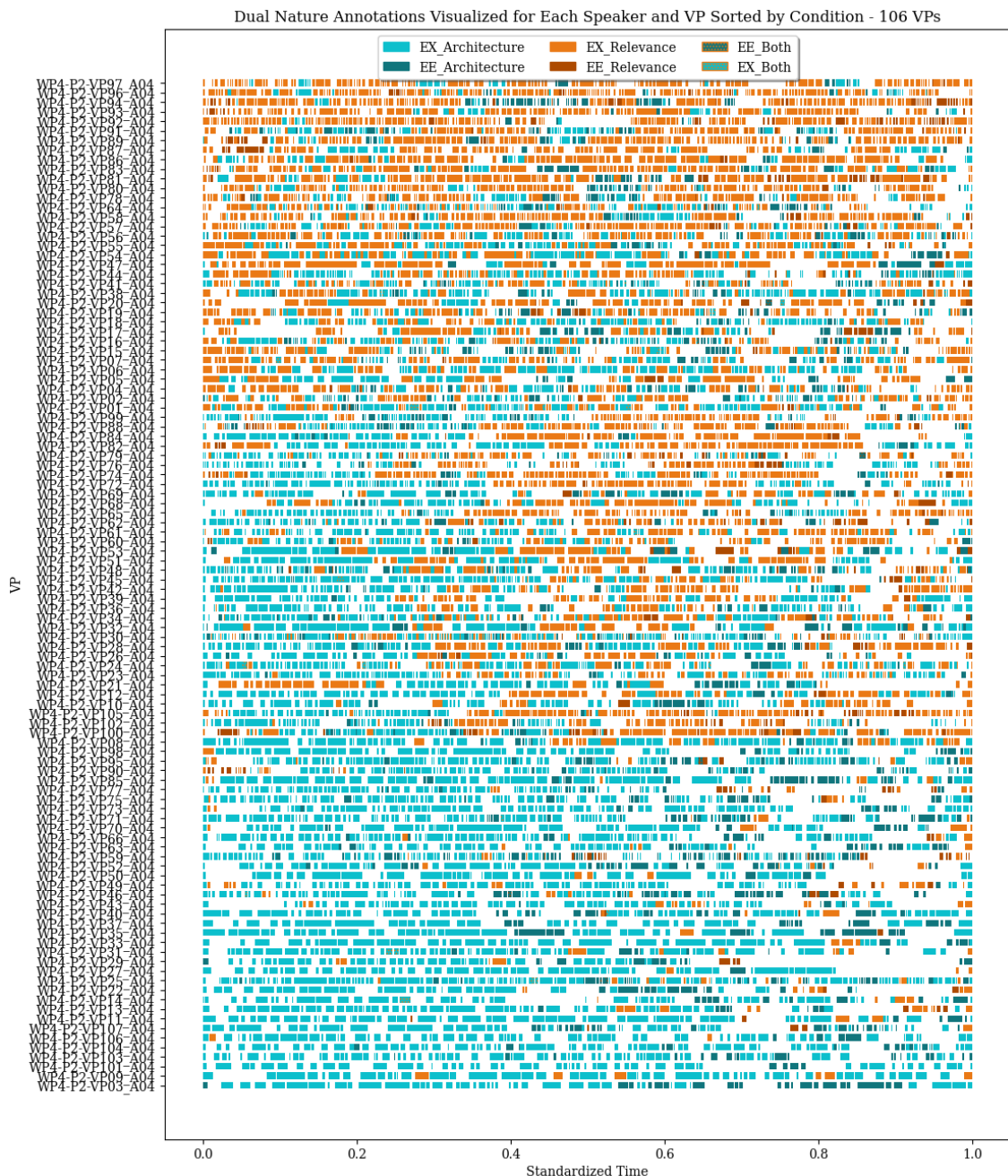
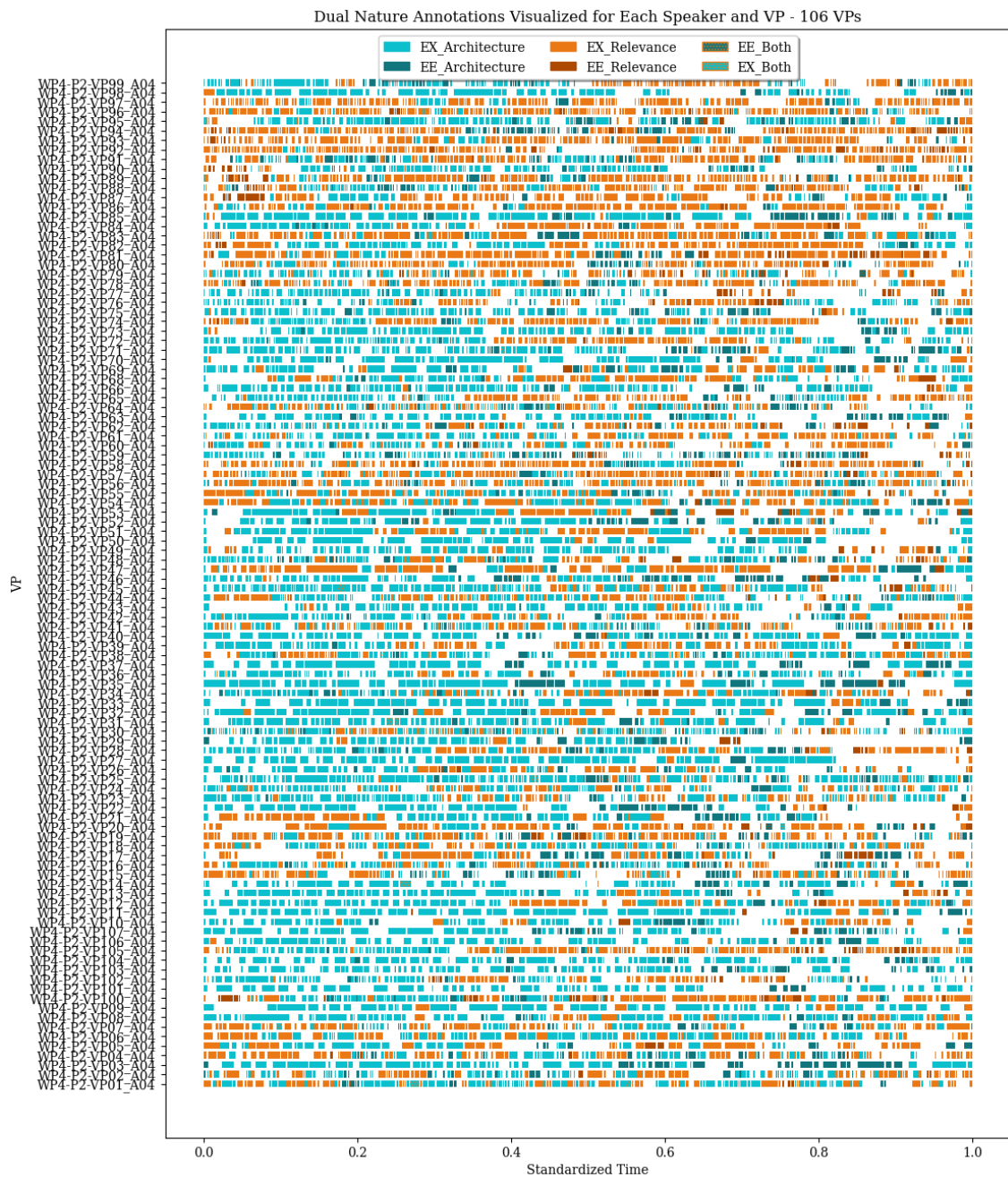


Figure E.18: BB4 individual visualisation (sorted)



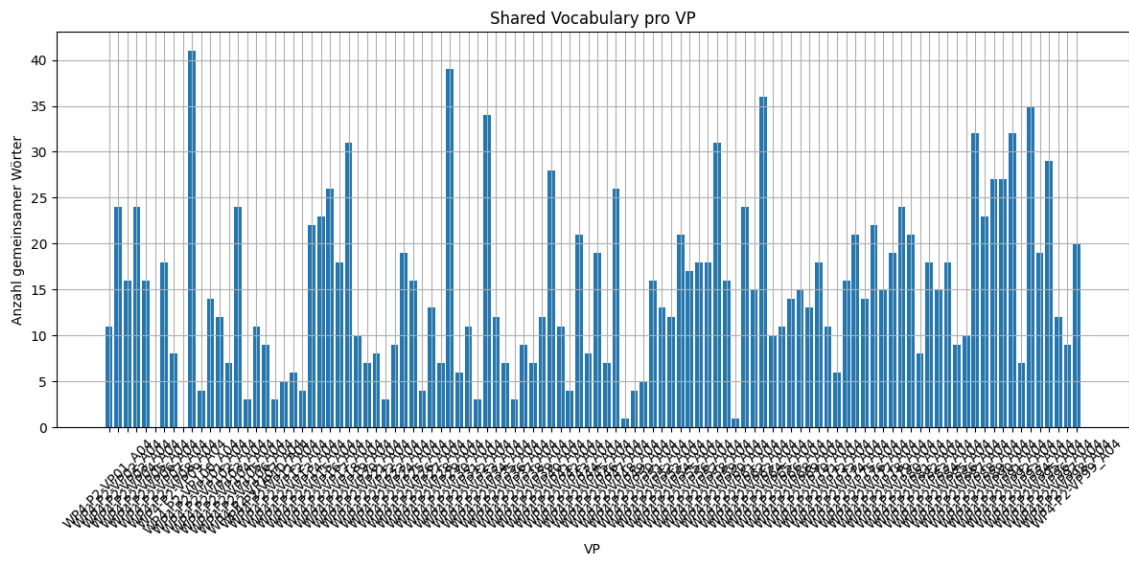
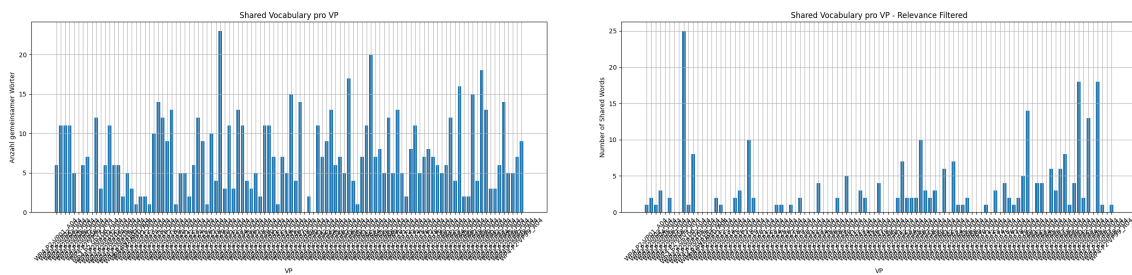


Figure E.20: BB₄ shared vocabulary analysis (combined)



(a) Shared vocabulary analysis – A

(b) Shared vocabulary analysis – R

Figure E.21: BB₄ shared vocabulary analysis by group

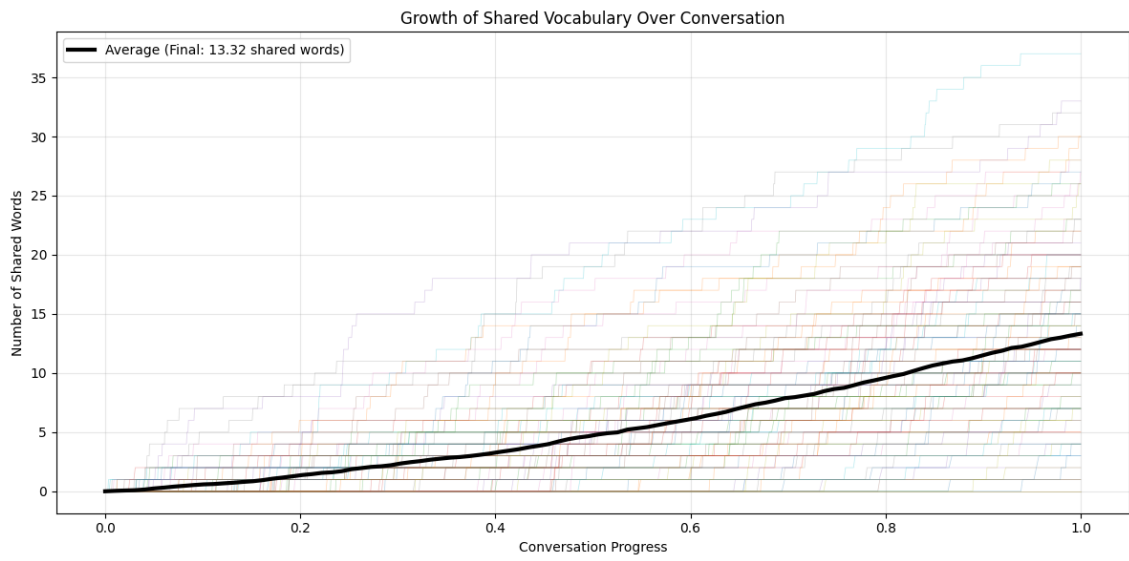
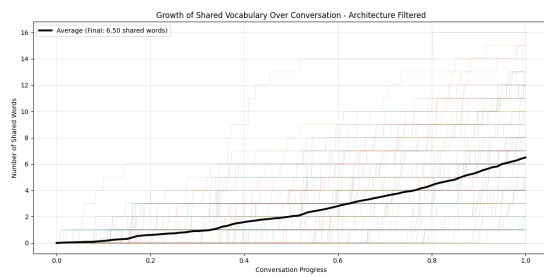
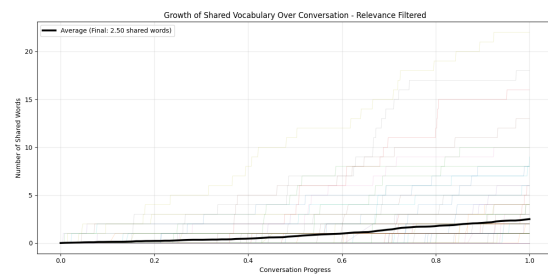


Figure E.22: BB₄ shared vocabulary evolution (combined)



(a) Shared vocabulary evolution – A



(b) Shared word evolution – R

Figure E.23: BB₄ shared vocabulary evolution by group

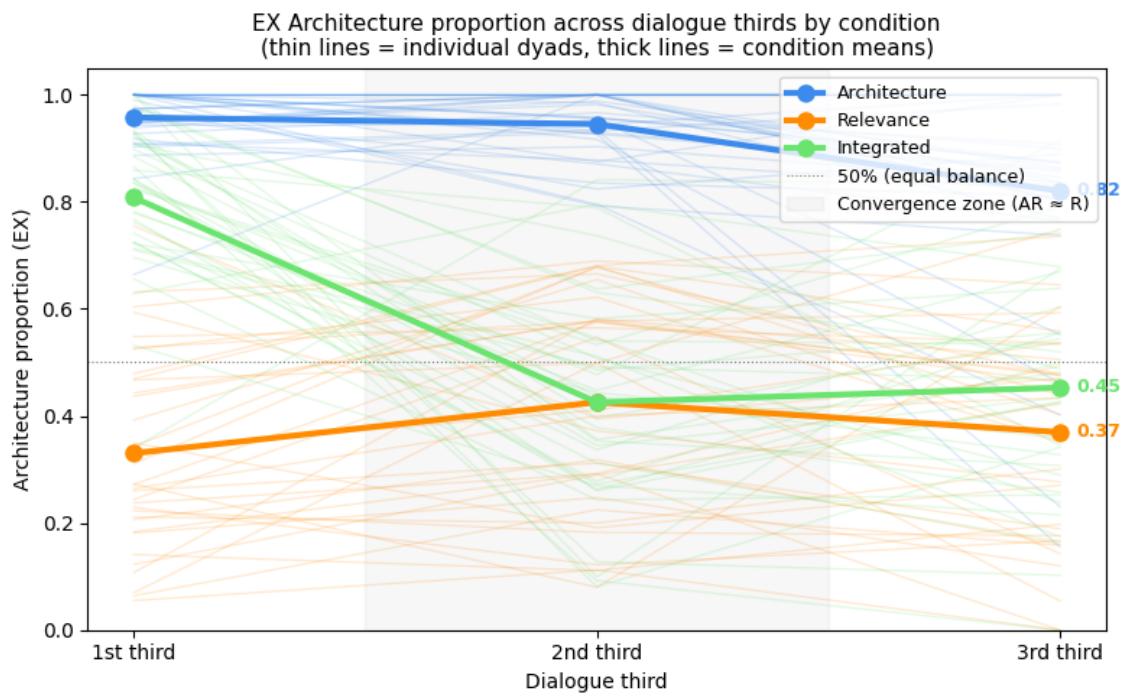


Figure E.24: BB4 correlation analysis (notebook cell o35)

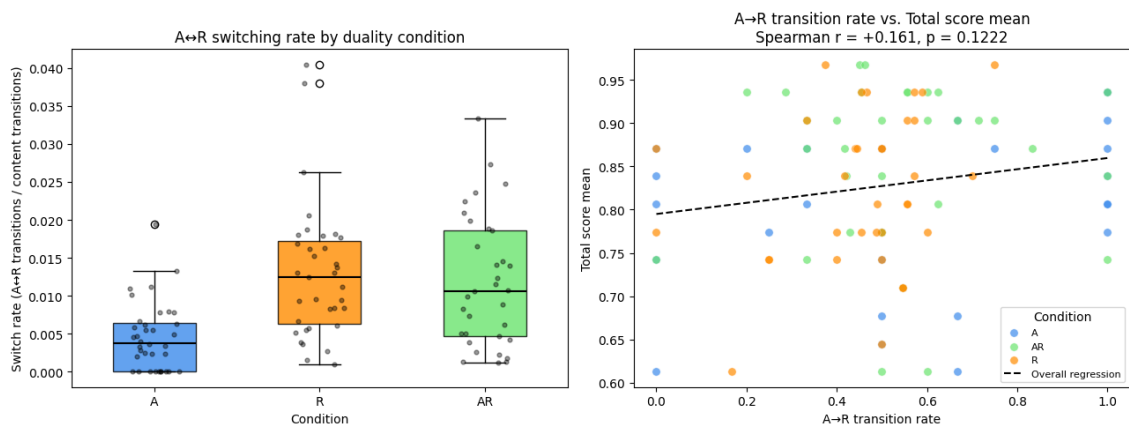


Figure E.25: BB4 DN cluster analysis (notebook cell o18)

F Exploratory Analyses

The figures in this chapter document exploratory analyses that were not part of the confirmatory research designs of BB2–BB4. They are included here to support transparency and to indicate directions for future work.

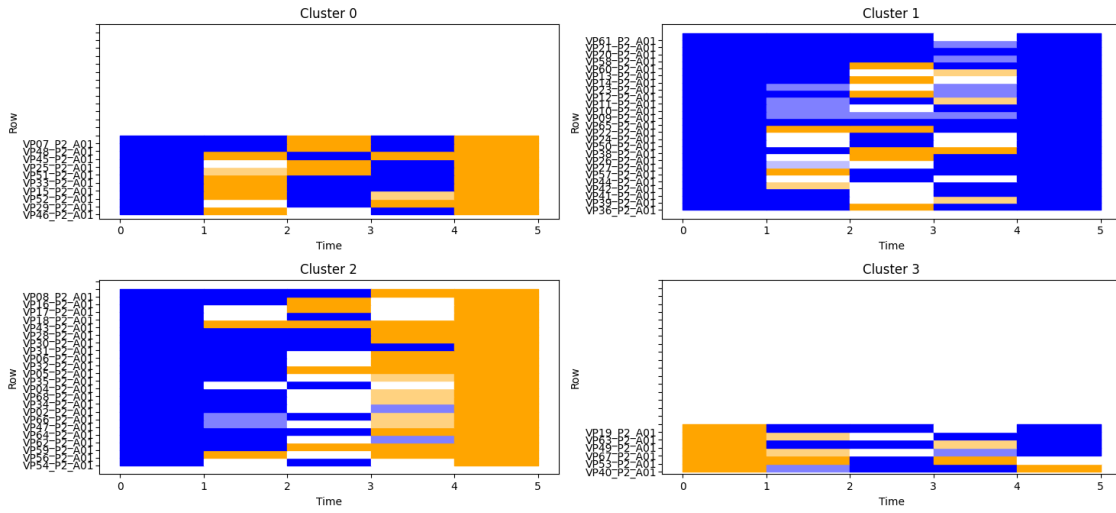


Figure F.1: *Exploratory.* K-means clustering of explanations by their Architecture and Relevance distributions across the explanations, with the cluster count determined via optimal-cluster-number estimation. Each row represents one explanation (labelled by ID); columns represent the five temporal segments used in the feature vector. Blue encodes Architecture share, orange encodes Relevance share, with shade intensity proportional to the share in that third. The analysis is illustrative rather than confirmatory: it demonstrates that distribution-based clustering can surface candidate explanation patterns, and motivates future work at finer granularity levels where such patterns might be identified as *optimal* with respect to downstream understanding outcomes.

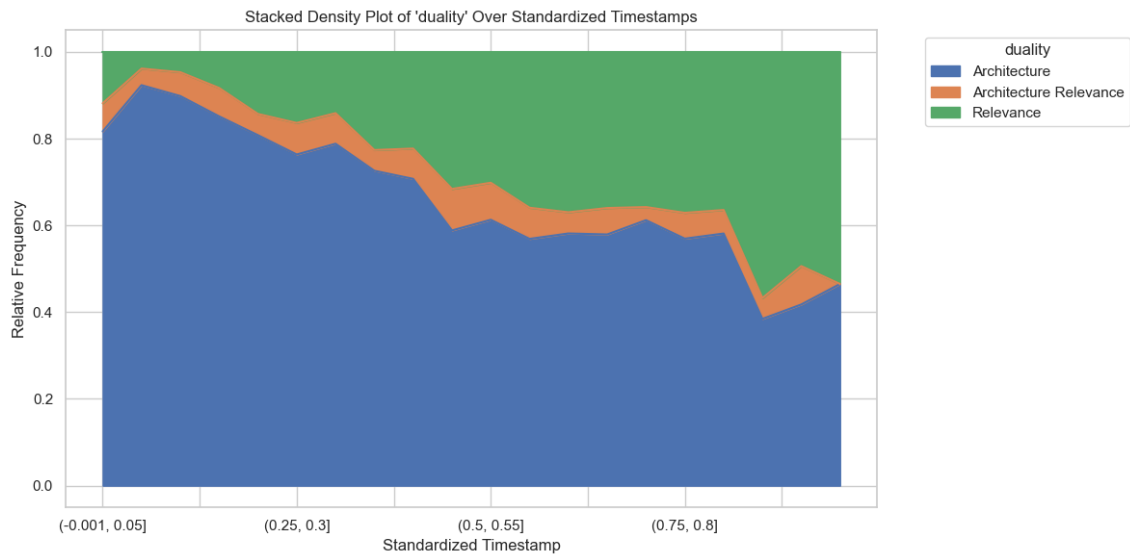


Figure F.2: A stacked density analysis gives a little more detailed insights into how the content of the explanation unfolded. Potentially useful for detecting interesting anomalies or reoccurrences for qualitative investigations (in BB2 cooperation analysis (notebook cell o21))

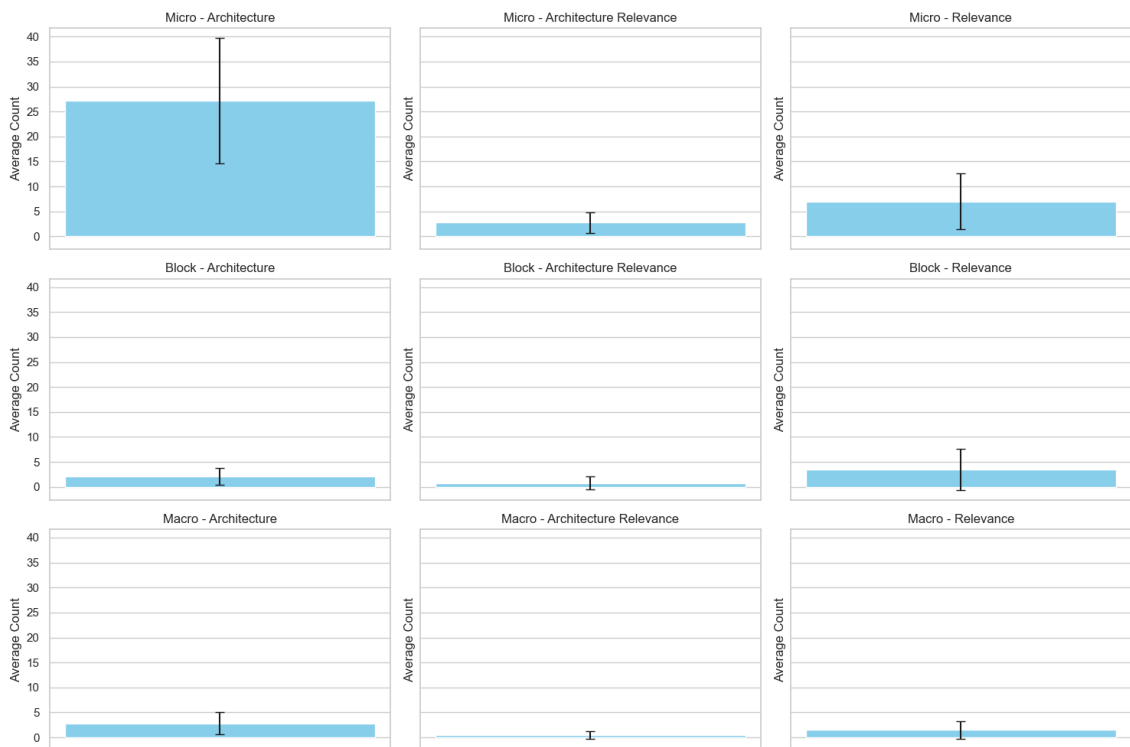


Figure F.3: WP2 Joint Dataset (Project A01+A01). Semantic node analysis was used as a foundation for programmatically determining code categories aligning with the block model cells. Visualization shows the distribution of codes (blockmodel x dual nature, in work package 2 cooperation analysis (notebook cell o23))

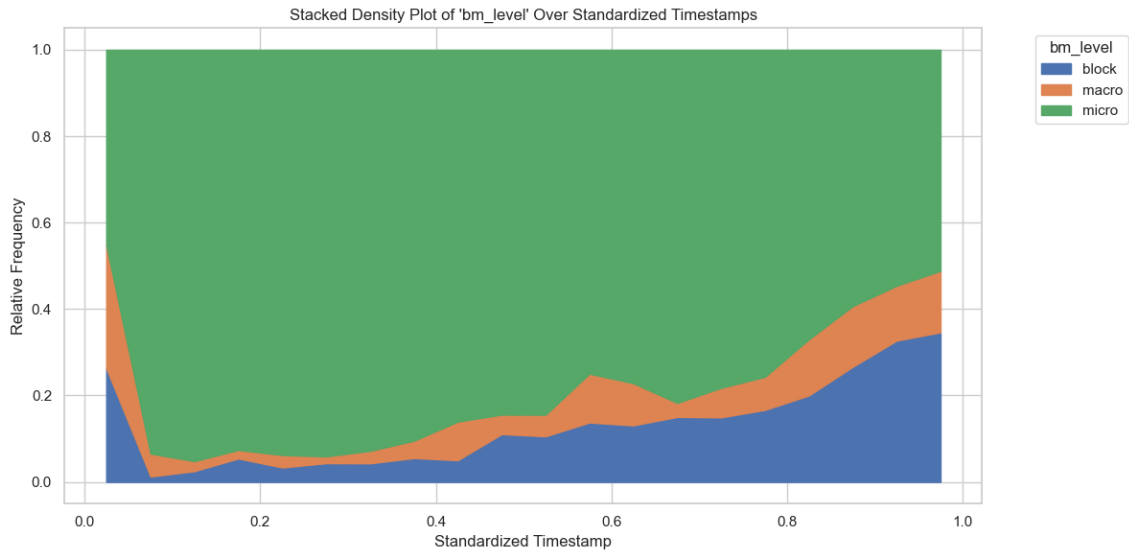


Figure F.4: WP2 Joint Dataset (Project Ao1+Ao1). Stacked density plot of block-model codes across the corpus. In work package 2 cooperation analysis (notebook cell o24)

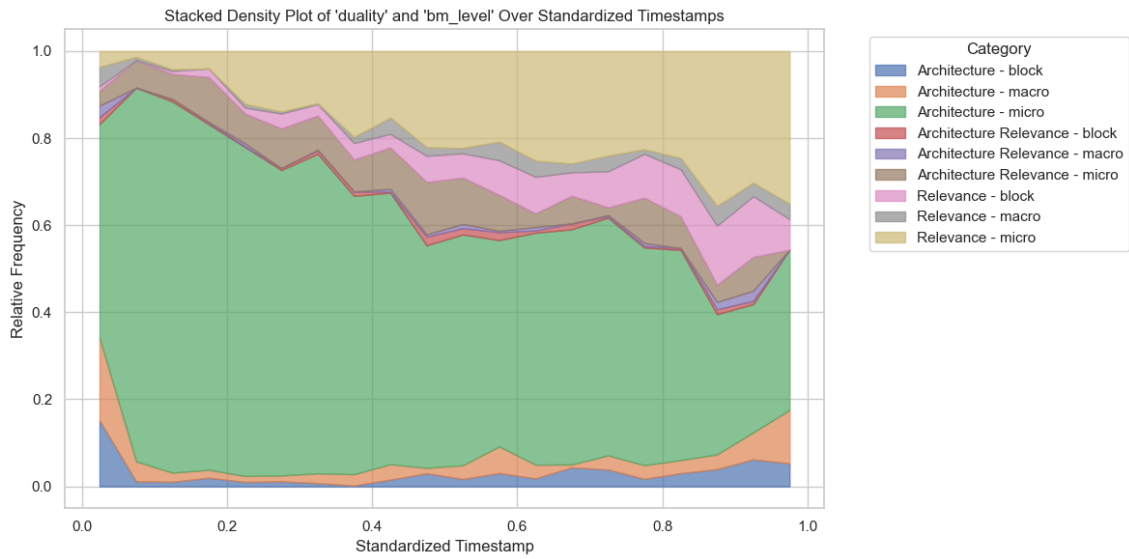


Figure F.5: WP2 Joint Dataset (Project Ao1+Ao1). Stacked density plot of block-model codes x dual nature codes across the corpus. In work package 2 cooperation analysis (notebook cell o25)

Cluster 0 Medoid Graph: VP29_P2_A01



Figure F.6: WP2 Joint Dataset (Project A01+A04). Directed graph visualization of explanation content as semantic node ((Fisher et al., n.d.)) traversal. Nodes represent unique semantic units; edges indicate traversal order, including revisits to previously addressed nodes. The displayed graph is the medoid of a k-means cluster. In work package 2, cooperation analysis (notebook cell o42).

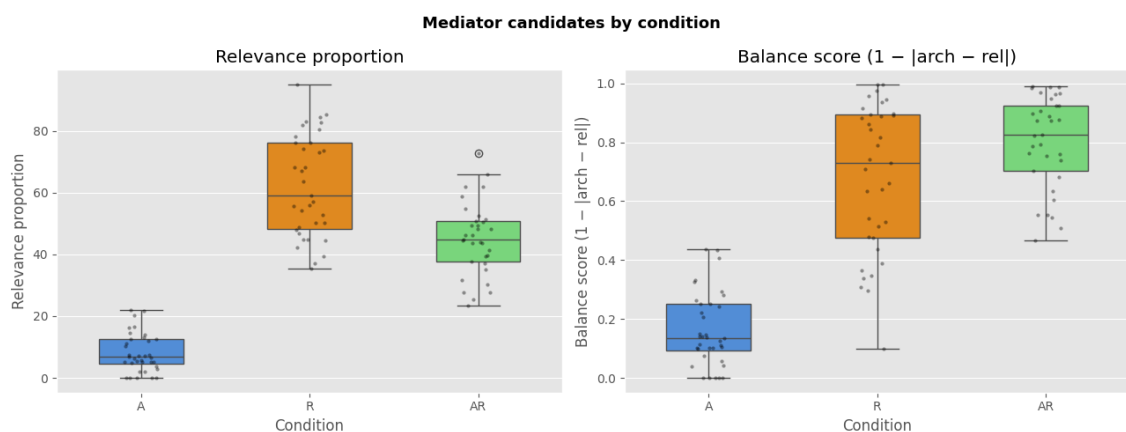


Figure F.7: Potential mediators that were tested in an exploratory manner to assess whether they predict understanding outcomes. Results were uninformative. Work package 4 correlation analysis (notebook cell oo9)

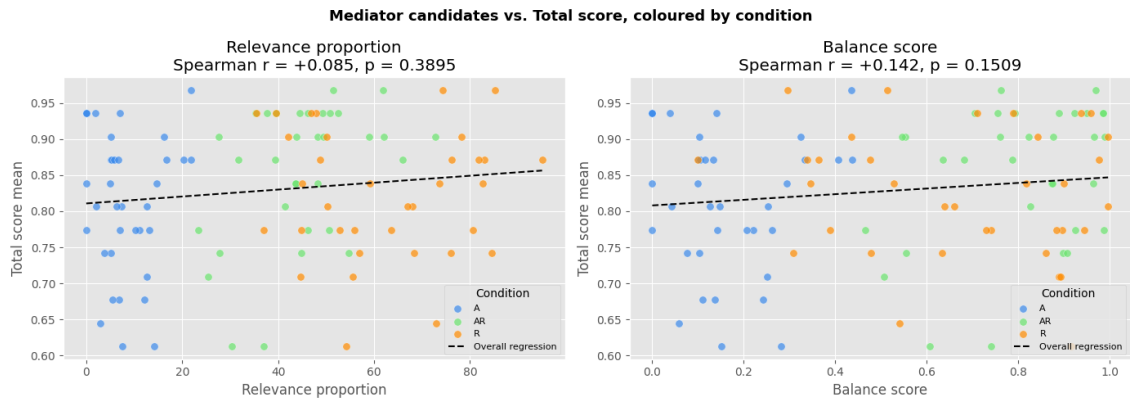


Figure E.8: Potential mediators that were tested in an exploratory manner to assess whether they predict understanding outcomes. Results were uninformative (across all conditions). Work package 4 correlation analysis (notebook cell 014)

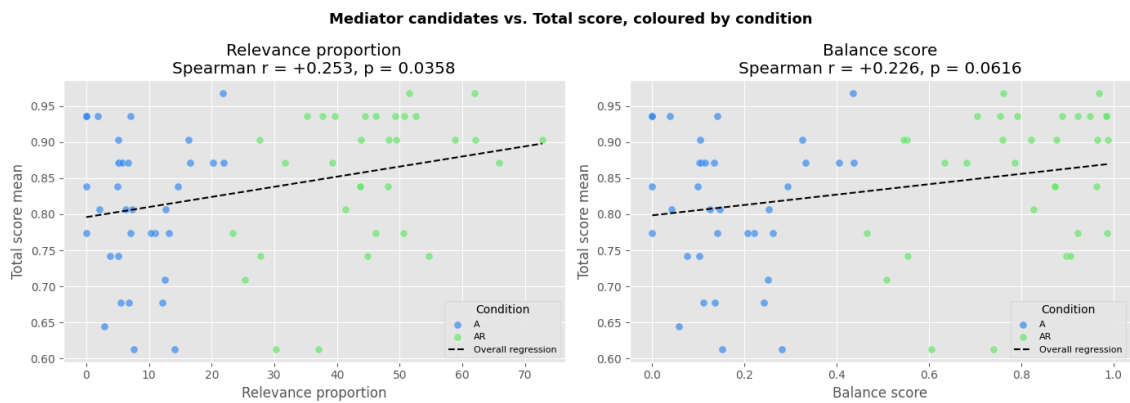


Figure E.9: Potential mediators that were tested in an exploratory manner to assess whether they predict understanding outcomes. Results were uninformative (A vs. AR conditions). In work package 4 correlation analysis (notebook cell 015)

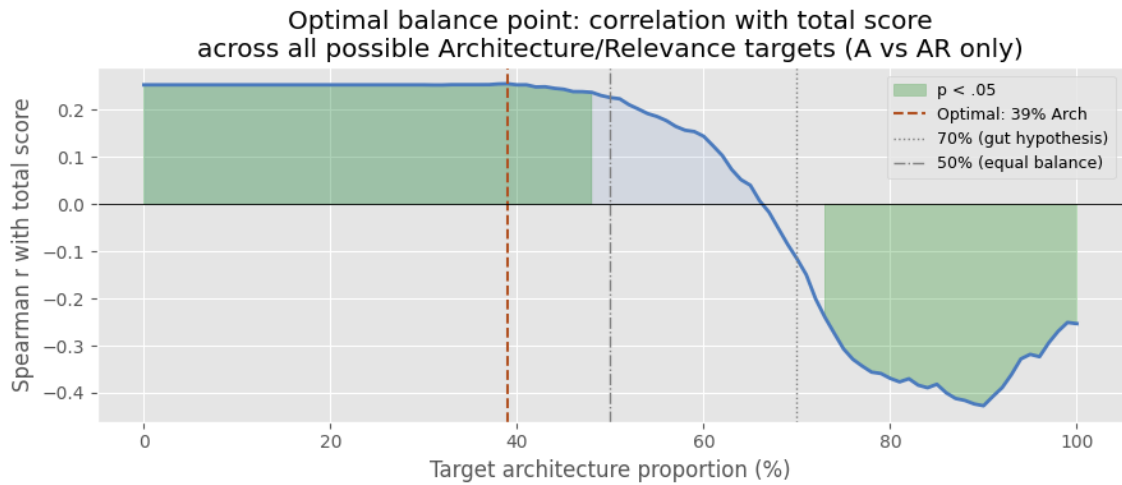


Figure F.10: Exploratory analysis. Programmatical approach for the assessment of an optimal balance point in relation to understanding outcomes. This optimum would provide a percentage of how much Architecture, in this instance, needs addressing for an optimal understanding outcome. Potentially relevant, if datasets grow larger. Work package 4 correlation analysis (in notebook cell o22)

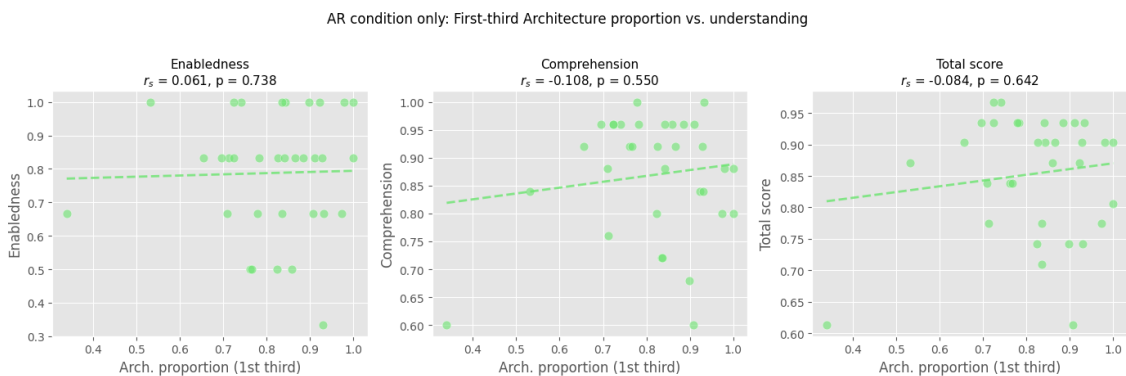


Figure F.11: Exploratory analysis. Does laying out the tools influence understanding outcomes? In work package 4 correlation analysis (notebook cell o33)