Goals as Reference Points: Empirical Evidence from a Virtual Reward System

Tobias Mutter
University of Paderborn
Paderborn, Germany
tobias.mutter@wiwi.upb.de

Dennis Kundisch
University of Paderborn
Paderborn, Germany
dennis.kundisch@wiwi.upb.de

Abstract

Heath et al. (1999) propose a prospect theory model for goal behavior which offers insights on how goals affect individual and group performance. Their analytical model is based on the assumption that goals inherit the main properties of the prospect theory value function, i.e., reference point dependence, loss aversion, and diminishing sensitivity. Evidence from laboratory experiments as well as first evidence from the field support this modeling choice. In this paper, we complement this work by investigating whether the main properties of the prospect theory value function transfer to goal behavior in the field. In particular, we analyze how individual performance is affected by the presence of goals. For our research, we take user activity data from a popular German Question & Answer community and analyze how users adjust their contribution behavior in the days surrounding goal achievement, where goals are represented by badges. We find that users gradually increase their performance in the days prior to earning a badge, with performance peaking on the day of the promotion. In subsequent days, user performance gradually diminishes again, with the decline being strongest on the day immediately following the badge achievement. Overall, user performance is higher *in the days preceding badge achievement, compared to subsequent days. These findings reflect the characteristic S-shape of the prospect theory value function which is convex below the reference point and concave above it. Our results thus support the transferability of the main properties of the prospect theory value function to goal behavior in the field and suggest a distinct shape of the value function around goals.*

Keywords: Goal-Setting Theory, Prospect Theory, Value Function, Gamification, Badges

Introduction

It is well established in the goal literature that individuals perform better when given more specific and challenging goals compared with being told to ‘do your best’ (e.g., Heath et al. 1999, Locke & Latham 1990, 1991, 2002, Mento et al. 1987). Goal setting theory (Locke & Latham 1990, 2002) proposes three key mechanisms for this behavior: goals (1) activate individuals to increase their effort; (2) lead to greater persistence; and (3) direct attention toward goal-relevant activities (Heath et al. 1999, Locke & Latham 2002). The literature distinguishes between two main types of goals: extrinsic rewards, and ‘mere’ goals (Heath et al, 1999). Extrinsic rewards are associated with external objects and have a direct impact on physiological well-being, while ‘mere’ goals represent ‘specific levels of performance (e.g., finishing a manuscript in 3 days as opposed to 5)’, without discrete pay-offs (Heath et al. 1999, p. 80). The effectiveness of extrinsic rewards may be explained with economic calculus, whereas ‘mere’ goals require a psychological explanation (Heath et al. 1999).

Heath et al. (1999) develop an analytical model for ‘mere’ goals in which the three key motivational phenomena associated with goal setting - effort, persistence, and attention – can be explained by one single underlying process, the value function. In their prospect theory model of goal behavior, goals serve as reference points for the valuation of outcomes, and since the content and intensity of goals alter the valuation of goals, the behavior in response to goals can be explained by loss aversion and diminishing sensitivity (Kahneman & Tversky 1979). In other words, according to Heath et al (1999, p93), goals ‘inherit the properties of [Prospect Theory’s] value function’, and their model offers insights on how ‘mere’ goals affect performance on the individual and on the group level.

Heath et al. (1999) provide evidence from laboratory experiments to support their modeling choices and, in addition, discuss the applicability of their model on the empirical literature on goal setting. However, in order to ensure generalizability of results it is also important to confirm laboratory findings with evidence from the field (Levitt & List 2007). Recently, Markle et al. (2014) and Allen et al. (2014) were the first to provide evidence for Heath et al.'s (1999) model in a field environment. Both studies use data from marathon runners. Markle et al. (2014, p. 1) find that 'satisfaction as a function of relative performance (the difference between a runner's time goal and her finishing time) exhibits loss aversion and diminishing sensitivity'. At the same time, their results reveal a discontinuity (or jump) at the reference point, which indicates that the value function adopts a distinct shape around a goal. Allen et al. (2014, p. 3) analyze the finishing times of marathon runners and find 'a lumpy distribution of finishing times, with bunching just ahead of round numbers' (e.g., at each 30 and 60-minute mark). However, they find that the shape of the distribution of the finishing times can be modeled with standard prospect theory parameter estimates. With our paper we complement the work of Markle et al. (2014) and Allen et al. (2014) by providing field evidence for the applicability of Heath et al.'s (1999) model to predict user performance in the presence of 'mere' goals. In particular, we answer the following research question: *Do individual performance patterns exhibit the main properties of the prospect theory value function (i.e., reference point dependence, loss aversion, and diminishing sensitivity) in the days surrounding 'mere' goal achievement?*

For our empirical analysis, we subject user activity data from a popular German Question & Answer (Q&A) community to an analysis of how users adjust their contribution behavior in the presence of 'mere' goals, represented by badges. The Q&A community deals exclusively with leisure-related topics (e.g., beauty, computers, gardening) and has set up a virtual reward system to activate its members. On performing certain activities, users are rewarded with points with which, over time, they can earn up to 20 successive badges that confer no more than an individual's 'status' on the platform. The platform thus provides a suitable research environment for investigating the effect of 'mere goals' on performance because we can rule out potential confounding effects caused by any type of external rewards (Heath et al, 1999). Our dataset includes detailed information about all user activity on the platform between February 2007 and May 2008. Overall, we analyze 5,828 users over a time period of 462 days. In particular, we investigate how user performance changes in the days before and after they earn a badge.

We take the two main activities on the platform (asking and answering questions) to measure user performance on a daily level. We find that users progressively increase their performance in the days before earning a badge, with performance peaking on the day of their promotion, and progressively declining over the subsequent days. The strongest drop in the level of performance occurs on the day after a badge is gained. Overall user performance levels are higher in the run-up to earning a badge than in the aftermath. These findings reflect the characteristic S-shape of the prospect theory value function, i.e. convex below the reference point and concave above it. In line with Markle et al. (2014) and Allen et al. (2014), our results thus support the transferability of the main properties of the prospect theory value function to goal behavior in the field and suggest a distinct shape of the value function around goals. With this paper we make novel and significant contributions to research in two ways: (1) by providing field evidence for the applicability of the Heath et al. (1999) model to predict user performance in the presence of 'mere' goals; and (2) we contribute new insights to the more recent literature on gamification which explores the effect of game mechanisms like badges on user activity levels in online communities (e.g., Hamari & Eranti 2011, Blohm & Leimeister 2013).

Theoretical Background

Three strands of literature are relevant to our study. The first is related to research in the field of goal setting, the second develops a prospect theory model for goal setting, while the third highlights our contribution to the gamification literature.

Goal Setting Research

Research on goal setting dates back to the end of the 19th century (Latham & Locke 2007). In the first 60 years, research was mostly conducted in an ad hoc manner resulting in numerous, but unrelated, studies (Latham & Locke 2007). The development of the goal setting theory by Locke & Latham in the nineteen-sixties represents an important milestone and provides the first theoretical framework to guide studies in the field (Locke & Latham 1990, 2002). To summarize the findings of the goal setting literature, Locke & Latham performed a comprehensive literature review covering 239 laboratory and 156 field studies (Locke & Latham 1990, 1991). These studies 'have been conducted with 88 different tasks including bargaining, driving, faculty research, health promoting behaviors, logging, maintenance and technical work, managerial work, management training, and safety' (Locke & Latham 1991, p. 216). The two main findings emerging from these goal setting studies are, first, that assigning individuals goals increases performance, even when the level of difficulty is set high; and second, specific and challenging goals elicit superior performances compared with being told to 'do your best' (Locke & Latham 2002). The generalizability of this behavior is such that Mento et al. (1987, p. 74) are able to state that '[i]f there is ever to be a viable candidate from the organizational sciences for elevation to the lofty status of a scientific law of nature, then the relationships between goal difficulty, difficulty/specificity, and task performance are most worthy of serious consideration.' As explanation for the two key findings of the literature, research (e.g. Heath et al. 1999, Locke & Latham 2002) identifies the causal relationship between goals and the three associated motivational phenomena – effort, persistence and attention – describing these as mechanisms that links them to goals (Wu et al. 2008) in the following manner:[1]

Goals increase Effort

According to Locke & Latham (1990), goals increase the effort invested in an activity by providing a focus, and that this has the effect of increasing performance. Moreover, the more difficult a goal the more effort is invested. A plethora of experiments covering a wide range of both physical (e.g., lifting weight or squeezing a grip) and cognitive tasks (e.g., calculus or brainstorming) (e.g., Ness & Patton 1979, Botterill 1977, Bryan & Locke 1967, Bandura & Schunk 1981, Garland 1982) showed that goals consistently increase effort.

Goals increase Persistence

Goals have also a positive impact on persistence. Locke & Latham (1990, p. 95) state that '[h]ard goals ensure that individuals will keep working longer than they would with vague or easy goals. Hard or challenging goals inspire the individual to be tenacious in not settling for less than could be achieved.' This has also been shown in a variety of studies on both physical (e.g., pain tolerance) and cognitive tasks (e.g., solving anagrams or studying) (e.g., Stevenson et al. 1984, Sales 1970, Laporte & Nath 1976).

Goals direct Attention

According to Locke & Latham (1990, 1992), goals direct attention toward goal-relevant activities and away from goal-irrelevant activities. For example, Locke & Bryan (1969) analyzed how subjects on a driving course changed their driving behavior. Having received feedback on various aspects of their driving performance (e.g., steering, braking, or acceleration) whilst simultaneously being set a goal to improve only one particular aspect of driving, they found that subjects improved their performance only on the aspect for which the goal was set. Numerous other studies support this finding (e.g., Nemeroff & Cosentino 1979, Rothkopf & Billington 1979, Wyeret al. 1982)

A Prospect Theory Model for Goal Setting

Building on previous research on goal setting, Heath et al. (1999) develop an analytical model which integrates and explains three behavioral phenomena (i.e., effort, persistence, and attention) within one single underlying explanatory process — the prospect theory value function.

Value Function

The prospect theory value function describes how goals alter the valuation of outcomes by acting as a reference point (Kahneman & Tversky 1979, Tversky & Kahneman 1992). The theory has been successfully tested in various fields including economics (Benartzi & Thaler 1995, Odean, 1998), medicine (McNeil, Pauker, & Tversky 1988), consumer behavior (Thaler 1985), social psychology, and political science (Kramer, 1989, Quattrone & Tversky 1988). Heath et al. (1999) propose a prospect theory model of goal behavior where goals represent reference points which help assess one's performance in terms of gain (success) and loss (failure), in a manner consistent with the value function.

The value function $v(x)$ can be characterized by three key principles:[2]

(1) *Reference point*: The reference point enables a spatial attribution (or conceptualization) of outcomes into regions of gain and loss. This implies that individuals evaluate outcomes as gain or loss relative to a reference point.

(2) *Loss aversion*: Losses are perceived as more painful than similar-sized gains are perceived as pleasurable ($v(x) < |v(-x)|, x > 0$). For example, the negative utility of losing $5 is larger than the positive utility of gaining $5 ($v(5) < |v(-5)|$).

(3) *Diminishing sensitivity*: The impact of an additional unit of outcome decreases with increasing distance from the reference point. For example, the additional utility gained from getting $120 instead of $110 is smaller than the additional utility from getting $20 instead of $10 (in general, $v''(x) < 0,\ x > 0$). Furthermore, the negative utility from losing -$120 instead of -$110 is smaller than the negative utility from losing -$20 instead of -$10 (in general, $v''(x) > 0,\ x < 0$). The value function is S-shaped due to diminishing sensitivity – convex below the reference point and concave above it.

Implications of the Value Function

According to Heath et al. (1999, p. 83) 'goals "inherit" the properties of the value function'. This has a number of implications for how goals affect performance.

Implications of the Reference Point

Goals which act as reference point help classify outcomes as success or failure. Individuals experience a positive or negative emotion based on this assessment. This has implications on how individuals feel about their performance. If individuals are unable to achieve a goal because it is too difficult, they end up dissatisfied with their performance and perceive their efforts as failure. If the level of difficulty is adequate and individuals are able to achieve their goal, they are satisfied with their performance and perceive it as success.

Implications of Loss Aversion

Loss aversion implies that loss (failure) is more painful than similar sized gain (success) is pleasurable ($v(x) < |v(-x)|, x > 0$). This implies that individuals who are below their goal by x units work harder to increase their performance than people who are above their goal by x units.

Implications of Diminishing Sensitivity

Diminishing sensitivity implies that the impact of goals on effort depends on whether individuals are above or below their goal. Individuals are expected to increase their effort with proximity towards their goal but to reduce it with increasing distance away from their goal. It also implies that individuals who are at greater distance from a goal might struggle to motivate themselves because they do not feel that they are making noticeable progress towards their goal.

The Effect of Goals on Performance

Heath et al. (1999) make three key assumptions for how the three main behavioral phenomena associated with goal setting (i.e., attention, effort, and persistence) can be explained with the value function of prospect theory. They assume that goals represent a reference point, that the value function describes how the value of outcomes is altered by goals, and, that individuals weigh up the benefits and costs gained from one additional unit of performance to determine their optimal performance level. To make this more explicit:

(1) The previously defined value function $v(x)$ and the selected goal g (reference point) determine the benefits $b_g(x)$ of an individual completing x units of performance and $b_g(x) = v(x - g)$.

(2) Let $c(x)$ describe the cost of performing x units. The cost of one additional unit of performance (e.g., providing one more answer) increases as the overall performance level increases (e.g., because of fatigue or boredom) which is equivalent to $c'(x) > 0$.

(3) Individuals stop to increase their performance when *marginal costs* equal *marginal benefits*. This implies that individuals are myopic, ignore the effect of

expectancies and only compare the marginal costs and benefits of one additional unit of performance.

Figure 1 and Figure 2 illustrate how the three key phenomena (i.e., effort, persistence, and attention) can be explained based on the Heath et al.'s model (1999). Figure 2 shows two benefit functions $b_{g_1}(x)$ and $b_{g_2}(x)$ for two individuals with the same ability but with two different goals g_1 and g_2 where $g_2 > g_1$. Figure 2 presents the corresponding marginal benefit functions $b'_{g_1}(x)$ and $b'_{g_2}(x)$ for the two goals g_1 and g_2 and the marginal cost function $c'(x)$.

(1) *Effort*: The slope of the benefit function $b_g(x)$ describes the marginal benefit of performing one additional unit of performance. Individuals are expected to increase their effort level (i.e., bear higher costs for performance) when marginal benefits are high. By comparing the slopes of the benefit functions $b_{g_1}(x)$ and $b_{g_2}(x)$ at a given level of performance, Heath et al. (1999) determine under which goal condition (i.e., $g_1$or g_2) an individual is willing to take higher cost and exert higher effort. For example, at $x = l$ the slope of the benefit function with the easier goal g_1 is larger $b'_{g_1}(l) > b'_{g_2}(l)$, but with increasing performance it is the other way around with $b'_{g_1}(m) < b'_{g_2}(m)$ at $x = m$ and $b'_{g_1}(n) < b'_{g_2}(n)$ at $x = n$ where $n > m > l$ (see Figure 1). At $x = l$ both individuals are in the area of failure but the individual with the easier goal g_1 is closer to the goal, thus her marginal benefits are higher. At $x = m$ both individuals are d units away from their goal but only the individual with the higher goal g_2 is in the failure region and thus her marginal benefits are higher. At $x = n$ both individuals are in the success region but the individual with the higher goal g_2 is closer to the reference point and thus perceives her marginal benefits as higher.

(2) *Persistence*: Individuals constantly compare the costs of an additional unit of performance $c'(x)$ with the benefits $b_g'(x)$ derived from it, and persist at a task as long as $b_g'(x) \geq c'(x)$. Important to note is that the marginal cost function $c'(x)$ does not depend on the goal while the marginal benefit function $b_g'(x)$ does. This is also the reason why individuals who set a challenging goal show greater persistence than individuals who set easy or modest goals as long as goals are not 'too difficult'. Heath et al. (1999, p. 97) explain this phenomenon in terms of marginal benefits and state that

> 'People who set modest goals and exceed them soon find themselves in the domain of gains where marginal benefits decrease swiftly because of diminishing sensitivity. Individuals who set aggressive goals remain longer in the domain of losses where marginal benefits are high (because of loss aversion) and increasing (because of convexity)'.

This mechanism is illustrated in Figure 2. The individual with the higher goal g_2 persists longer than the individual who pursues the easier goal g_1.

(3) *Attention*: Goals acting as reference points are more motivating because they focus attention on a specific task. Heath et al. (1999) argue that individuals are strategically allocating their time to different tasks, and this allocation is dependent on the relative perceived importance of each task. The relative task importance rises if the marginal benefits $b_g'(x)$ for a specific task are seen to increase. This should also increase the attention allocated to this task and lower the attention for other tasks.

Impact of Goals on Performance

Performance can be defined as 'the result of work over time, and work depends on effort and persistence' (Heath et al. 1999, p.98). The model introduced previously predicts that 'if an individual performs above a specific goal, she would have performed even better if she had increased her goal slightly, but that if she performs below a goal, she would have performed better if she had lowered her goal slightly' (Heath et al. 1999, p. 98). For example, the individual who has set goal g_1 and performs $x^*_{g_1}$ units where $b'_{g_1}(x^*_{g_1}) = c'(x^*_{g_1})$ (see Figure 2). If she had set goal g_2 instead of g_1 her overall performance would have increased to $x^*_{g_2}$ where $b'_{g_2}(x^*_{g_2}) = c'(x^*_{g_2})$. However, if the individual would be unable to accomplish goal g_1, setting goal g_2 instead of g_1 would have the opposite effect and led to reduced overall performance, because the marginal benefit function $b'_{g_2}(x)$ lies below $b'_{g_1}(x)$ when $x < g_1$.

The prospect theory model of goal behavior refines the general statement from the goal setting literature that more challenging goals lead to better performance (Locke & Latham 1991). Heath et al.'s model (1999) allows making predictions on the individual and the group level and provides new and valuable insights on how goals affect performance.[3]

Predictions

The model developed by Heath et al. (1999) allows making predictions about how adjustments to the level of difficulty of a goal affect individual and group performance and subsequently whether the level of difficulty should be increased or reduced to improve performance. In addition, the model offers detailed insights on how marginal effort changes with growing or decreasing proximity to a goal. Thereby, the model provides a theoretical foundation for the 'goal-gradient hypothesis' which states that the effort invested in reaching a goal increases with proximity towards its (Kivetz et al. 2006). Further, the value function provides an explanation for the 'starting problem' which states that challenging goals might be demotivating in the beginning because the perceived progress towards a goal is negligibly small. The theoretical explanation for the starting problem is the concept of 'diminishing sensitivity', one of the components of Kahneman and Tversky's (1979) value function. At the same time, the model provides a theoretical explanation for the implementation of subgoals as an approach to resolve the starting problem. Finally, the value function predicts that 'performance will cluster around (or 'pile up') around difficult but attainable goals' (Heath et al. 1999, p. 103). The reason is that the marginal benefits of an additional unit of performance increase with proximity towards the goal and decrease thereafter.

Empirical Evidence

Heath et al. (1999) provide laboratory evidence to support their modelling choices and to test the predictions from their model (e.g., starting problem and subgoals). In particular, they use paper based experiments to test the transferability of each of the three main properties of the prospect theory value function to goal behavior (i.e., reference point dependence, loss aversion, and diminishing sensitivity), to illustrate the starting problem and the potential benefits of subgoals.

Markle et al. (2014) and Allen et al. (2014) were the first to provide evidence for Heath et al.'s (1999) model in a field environment. Both studies are based on data from a large-scale field study of marathon runners. (Markle et al. 2014, p. 1) find that 'satisfaction as a function of relative performance (the difference between a runner's time goal and her finishing time) exhibits loss aversion and diminishing sensitivity, consistent with the prospect theory value function'. Furthermore, their results reveal a discontinuity (or jump) at the reference point which suggests a distinct shape of the value function around goals. Finally, they reveal that goal importance amplifies loss aversion, 'that multiple reference points simultaneously impact runner satisfaction, and that loss aversion is overestimated in predictions of satisfaction, but still present in actual experienced satisfaction' (Markle et al. 2014, p. 1). Allen et al. (2014, p. 3) analyze the finishing times of marathon runners and find ‘a lumpy distribution of finishing times, with bunching just ahead of round numbers’ (e.g., at each 30 and 60-minute mark). They provide evidence that this effect is caused in ‘part by pacing and planning and in part by effort provision over the final 2.195 kilometers of the marathon’ (Allen et al. 2014, p. 3). Finally, they build different models of reference dependence and illustrate that the shape of the distribution of the finishing times can be modeled with a simple model of prospect theory using standard prospect theory parameter estimates from the lab. With our paper we build on these studies by providing field evidence for the applicability of Heath et al.’s model (1999) to predict user performance in the presence of ‘mere’ goals.

Gamification and Badges

Gamification refers to ‘using game design elements in non-gaming contexts’ (Deterding et al. 2011). In the context of online communities or social media sites, gamification is used in order to activate user contribution behavior and encourage the social interaction between users (Hamari 2013). One popular game element are so-called badges (Hamari et al. 2014). ‘Badges are given to users for particular contributions to a site, such as performing a certain number of actions of a given type’ (Anderson et al. 2013). They have been implemented in a variety of online contexts, including education (e.g., Khan Academy), social news (e.g., Huffington Post), knowledge-creation (e.g., Wikipedia), location-based social networking tools (e.g., Foursquare), and many others (e.g., Anderson et al. 2013, Denny 2013). Depending on the type of online community or social media site (i.e., whether leisure or job related) badges might represent either ‘mere’ goals or extrinsic rewards. Consequently, an explanation rooted in psychology or based on economic calculus might be more appropriate to model the impact of badges on user behavior.

Whilst a body of literature has recently emerged which analyzes the impact of badges on user contribution levels (e.g., Hamari 2013, Denny 2013, Mutter & Kundisch 2014a, 2014b), only one article by Anderson et al. (2013) develops an analytical model analyzing how badges affect user contribution behavior in the Q&A community Stack Overflow. Anderson et al. model badges as extrinsic rewards because their research environment is primarily used by computer programmers interested in programming issues and provides a platform for them to signal their ability to the labor market (Lerner & Tirole 2002). Anderson et al.'s model (2013) predicts that users increase their contribution quantity for the rewarded activity with proximity toward a badge, partially at the expense of the non-rewarded activities. After earning a badge, the contribution quantity of each activity jumps immediately back to its original level. To support the predictions from their model, Anderson et al. (2013) select a few badges from the platform and plot the number of user contributions in the presence of those badges. Their graph analysis suggests that in the days prior to earning a badge users increase the quantity of the type of activity needed to earn the next badge. In other words and in accordance with goal setting theory, the goals both energize and focus the attention and efforts of users.

In our work, we investigate empirically how the achievement of badges representing a 'mere' goal affects user contribution behavior. In contrast to Anderson et al. (2013), we use data from a leisure related Q&A community and thereby can rule out potential spillovers to the labor market or confounding effects caused by any other type of external reward. With our work, we contribute to the literature on gamification by providing empirical evidence for the explanatory power of Heath et al.'s model (1999) for the impact of badges as 'mere' goals on user contribution behavior.[4]

The next section introduces the research environment, followed by the explicit formulation of our research hypotheses.

Research Environment[5]

The website at the center of our analysis was launched in January 2006 and will remain anonymous at the owner's request. The platform offers registered and non-registered users the opportunity to ask questions to community members on everyday topics (e.g., beauty, computers, gardening). In other words, the platform deals exclusively with leisure-related topics, rather than ones related to the labor-market, which is why we define the goals as 'mere' goals. All registered users automatically participate in the virtual reward system of the community. For almost all their activities, registered users receive an incentive in the form of so-called status points. In Table 1, we present a list of the main activities and the corresponding status point scheme. Almost all (99%) status points are earned by users taking part in one of the two main activities, *asking* and *answering questions*. Some activities (e.g., *inviting new members to the platform*) play only a very minor role, accounting for less than 1% of the total number of accumulated status points.

More specifically, an overall 76% of the accumulated status points are earned with the activity *answering questions*. Depending on the quality of their answer, users can earn between 0 and 25 status points for a given answer. The answer quality is rated by both the questioner and by other members of the community. Users earn an average of 4 status points per answer. Apart from the activity *answering questions*, registered users can also get status points by *asking questions* to the community. If a question receives at least one answer or is rated as 'helpful' by at least one other user, the questioner receives between 1 and 4 status points. No status points are earned, however, if the question remains unanswered. On average, users earn 3 status points per question. As they accumulate status points, users automatically move up in an ascending ranking system of 20 hierarchical badges. For each badge, users need to earn a predetermined - but varying - number of status points. In Table 2 we provide a detailed list of available badges and the status points required for each badge.

The labels of the first nine badges are noticeably hierarchical, such as ‘Beginner’, ‘Student’, ‘Bachelor’ and so on. For example, the badge ‘Master’ requires an accumulation of at least 1,030 status points. With an average of 4 status points per answer given, users would have to answer more than 250 questions to get the ‘Master’ badge. The list with the badges and the required status points for each badge are also publicly available on the platform. The badge and the total number of earned status points are displayed in the personal profile of each user. Both pieces of information are also publicly visible to other platform users or guests whenever a user poses or answers a question.

Hypotheses Development

Performance is defined as the result of work over time (Heath et al. 1999) and is dependent on effort and persistence. This means that users’ performance levels on any given day are equal to the amount of time they spend on the platform combined with the effort they make to contribute to the online activities of the community. In other words we can use the number of answers and questions per user per day as a proxy for user performance (see Table 1). According to theory, users decide every day on their optimal performance level by assessing the marginal benefits and costs of one additional unit of performance. The slope of the benefit and cost function depends on the performance level on that day, however, only the slope of the benefit function alters with the distance toward the next badge (i.e. marginal benefits for one activity increase with proximity towards a badge). In contrast, the slope of the cost function remains roughly similar each day and does not depend on the proximity towards the next badge. If badges do act as reference points, and the diminishing sensitivity principle applies, then we would expect to see the slope of the benefit function increase continuously with proximity to the next badge, and decrease with greater distance away from the badge. Thus, users are expected to increase their performance on each successive day as they come closer to the next badge and to decrease their daily performance successively in the following days. To test the transferability of these two properties of the value function we therefore formulate the following two hypotheses:

HYPOTHESIS I: Users increase their average performance level successively with increasing proximity towards a badge.

HYPOTHESIS II: Users decrease their average performance level successively after a badge has been reached.

Moreover, the remaining property assumes that individuals are risk averse. If users are risk averse, the average performance in the days shortly before earning a badge should be higher than the average performance in subsequent days. Therefore, we formulate a third hypothesis to test the remaining assumption:

HYPOTHESIS III: The average performance level is higher before a badge has been reached than after the event.

Dataset, Sample & Descriptive Statistics

Dataset

We are very fortunate in having a unique dataset at our disposal – kindly provided by the operator of this Q&A community – as this allows us to analyze how users adjust their performance in the presence of badges. The entire dataset covers all user activity on the platform between the beginning of February 2007 and the end of May 2008, i.e., an observation period of 462 days. During this observation period, 316,142 unregistered visitors posed a question to the community, and 73,017 new users registered on the platform. Our dataset enables us to observe that these users replied to 874,927 posted questions with 2,520,192 answers. Due to the fact that we have data on the user level, we know exactly when a user registers on the platform, when and how often this user performs a certain activity, when and how many status points she earns for her activities, and when she earns a badge.

Sample

For our empirical analysis, we aggregate the activity data on a daily level to analyze how user contribution behavior changes in the days shortly before and after a badge is earned. We restrict our sample to those users who show some commitment to the community by earning at least one badge during their membership. We also drop from our sample all the observations of users who stop to perform any of the platform's activities and thus become inactive. This leaves us with an unbalanced panel of 5,939 users and 1,312,665 observations on a daily level over a period of 462 days.

Descriptive Statistics

Activity History of Users

Table 3 presents selected descriptive statistics for our sample. On average, we observe users for 221 days (*Sum of Active Days*) before they become inactive and stop contributing to the platform. During the observation period, users contribute an average of 382 answers each (*Sum of Answers*) and ask 65 questions (*Sum of Questions*). As can be seen from the quantiles of the distributions, there is a strong heterogeneity in the history of user participation. The median values differ substantially from the mean values for the main activities as well as for the number of days required to reach the next badge. This reveals that a substantial share of activities is performed by a small number of top contributors.

Distribution of Badges

The users in our sample earn a total of 16,976 badges over the observation period. Table 4 illustrates the distribution of earned badges across the users in our sample. When they register on the platform users automatically receive the badge 'Beginner', but from then on they need to collect more status points if they want to get the next badge. For the badge 'Student', users need to earn 210 status points (see Table 3). We observe 5,342 users who collect sufficient status points to earn this badge.

In general, the more challenging a badge the fewer users earn it, as illustrated in Table 4. Our sample includes also users who were already registered on the platform before our observation period started and who hold a more valuable badge than the badge 'Beginner' at the beginning of our observation period. Thus, we do not observe all the 5,939 users in our sample earning the badge 'Student' but only 5,342 users.

Performance Measures

We use the number of *Answers* and the number of *Questions* per day on the user level as measures for user performance. In Table 5, we provide mean, standard deviation, median, 95% quantile, 99% quantile, and maximum value for each of the two variables. Users provide on average 1.73 answers and ask on average 0.29 questions per day.

Empirical Analysis

Main Variables

We use the number of *Answers* and *Questions* per day to measure user performance. We create a set of dummy variables, covering five days before receive a badge (*Day Dummy (-5)* to *Day Dummy (-1)*), five days afterwards (*Day Dummy (+1)* to *Day Dummy (+5)*), and for the day of the promotion (*Day Dummy (0)*) to elucidate how users adjust their performance in the days shortly before and after earning a badge, and on the day of the promotion itself. Additionally, to account for potential fluctuations in activity levels caused by the day of the week we create a set of dummy variables for each day of the week. Furthermore, as users can only answer questions that are open on the platform at any one time, activity levels might vary depending on the overall number of questions that are open at that time. An unanswered question can stay open for a maximum of up to seven days. Hence we calculate the total number of questions per day as a measure of the overall activity level on the platform. By calculating the first differences of the time series we account for non-stationarity. We incorporate the first difference as well as seven lags of this variable into our model.

Model

We estimate a poisson fixed effects model for each of the two performance measures.[6] The model is illustrated in equation (1):

$$Y_{it} = \sum_{\tau=1}^{5} \beta_{-\tau} D_{t-\tau} + \sum_{\tau=0}^{5} \beta_{\tau} D_{t+\tau} + \sum_{\tau=1}^{6} \gamma_{\tau} WD_{\tau t} + \sum_{\tau=0}^{7} \delta_{-\tau} \Delta q_{t-\tau} + \mu_i + \varepsilon_{it} \qquad (1)$$

The variable Y_{it} represents the dependent variables. Each observation in the sample is identified exactly by the index it where i represents the individual and t the day in our observation period. The variable D_t represents a dummy variable for the day on which a user earns a badge. In addition, we include 5 lags ($\beta_{-1}, \ldots, \beta_{-5}$) and 5 leads ($\beta_{+1}, \ldots, \beta_{+5}$) of this variable to capture average activity levels across 5 days before and five days after the promotion. In addition, we add a set of weekday dummies $WD_{\tau t}$ and the first difference as well as 7 lags of the first difference of the overall number of questions on the platform ($\delta_0, \ldots, \delta_{-7}$). Finally, we include user-specific fixed effects μ_i and the error term ε_{it} in our model.

Identification

We use individual-specific fixed effects to account for unobserved time constant heterogeneity (Wooldridge 2010). In order to analyze how users adjust their performance in the presence of badges we investigate how average user performance alters in the five days before and after users earn a badge, and on the day of the promotion. If the estimators for the variables *Day Dummy (-5)* to *Day Dummy (0)* increase successively in the days prior to earning the next badge, peak on the day of the promotion, and decrease successively in the subsequent days *Day Dummy (+1)* to *Day Dummy (+5),* this would confirm that badges act as reference points and that the diminishing sensitivity principle applies (see *HYPOTHESIS I* and *HYPOTHESIS II*). Furthermore, if the average performance is higher in the five days before users earn a badge *Day Dummy (-5)* to *Day Dummy (-1)* compared to in the five subsequent days *Day Dummy (+1)* to *Day Dummy (+5)*, this would support the assumption that users are risk averse (see *HYPOTHESIS III*).

Results

The results for the two performance measures are illustrated in Table 6. The independent variables are presented in the first column, and the results for the number of *Answers* and *Questions* in column two and three. The estimators for the variables *Day Dummy (-5)* to *Day Dummy (+5)* reveal how the contribution quantity differs on each corresponding day.

For the number of *Answers,* all estimators for the day dummies have a positive sign and are significant on a one percent level. The estimators increase continuously from 0.361 or 43% in the five days before (*Day Dummy (-5)*) to 0.681 or 98% on the day of the promotion (*Day Dummy (0)*).[7] This means that the quantity of answers increases by 55 ppt. or by approximately 0.95 answers per day.

On the first day after the promotion, we observe a drop in the quantity of contributions. The estimator for the *Day Dummy (+1)* is 0.467 or 60% and thus substantially lower compared to the estimator for the *Day Dummy (0)*. The difference is -38 ppt. or by approximately 0.65 answers per day. In the following days, the estimators for the day dummy variables decrease continuously from 0.467 or 60% on the first day following the promotion (*Day Dummy (+1)*) to 0.155 or 17% in the five subsequent days (*Day Dummy (+5)*). This means that the quantity of answers decreases by 43 ppt. or by approximately 0.74 answers per day.

In the chart on the left of Figure 3 we illustrate the estimators for the day dummy variables in absolute terms. The dashed vertical line represents the day of the promotion. The chart illustrates that the number of *Answers* increases progressively with proximity to the next badge, peaks on the day of the promotion and decreases again in subsequent days, with the reduction being strongest on the first day after the badge has been earned. It is important to note that the contribution quantity does not drop immediately back to a constant base level on the first day following the promotion but, instead, it is the number of *Answers* that decreases continuously with each day.

In Table 7 we present the differences between the consecutive day dummy variables and the corresponding chi-squared test results. The differences between the estimators and the test results for the number of *Answers* and *Questions* are presented in column two and three.

For the number of *Answers*, the calculated differences between the estimators for the day dummy variables align perfectly with the chart in Figure 1. All differences are significant on a one percent level. As expected, the differences are positive in the days preceding the promotion and negative afterwards. More importantly, the differences are increasing in size with proximity to the day of the promotion, and decreasing subsequently. This reflects the slope of the prospect theory value function which is increasing in the domain of losses (failure) and decreasing in the domain of gains (success).

The results are equivalent for the second performance measure, number of *Questions*. The estimators for the variables *Day Dummy (-5)* to *Day Dummy (+5)* have a positive sign and are significant on a one percent level (see Table 6). The estimators increase continuously from 0.397 or 49% in the five days before (*Day Dummy (-5)*) to 1.332 or 279% on the day of the promotion (*Day Dummy (0)*). This means that the quantity of questions increases by 230 ppt. or by approximately 0.7 questions per day. On the first day after the promotion, we observe a drop in the quantity of contributions. The estimator for the *Day Dummy (+1)* is 0.653 or 92% and thus substantially lower compared to the estimator for the *Day Dummy (0)*. The difference is -187 ppt. or approximately 0.54 questions per day. On subsequent days, the estimators for the day dummy variables decrease continuously from 0.653 or 92% on the first day (*Day Dummy (+1)*) to 0.211 or 23% on the fifth day (*Day Dummy (+5)*). This means that the quantity of questions decreases by 69 ppt. or by approximately 0.3 questions per day, in the five days after a badge is earned.

In the chart on the right of Figure 3 we illustrate the estimators for the day dummy variables in absolute terms. The chart illustrates the increase in the number of *Questions* successively with proximity to the next badge, peaks on the day of the promotion and decreases successively in the days afterwards, with the reduction being the most pronounced on the first day after badge achievement. It is important to note that, as for the number *Answers*, the contribution quantity does not drop back immediately to a constant base level on the first day after the promotion but, instead, the number of *Questions* decreases continuously with each day. In Table 7 we present the differences between the consecutive day dummy variables and the corresponding chi-squared test results. Again, the calculated differences between the estimators for the day dummy variables align perfectly with the chart in Figure 3. All differences are significant on a one or five percent level. As expected, the differences are positive in the days preceding the promotion and negative afterwards. More significantly, the differences are increasing in size with proximity to the day of the promotion and decreasing in the days that follow. Thus we derive our first two results:

RESULT I: Users increase their performance successively in the days before earning a badge. From five days before up to the day of the promotion itself, the average number of answers per day increases by 55 ppt. and the average number of questions by 230 ppt.

RESULT II: Users decrease their performance level successively in the days after a badge has been earned with the strongest reduction occurring on the first day after badge achievement. From the first to the fifth day after the promotion, the average number of answers per day decreases by 43 ppt. and the average number of questions by 69 ppt.

Finally, we investigate if the average performance level is higher before users earn a badge compared to afterwards. First, we test for both performance measures if the estimator for the *Day Dummy (-1)* is larger than the estimator for the *Day Dummy (+1)*. For the number of *Answers* the difference is 25 ppt. and significant on a one percent level ($\chi^2(1) = 86$, $p<0.01$). For the number of *Questions* the difference is 34 ppt. and also significant on a one percent level ($\chi^2(1) = 36$, $p<0.01$). Second, we aggregate the day dummy variables in the five days before (*Day Dummy (-5)* to *Day Dummy (-1)*) and after (*Day Dummy (-5)* to *Day Dummy (-1)*) users earn a badge and thereby create two dummy variables covering either the five days before or after the day of the promotion. We update equation (1), estimate our model again for both performance measures, calculate the differences between the new dummy variables, and test for significance. For the number of *Answers* the difference is 36 ppt. and significant on a one percent level ($\chi^2(1) = 289$, $p<0.01$). For the number of *Questions* the difference is 30 ppt. and also significant on a one percent level ($\chi^2(1) = 96$, $p<0.01$). Thus, we derive our third result:

RESULT III: User performance is higher in the days before a badge is earned compared to the days afterwards. For the number of answers the difference between one or five days before a badge is earned compared to one or five days afterwards is 25 ppt. and 36 ppt. For the number of questions the difference is 34 ppt. and 30 ppt.

Summary of Findings

We find that users increase their performance level successively with proximity towards a badge (*RESULT I*) and decrease their performance level successively afterwards (*RESULT II*). Thus, we find support for *HYPOTHESIS I* and *HYPOTHESIS II* which represent two of the three properties of the value function, *reference point dependence* and *diminishing sensitivity*. In addition, *RESULT III* reveals that the average user performance is higher in the days preceding the promotion than in the following days. This confirms *HYPOTHESIS III* which covers the remaining property that individuals are risk averse. Overall, we find empirical evidence for the transferability of the main properties of the prospect theory value function to 'mere' goal behavior.

Robustness Checks

In order to demonstrate the robustness of our results we examine a number of robustness checks, run separately for each performance measure. (1) We estimate the model for each badge on the platform (see Table 2) separately; (2) we estimate our main model again and drop the observations from our sample where users took fewer than eleven days to earn the next badge; (3) to rule out that our results are driven by outliers we recode the values of both quantity measures which lie above the 99% quantile with the value of the quantile. For each of these robustness checks our main results remain qualitatively unchanged.

Conclusion

Heath et al. (1999) develop a prospect theory model of goal behavior to explain why 'mere' goals lead to higher performance. Their model inherits the properties of the prospect theory value function, notably reference point dependence, loss aversion, and diminishing sensitivity (Kahneman & Tversky 1979) and offers insights on how goals affect performance on the individual as well as on the group level. With this paper we provide field evidence for the applicability of Heath et al.'s model (1999) to predict user performance in the days surrounding goal achievement. In particular, we take user activity data from a popular German Q&A-community and analyze empirically how users alter their performance in the presence of 'mere' goals represented by badges. We find that users increase their performance successively in the days before earning a badge with performance peaking on the day of the promotion. In the days succeeding the promotion, users decrease their performance progressively while the reduction in user performance is strongest on the first day after badge achievement. Overall, the performance of users is higher in the days before earning a badge compared to subsequent days. These findings reflect the characteristic S-shape of the prospect theory value function which is convex below the reference point and concave above it. Thus, in line with Markle et al. (2014) and Allen et al. (2014), we provide empirical evidence for the transferability of the main properties of the prospect theory value function (i.e., reference point dependence, loss aversion, and diminishing sensitivity) to goal behavior in the field. Our results also suggest a distinct shape of the value function around goals because the reduction in user performance is strongest on the first day after successful badge achievement compared to the decline in subsequent days.

With this paper we contribute to research on goal setting by providing field evidence for the applicability of the model provided by Heath et al. (1999) for the prediction of user performance in the presence of 'mere' goals. Thus, we present further evidence that the prospect theory value function provides 'a unifying explanation for findings in both the goal literature and the judgment and decision making literature' (Heath et al. 1999, p. 80). Thereby, we contribute to the connection of the goal setting literature and the decision making literature which might bear the potential to derive new insights in the future for each. In addition, we also contribute to the recent literature on gamification (e.g., Blohm & Leimeister 2013, Hamari et al. 2014) by providing a theoretical explanation and empirical evidence for the impact of badges in the form of 'mere' goals on user contribution behavior.

Although our findings are overall consistent with the theory, we recognize that there might be other factors (e.g., the topic or thematic areas of the platform) that we have not accounted for but that might also be playing a role in our research setting. While the results from the Q&A community under study may not be directly applicable to other domains, our findings are, nevertheless, suggestive. Previous research in the domain of knowledge contribution in online communities has emphasized that user contribution behavior is influenced by both idealistic and altruistic factors (e.g., Krankanhalli et al. 2005, Jeppesen & Frederiksen 2006). Thus, it is not unreasonable to expect that the prospect theory model for goal behavior from Heath et al. (1999) applies to any other contexts in which ‘mere’ goals are used for motivational purposes.

The extension and refinement of the prospect theory model for goal behavior from Heath et al. (1999) represents, in our opinion, a promising avenue for future research. One starting point might be to investigate the precise causes of the discontinuity at the reference point. This is important for a more nuanced understanding of the distinct shape of the value function for goal behavior. Another opportunity might be to systematically adjust the goals’ level of difficulty on the individual level to find further evidence for the explanatory power of the value function or to test the model predictions for subgoals, for example, including to alleviate the ‘starting problem’ when individuals face very challenging goals.

Figures

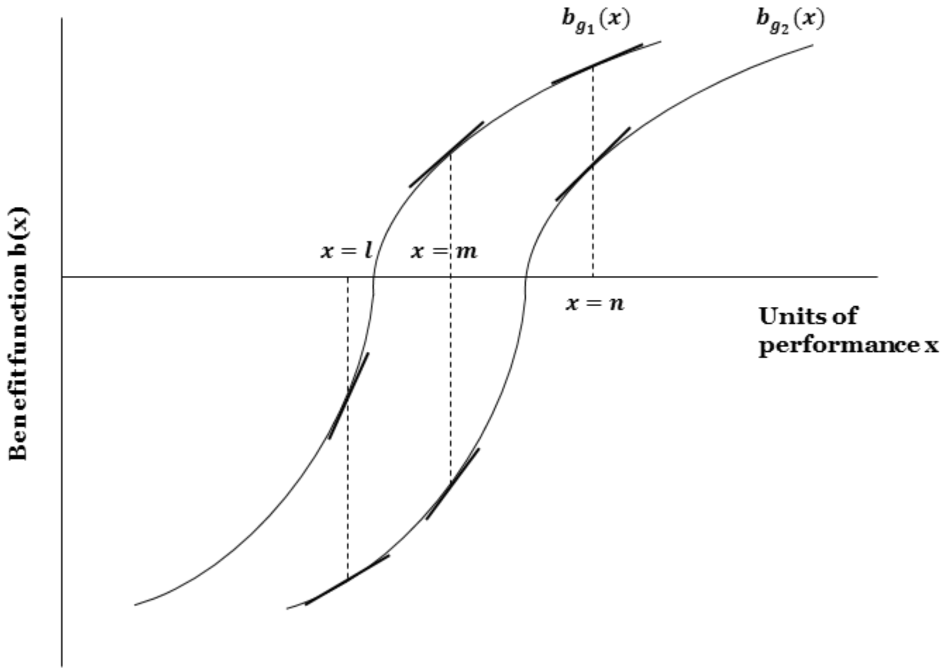

Figure 1: Value functions for two different goals (Heath et al. 1999, p. 87)

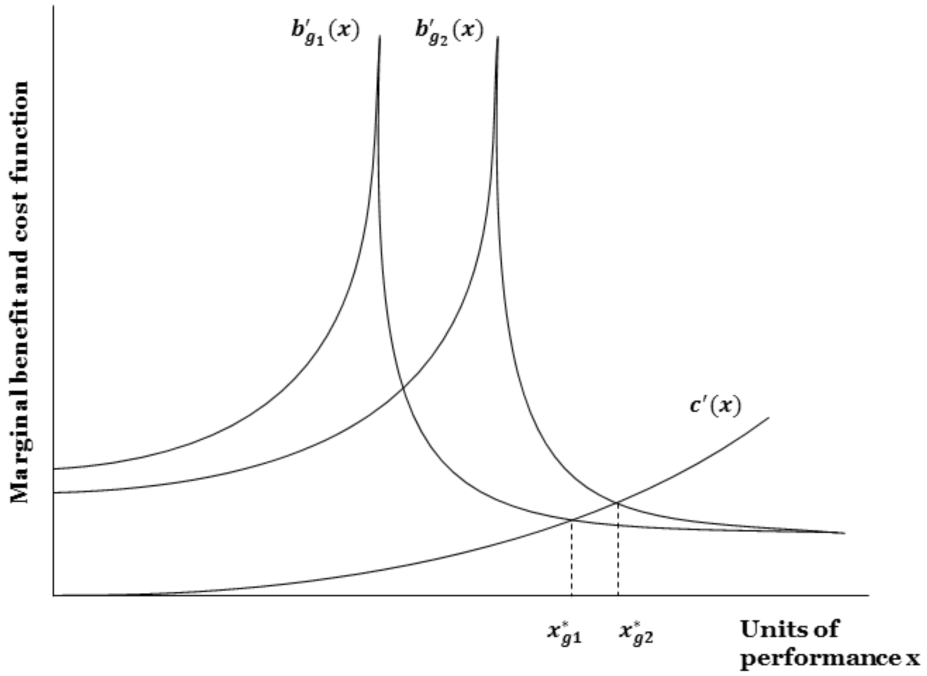

Figure 2: First derivatives of value function for two different goals (Heath et al. 1999, p. 97)

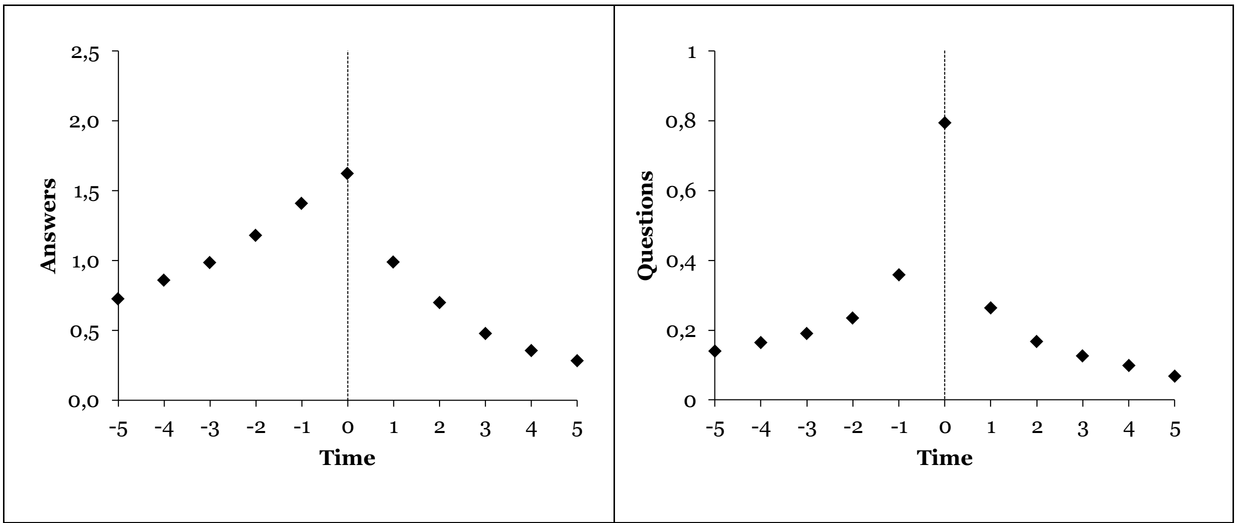

Figure 3: User Performance per Day

Tables

Table 1: Status Point Scheme			
Main Activities	Status Points per Activity	Average of Status Points Received	Ratio of Total Status Points
Answering Questions	0 – 25	4	76%
Asking Questions	0 – 4	3	23%

Table 2: List of Badges			
Label of Badge	Required Status Points	Label of Badge	Required Status Points
Beginner	0	Robert Koch	8,240
Student	210	Immanuel Kant	8,740
Bachelor	530	Archimedes	9,240
Master	1,030	Max Planck	9,740
Research Assistant	1,630	Isaac Newton	10,240
Doctor	2,430	T. A. Edison	10,740
Assistant Professor	3,330	Pythagoras	11,240
Professor	4,240	Galileo Galilei	11,740
Nobel Laureates	5,240	Leonardo da Vinci	12,240
Albert Schweitzer	7,740	Albert Einstein	>12,740

Table 3: Users' Activity History							
Variables	Mean	Min	Q25	Median	Q75	Max	Sum
Sum of Active Days	221	1	107	210	326	462	1,312,665
Sum of Answers	382	0	58	124	334	16,834	2,263,940
Sum of Questions	65	0	10	28	68	1,943	384,252
Number of Days for Promotion	46	1	11	22	53	461	-

Table 4: Distribution of Badges			
Label of Badge	Number of Promotions	Label of Badge	Number of Promotions
Beginner	-	Robert Koch	247
Student	5,342	Immanuel Kant	224
Bachelor	3,313	Archimedes	208
Master	2,086	Max Planck	193
Research Assistant	1,486	Isaac Newton	183
Doctor	999	T. A. Edison	163
Assistant Professor	726	Pythagoras	161
Professor	539	Galileo Galilei	150
Nobel Laureates	421	Leonardo da Vinci	141
Albert Schweitzer	266	Albert Einstein	128

Table 5: Quantity of Users' Contributions						
Variables	Mean	Std.	Median	Q95	Q99	Max
Answers	1.73	5.60	0	10	30	218
Questions	0.29	1.17	0	2	5	254

Table 6: Analysis of Performance Measures				
Variables	*Answers*		*Questions*	
Day Dummy (-5)	**0.361****	(0.0274)	**0.397****	(0.0435)
Day Dummy (-4)	0.416**	(0.0312)	0.454**	(0.0456)
Day Dummy (-3)	0.466**	(0.0377)	0.509**	(0.0516)
Day Dummy (-2)	0.536**	(0.0395)	0.598**	(0.0527)
Day Dummy (-1)	0.613**	(0.0402)	0.814**	(0.0492)
Day Dummy (0)	**0.681****	(0.0417)	**1.332****	(0.0481)
Day Dummy (+1)	**0.467****	(0.0414)	**0.653****	(0.0524)
Day Dummy (+2)	0.350**	(0.0401)	0.461**	(0.0564)
Day Dummy (+3)	0.252**	(0.0340)	0.365**	(0.0496)
Day Dummy (+4)	0.193**	(0.0280)	0.294**	(0.0472)
Day Dummy (+5)	**0.155****	(0.0236)	**0.211****	(0.0423)
Control Variables	✓		✓	
Individual Fixed Effects	✓		✓	
Observations	1,239,912		1,193,901	
Number of Users	5,753		5,455	
-Ln Likelihood	2,639,181		691,263	
Cluster Robust Standard Errors in Parentheses, ** p<0.01, * p<0.05				

Table 7: Differences Day Dummies		
Differences Estimators	*Answers*	*Questions*
DD (-4) - DD (-5)	0.0815** $\chi^2(1) = 164$	0.0881* $\chi^2(1) = 56$
DD (-3) - DD (-4)	0.0770** $\chi^2(1) = 112$	0.0875* $\chi^2(1) = 42$
DD (-2) - DD (-3)	0.1168** $\chi^2(1) = 237$	0.1558** $\chi^2(1) = 128$
DD (-1) - DD (-2)	0.1365** $\chi^2(1) = 364$	0.4390** $\chi^2(1) = 761$
DD (0) - DD (-1)	0.1291** $\chi^2(1) = 271$	1.529** $\chi^2(1) = 4{,}869$
DD (+1) - DD (0)	-0.3806** $\chi^2(1) = 2{,}318$	-1.865** $\chi^2(1) = 6{,}052$
DD (+2) - DD (+1)	-0.1756** $\chi^2(1) = 715$	-0.3364** $\chi^2(1) = 500$
DD (+3) - DD (+2)	-0.1324** $\chi^2(1) = 334$	-0.1454** $\chi^2(1) = 132$
DD (+4) - DD (+3)	-0.0741** $\chi^2(1) = 153$	-0.0990** $\chi^2(1) = 70$
DD (+5) - DD (+4)	-0.0453** $\chi^2(1) = 75$	-0.1060** $\chi^2(1) = 91$
Chi-Squared Test, ** $p<0.01$, * $p<0.05$		

Footnotes

[1] The following paragraphs provide a brief summary of the discussion in Heath et al. (1999) and Wu et al. (2008).

[2] In the following paragraphs, we describe the model devised by Heath et al. (1999). A more detailed description of the model can be found in Wu et al. (2008).

[3] Heath et al. (1999) provide an extensive discussion of how their model can be used to make predictions on the group level and how both the group and the individual level are related.

[4] In a concurrent but independent work, Goes et al. (2014) examine the impact of badges on user contribution behavior in the context of an IT related community. The authors find that users increase their contribution level before they earn a badge and substantially reduce it afterwards. Our study differs in context, data granularity and scope. Goes et al. (2014) use data from the IT community on a weekly level while we work with daily data from a leisure related community where users can ask everyday questions (e.g., on beauty, computers, gardening). We explicitly investigate whether the main properties of the prospect theory value function (i.e., reference dependence, loss aversion, and diminishing sensitivity) transfer to goal behavior. In contrast, Goes et al. (2014) are interested in the overall impact of goals on performance, and, although using Heath et al.'s model (1999) to support their hypotheses development they build on the assumption that the value function explains goal behavior.

[5] Two related papers by Mutter & Kundisch (2014a, 2014b) are drawing on the same research environment. Despite some overlap in the underlying dataset, the related studies differ in their scope, each addressing independent research questions.

[6] We estimate a poisson model to consider the distribution properties of both dependent variables (i.e., only non-negative integer values and large number of zeros). To account for overdispersion and autocorrelation in the data, we use cluster robust standard errors (Cameron & Trivedi 2013).

[7] To get an approximation for the absolute effect size (e.g., number of answers per day) we multiply the relative effect (or semielasticity) with the mean value of the corresponding variable (see Table 6). For example, for the *Day Dummy (0)* we get the absolute effect of $\left((\exp(0.681) - 1) \times 100\right) \times 1.73 = 1.69$ answers per day.

References

Allen, E. J., P. M. Dechow, D. G. Pope, G. Wu. 2013. Reference-Dependent Preferences: Evidence from Marathon Runners. NBER Working Paper No. 20343. Available at NBER: http://www.nber.org/papers/w20343 (Accessed: 15/01/2015).

Anderson, A., D. Huttenlocher, J. Kleinberg, J. Leskovec. 2013. Steering User Behavior with Badges. *Proceedings of the 22nd International Conference on World Wide Web*, Rio de Janeiro, Brasil.

Bandura, A., D. H. Schunk. 1981. Cultivating competence, self-efficacy, and intrinsic interest through proximal self-motivation. *Journal of Personality and Social Psychology* 41(3) 586–598.

Benartzi, S., R. H. Thaler. 1995. Myopic loss-aversion and the equity premium puzzle. *Quarterly Journal of Economics* 110(1) 75–92.

Botterill, C. B. 1977. Goal setting and performance on an endurance task. University of Alberta, unpublished doctoral dissertation.

Blohm, I., J. M. Leimeister. 2013. Gamification. Design of IT-Based Enhancing Services for Motivational Support and Behavioral Change. *Business & Information Systems Engineering* 5(4) 275–278.

Brown, J. S. 1948. Gradients of Approach and Avoidance Responses and their Relation to Level of Motivation. *Journal of Comparative and Physiological Psychology* 41(6) 450–465.

Cameron, A. C., P. K. Trivedi. 2013. *Regression Analysis of Count Data*. Cambridge University Press.

Deterding, S., R. Khaled, L. E. Nacke, D. Dixon. 2011. Gamification: Toward a Definition. *Gamification Workshop Proceedings*, Vancouver, BC, Canada.

Denny, P. 2013. The Effect of Virtual Achievements on Student Engagement. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, 763–772.

Drèze, X., J. C. Nunes. 2006. The Endowed Progress Effect: How Artificial Advancement Increases Effort. *Journal of Consumer Research* 32(4) 504–512.

Drèze, X., J. C. Nunes. 2011. Recurring Goals and Learning: The Impact of Successful Reward Attainment on Purchase Behavior. *Journal of Marketing Research* 48(2) 268–281.

Garland, H. 1982. Goal levels and task performance: A compelling replication of some compelling results. *Journal of Applied Psychology* 67(2) 245–248.

Goes, P., C. Guo, M. Lin. 2014. Do Incentive Hierarchies Induce User Effort? Evidence from an Online Knowledge Exchange. Available at SSRN: http://ssrn.com/abstract=2413052 (Accessed: 01/09/2014).

Hamari, J., V. Eranti. 2011. Framework for Designing and Evaluating Game Achievements. *Proceedings DiGRA 2011: Think Design Play,* 122–134.

Hamari, J. 2013. Transforming Homo Economicus into Homo Ludens: A field Experiment on Gamification in a Utilitarian Peer-to-Peer Trading Service. *Electronic Commerce Research and Applications* 12(4) 236–245.

Hamari, J., J. Koivisto, H. Sarsa. 2014. Does Gamification Work? – A Literature Review of Empirical Studies on Gamification. *Proceedings of the 47th Hawaii International Conference on System Sciences*.

Heath, C., R. P. Larrick, G. Wu. 1999. Goals as Reference Points. *Cognitive Psychology* 38(1) 79–109.

Heilizer, F. 1977. A Review of Theory and Research on the Assumptions of Miller's Response Competition (Conflict) Models: Response Gradients. *The Journal of General Psychology* 97(1) 17–71.

Hull, C. L. 1932. The Goal-Gradient Hypothesis and Maze Learning. *Psychological Review* 39(1) 25–43.

Jeppesen, L. B., L. Frederiksen. 2006. Why Do Users Contribute to Firm-Hosted User Communities? The Case of Computer Controlled Music Instruments. *Organization Science* 17(1) 45–63.

Kahneman, D., A. Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* 47(2) 263–291.

Kankanhalli, A., B. Tan, K. K. Wei. 2005. Contributing Knowledge to Electronic Knowledge Repositories: An Empirical Investigation. *MIS Quarterly* 29(1) 113–143.

Kivetz, R., O. Urminsky, Y. Zheng. 2006. The Goal-Gradient Hypothesis Resurrected: Purchase Acceleration, Illusionary Goal Progress, and Customer Retention. *Journal of Marketing Research* 43(1) 39–58.

Kramer, R. M. 1989. Windows of vulnerability or cognitive illusions? Cognitive processes and the nuclear arms race. *Journal of Experimental Social Psychology* 25(1) 79–100.

LaPorte, R.E., R. Nath. 1976. Role of performance goals in prose learning. *Journal of Educational Psychology* 68(3) 260–264.

Latham, G. P., E. A. Locke. 1991. Self-regulation through goal setting. *Organizational behavior and human decision processes* 50(2) 212–247.

Latham, G. P. E. A. Locke. 2007. New developments in and directions for goal-setting research. *European Psychologist* 12(4) 290.

Levitt, S. D., J. A. List. 2007. What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives* 21(2) 153–174.

Locke, E. A., J. F. Bryan. 1969. The directing function of goals in task performance. *Organizational Behavior and Human Performance* 4(1) 35–42.

Locke, E. A., G. P. Latham. 1990. *A theory of goal setting & task performance*, Prentice-Hall.

Locke, E. A., G. P. Latham. 2002. Building a Practically Useful Theory of Goal Setting and Task Motivation: A 35-Year Odyssey. *American Psychologist* 57(9) 705–717.

Markle, A., G. Wu, R. J. White, A. M. Sackett. 2014. Goals as Reference Points in Marathon Running: A Novel Test of Reference Dependence. Fordham University Schools of Business Research Paper No. 2523510. Available at SSRN: http://ssrn.com/abstract=2523510 or http://dx.doi.org/10.2139/ssrn.2523510 (Accessed: 12/05/2014).

McNeil, B. J., S. G. Pauker, A. Tversky. 1988. On the framing of medical decisions. D. E. Bell, H. Raiffa, A. Tversky, eds. *Decision making: Descriptive, normative, and prescriptive interactions*, Cambridge: Cambridge Univ. Press, 562–568.

Mento, A. J., R. P. Steel, R. J. Karren. 1987. A meta-analytic study of the effects of goal setting on task performance: 1966-1984. *Organizational Behavior and Human Decision Processes* 39(1) 52–83.

Mutter, T., D. Kundisch. 2014a. Don't Take Away my Status!–Evidence from the Restructuring of a Virtual Reward System. *Computer Networks* 75 477–490.

Mutter, T., D. Kundisch. 2014b. Behavioral Mechanisms Prompted by Badges: The Goal-Gradient Hypothesis. *Proceedings of the 2014 International Conference on Information Systems*, Auckland, New Zealand.

Nemeroff, W.F., J. Cosentino. 1979. Utilizing feedback and goal setting to increase performance appraisal interviewer skills of managers. *Academy of Management Journal* 22(3) 566–576.

Ness, R.G., R.W. Patton 1979. The effect of beliefs on maximum weight-lifting performance. *Cognitive Therapy and Research* 3(2) 205–211.

Odean, T. 1998. Are investors reluctant to realize their losses? *Journal of Finance* 53(5) 1775–1798.

Quattrone, G. A., A. Tversky. 1988. Contrasting rational and psychological analyses of political choice. *American Political Science Review* 82(03) 719–736.

Rothkopf, E. Z., M. J. Billington, 1979. Goal-guided learning from text: Inferring a descriptive processing model from inspection times and eye movements. *Journal of Educational Psychology* 71(3) 310–327.

Sales, S. M. 1970. Some effects on role overload and role underload. *Organizational Behavior and Human Performance* 5(6) 592–608.

Stevenson, M. K., F. H. Kanfer, J. M. Higgins. 1984. Effects of goal specificity and time cues on pain tolerance. *Cognitive Therapy and Research* 8(4) 415–426.

Thaler, R. H. 1985. Mental accounting and consumer choice. *Marketing Science* 4(3) 199–214.

Wooldridge, J. M. 2010. *Econometric Analysis of Cross Section and Panel Data*. The MIT Press, Cambridge.

Wu, G., C. Heath, R. Larrick. 2008. A prospect theory model of goal behavior. Working Paper, Graduate School of Business, University of Chicago (Version April 22, 2008)

Wyer, R. S., T. K. Srull, S. E. Gordon, J. Hartwick. 1982. Effect of processing objectives on the recall of prose material. *Journal of Personality and Social Psychology* 43(4) 674–688.