Journal of Economic Behavior and Organization 219 (2024) 25–37

Contents lists available at ScienceDirect

Journal of Economic Behavior and Organization

journal homepage: www.elsevier.com/locate/jebo

Research Paper

Expectations or rational expectations? A theory of systematic goal deviation ☆

Benjamin Young

University of Technology Sydney, Sydney, 2007, Australia

ARTICLE INFO

JEL classification:
D84
D90

Keywords:
Goals
Goal deviation
Reference points
Rational expectations
Optimal expectations

ABSTRACT

A planner uses goals to manage a preference disagreement over effort provision with a doer. Goals set output expectations for the doer which affect her behavior due to reference-dependent, loss-averse preferences over output. We characterize the planner's optimal goal and explore when it is *aspirational* versus *achievable*. Specifically, we show that the optimal goal is achieved by the doer only if the extent of preference disagreement is relatively small. Instead, when the extent of preference disagreement is large, the doer falls short of the optimal goal. The stochasticity of output plays an important role in generating this prediction within our model.

1. Introduction

This paper is motivated by three empirical facts. First, people set goals, either for themselves or for others.[1] Second, goals serve as a motivating force for individuals.[2] Finally, individuals often deviate from the goals they face.[3] It is natural to hypothesize that such systematic goal deviation is driven by mistaken or 'naive' decision-making. Instead, this paper argues that goal deviation can stem from the optimal decision-making of sophisticated individuals. Hence, goal deviation does not necessarily identify a 'planning fallacy' but, rather, one should *expect* to observe persistent goal deviation in practice.

To motivate the model of goal setting that we provide in this paper, consider the following example based on the experimental study of Clark et al. (2020). A student is studying for their final exam. One study tool at their disposal is to complete practice exams. The more practice exams that the student completes the better their performance will be in the course. This increase in performance,

☆ I would like to thank the co-editor, Friederike Mengel, an associate editor, and two anonymous reviewers for their comments that vastly improved the paper. I am extremely grateful to Roland Bénabou for his advise and support. I would also like to thank Stephen Morris, Alex Jakobsen, Emil Verner, Benjamin Balzer, and seminar participants at Princeton University for their useful feedback and advice. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

E-mail address: benjamin.young@uts.edu.au.

[1] There is evidence that individuals with a need for commitment in domains such as weight loss (Toussaert, 2018) or energy consumption (Harding and Hsiaw, 2014) are willing to subject themselves to goal plans.

[2] Goals can increase both effort provision and performance, especially when goals are task-based (Clark et al., 2020) or set daily (or 'narrowly') (Koch and Nafziger, 2020). See Locke and Latham (2002) for a review of the efficacy of goals in the psychology literature.

[3] Markle et al. (2015) find that only 25% of a group of surveyed marathon runners achieved their self-set goal for finishing time. In another paper, Allen et al. (2016) make explicit note of this by stating that "goals are clearly related to expectations…but goals are not rational expectations." Della Vigna and Malmendier (2006) find evidence that a significant fraction of individuals overpay for the gym based on their actual attendance, which is indirect evidence that people may fall short of their goal to 'get fit'.

https://doi.org/10.1016/j.jebo.2024.01.003
Received 27 June 2022; Received in revised form 14 December 2023; Accepted 4 January 2024

Available online 18 January 2024

however, is stochastic since practice exams are only imperfect representations of the types of questions one might expect to see on the final exam. Suppose that the student has a motivation problem: they over-weight the costs of studying (or completing practice exams) relative to the benefits on course performance. The main question of this paper is, if goals in terms of how many practice exams to complete can be used to motivate this student, how should they be set? In particular, should the goal always be achieved by the student or can aspirational goals that are not achieved do better?

As elucidated in the motivating example, we focus on deviation from input (or effort) goals. This method of process-oriented goal setting is commonly espoused as individuals have more control over inputs (i.e., effort provision) relative to output itself, which is also affected by extraneous noise. While this has been empirically validated, there is still significant deviation from input goals in practice (Clark et al., 2020).[4] Our model provides a rationale for why goals that are systematically deviated from are not only effective motivators, but actually constitute optimal goals.

The model is as follows. There is a planner and a doer. The doer chooses a level of effort that stochastically generates output in the form of either success or failure. Output is valued by both the planner and the doer. The two parties, however, disagree on how costly effort is: the doer over-weights the cost of effort relative to its benefits in comparison to the planner. This preference disagreement provides a channel through which imposing goals on the doer can be useful to the planner.

The planner sets a goal for the doer in the form of a recommended level of effort. Since output is a stochastic function of effort, this recommended effort level forms a stochastic reference point (Kőszegi and Rabin, 2006, 2007) for the doer in the form of an expected distribution over output. Since the doer is assumed to have reference-dependent, loss-averse preferences over stochastic output, this reference-point formation allows goals to serve as a motivating force for the doer.[5] Specifically, setting a goal provides the doer with incentives to increase her own effort provision in order to minimize stochastic psychological losses.

We place no restrictions on the goal that the planner can select. As such, the best response of the doer to a goal does not necessarily have to coincide with the goal itself; that is, we allow for the possibility of goal deviation in equilibrium. We say that goals are *aspirational* if the doer exerts less effort than the goal set by the planner. Instead, we say that goals are *achievable* if the doer exerts more effort than the goal set by the planner. Finally, we say that a goal satisfies *rational expectations* if the effort choice of the doer and the goal choice of the planner coincide. While most previous work that applies stochastic reference points also assumes they are derived from rational expectations, we do not. Thus, a theoretical contribution of our paper is to utilize Kőszegi and Rabin (2006)'s framework of how expectations affect utility (i.e., stochastic reference points) while providing an alternative theory to rational expectations of how such expectations are formed.

We find that goals, generically, do not satisfy rational expectations in our framework. That is, the planner's optimal goal generally does not coincide with the corresponding level of effort exerted by the doer in response to this goal. Hence, goal deviation is a rule, rather than an exception to the rule. More specifically, we find that goals are aspirational when there is a large preference disagreement between the planner and the doer, while goals are achievable when this preference wedge is small. Intuitively, this is because the planner needs to set exacting goals when the preference wedge is large in order to maximally motivate the doer. Instead, goals become less onerous as preference disagreement dissipates since the doer requires less motivation from the planner's perspective. We show that this finding is robust to a number of alternative formulations of the model.

We explore some additional predictions of our model. Interestingly, we find that the planner may be able to achieve her first-best payoff while, simultaneously, the doer is falling short of her goals. Hence, observing goal deviation does not necessarily imply that individuals are engaged in sub-optimal behavior. Indeed, goals may be a sufficient form of soft commitment power such that intervening with an individual facing onerous goals could actually be *welfare decreasing*. This directly challenges the notion that *achievability* is a universally desirable criterion for effective goal-setting. Moreover, this finding suggests that the presence of a preference disagreement between the planner and the doer does not necessarily imply that the planner is willing to pay to force the doer to commit to a course of action.[6] These are important implications for all individuals that set goals, whether it be for themselves or for other people (such as managers setting goals for employees).

The predictions of this model depend crucially on the assumption that output is a stochastic function of effort. To elucidate this point, we provide a version of the model that is identical except for the fact that output is a deterministic function of effort. We show that the set of effort levels that the planner can induce the doer to exert through the use of goals is the same across the stochastic output and deterministic output versions of the model. However, any implementable effort level in the deterministic setting can be induced through the use of a rational-expectations goal. As such, stochasticity is a necessary ingredient for generating equilibrium goal deviation in our framework. This suggests that simply assuming that goals should consistute rational expectations (consistent with Kőszegi and Rabin (2006)) may be suitable in deterministic environments but less suitable in stochastic environments.

The paper proceeds as follows. In Section 2 we discuss the relevant literature. Section 3 provides the formal details of the model. The main findings of the baseline version of the model are presented in Section 4. Finally, Section 5 concludes. All proofs are relegated to the appendix.

[4] For example, in their experiment which fits with the motivating example in the preceding paragraph, Clark et al. (2020) find that task-based or input goals are more effective motivators and are more commonly achieved than their outcome-based counterparts. Nonetheless, 47% of subjects still did not achieve their goal regarding the number of practice exams to complete.

[5] There is evidence that expectations can affect preferences through the channel of reference-dependence. For example, Abeler et al. (2011) find evidence that expectations can affect individual incentives to exert effort, and Karle et al. (2015) show that expectations can affect preferences over consumption.

[6] This is consistent with the empirical observation that the take-up of commitment devices is often low, even when individuals *should* demand them (Ashraf et al., 2006). Perhaps soft commitment in the form of unobserved goal-setting helps to explain some of this missing demand.

2. Literature review

The potential link between goal-setting and reference-dependent, loss averse preferences (as in Kahneman and Tversky (1979)) has long been recognized. In an early contribution, Heath et al. (1999) argue that a goal may form a reference point and, given the presence of loss aversion, an individual will not wish to fall short of this goal. However, their analysis assumes the presence of a goal-induced reference point and does not formally model the process by which goals are formed.

This paper most closely complements a small but growing economics literature on the efficacy of endogenously-determined goals for ameliorating self-control problems in the form of present bias (Phelps and Pollak, 1968; Laibson, 1997).[7] Investigations have been conducted in one-shot settings (Suvorov and Van de Ven, 2008; Koch et al., 2014; Koch and Nafziger, 2016) and repeated settings (Hsiaw, 2013, 2018; Koch and Nafziger, 2020). To model goals, these papers utilize the theory of rational-expectations reference-point formation provided in Kőszegi and Rabin (2006). Consequently, goals are always precisely achieved in equilibrium. In contrast, we do not constrain attention to goals that satisfy rational expectations in this paper. As such, we allow for the possibility of goal deviation as an equilibrium phenomenon.[8]

There are a number of papers that do not *a priori* restrict goals to satisfy rational expectations. Some of these theoretical models find little evidence for the suitability of setting goals that are not achieved (Wu et al., 2008; Brookins et al., 2017). There are, however, a number of exceptions to this. Corgnet et al. (2015) model a situation in which there is information that goals cannot be made contingent upon. Thus, the optimal goal can involve deviation depending on the realization of this information. Similarly, we highlight the important role of uncertainty in goal deviation. We differ, however, in that their notion of uncertainty realizes *before* the choice of effort while uncertainty in our framework is realized *after* the choice of effort. Thus, we complement their work by showing uncertainty can lead to goal deviation even when goals can always feasibly be achieved. Kaiser et al. (2021) establish both theoretically and empirically that the ability to revise one's goals plays an important role in goal deviation. We contribute to their findings by showing that goal deviation can result even when goals cannot be explicitly revised to the individual's benefit.

Carrillo and Dewatripont (2008) provide a model of 'promises', where an individual experiences disutility if she does not take an action she promises to do. They find that, when the marginal cost of downward deviation from a promise is small, agents optimally deviate from promises. In this paper we highlight the importance of an alternative channel to the shape of the psychological cost function: the stochasticity of output as a function of effort. Jain (2009) explores the use of goals in a setting in which individuals are naive regarding their future preferences over effort. Such naiveté permits unrealistic goals to constitute optimal behavior. The current paper shows that naiveté is not a necessary condition for goal deviation: the planner has a sophisticated understanding of the doer's preferences over effort provision. Rather, goal deviation can result from the optimal behavior of a sophisticated decision-maker.[9]

Finally, both Koch and Nafziger (2011) and Koch and Nafziger (2021) analyze models in which output is a stochastic function of effort, where goals take the form of an expected level of output (i.e., are performance goals). As such, ex-post goal deviation is *guaranteed* in these models and, although not explored in these papers, goals may also be deviated from in expectation (i.e., need not be achieved on average). However, both papers have different objectives from the current paper: Koch and Nafziger (2011) focuses on the motivating force of goals and Koch and Nafziger (2021) focuses on the optimality of narrow- versus broad-bracketed goals. Instead, we are directly interested in the extent to which input or effort goals may be deviated from. This is because individuals may have full control over effort and, yet, we still observe goal deviation from effort goals in practice (Clark et al., 2020). We achieve this by separating goals (measured at the level of inputs) from their psychological-utility consequences (measured as loss utility over stochastic output), which stands in contrast to models of performance goals which conflate the two. Thus, our model complements these papers by establishing a novel sense in which stochastic output plays a role in generating goal deviation as an equilibrium prediction.

3. The model

3.1. Formal details

Environment: There is a planner (denoted by P) that sets goals for a doer (denoted by D). The doer chooses a level of effort $e \in [0,1]$. Effort stochastically generates output $y \in \{0, v\}$, where $v > 0$. More precisely, given effort level $e \in [0,1]$, $y = v$ with probability e and $y = 0$ with probability $1 - e$. Let $F(y|e)$ denote the distribution over output given effort level $e \in [0,1]$.

Preferences: There is a wedge between the preferences of the planner and the doer. The planner perceives the cost of effort level e to be $C(e)$. The doer, instead, perceives the cost of effort to be $C(e)/\beta$, where $\beta \in (0,1]$. We assume that C is strictly increasing, strictly convex, continuously differentiable, $C'(0) = 0$, and $\lim_{e \to 1} C'(e) = +\infty$. Taken altogether, these assumptions ensure that both the planner's desired effort level and the doer's actual level of effort are always interior. Hence, the doer over-weights the cost of effort

[7] Goals are, of course, not the only way to ameliorate self-control problems. Indeed, contracts may also be useful in such scenarios (DellaVigna and Malmendier, 2004; Kaur et al., 2015).

[8] Hsiaw (2013) does consider the case in which the goal-setter has only *partial* commitment power: with some probability she can renegotiate the goal against which she evaluates herself so that goals need not satisfy rational expectations. If this is the case, the renegotiated goal is set *as low as possible* so that goals are only ever exceeded. Thus, this does not capture situations in which individuals fall short of their goals.

[9] Additionally, one may expect an individual to eventually 'learn away' an optimistic naiveté bias (Ali, 2011). This seems to be a particularly pertinent issue in the case of goal-setting, where goals are formed and re-shaped dynamically (e.g. New Year's resolutions, professionals setting goals, marathon runners, etc). Instead, in our model, goal deviation is never learned away as it constitutes optimal behavior.

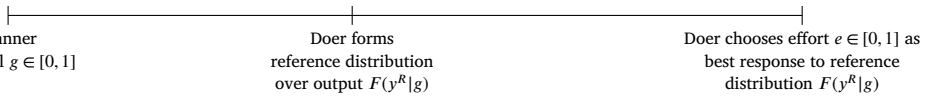

Fig. 1. Timing of the Model.

relative to the planner. Specifically, given effort e and output level y, the payoff to the planner is $u_P(e,y) \equiv y - C(e)$, while the payoff to the doer is $u_D(e,y) \equiv y - C(e)/\beta$, where $\beta \in (0,1]$. Hence, the optimal level of effort from the planner's perspective, e_P^*, solves

$$\max_{e \in [0,1]} ev - C(e), \tag{1}$$

which implies e_P^* is the unique solution to $C'(e) = v$. Instead, the optimal effort level from the perspective of the doer, e_D^*, solves

$$\max_{e \in [0,1]} ev - C(e)/\beta, \tag{2}$$

which implies e_D^* is the unique solution to $C'(e) = \beta v$. Since C is strictly convex, we have that $e_D^* < e_P^*$ whenever $\beta \in (0,1)$, with e_D^* moving further away from e_P^* as β decreases. As such, we say that β captures the *extent of preference disagreement* between the planner and the doer (the lower is β, the greater the disagreement). This preference disagreement is what the planner attempts to ameliorate through goal-setting.

Timeline: Fig. 1 summarizes the timing of the model, including the decisions that the planner and the doer make. First, the planner sets a goal $g \in [0,1]$. This sets a reference point for the doer in terms of a distribution over output consistent with the goal. The doer, who holds reference-dependent preferences over stochastic output expectations, then chooses effort as a best-response to the reference distribution of output implied by the goal.

We now describe each aspect of the model in further detail.

Reference Dependence and Loss Aversion: The doer has reference-dependent and loss-averse preferences over output. In particular, if y^R is a reference level of output and output level y realizes, the doer experiences psychological utility (in addition to material utility) of

$$\mu(y|y^R) \equiv \begin{cases} 0 & \text{if } y \geq y^R \\ \lambda(y - y^R) & \text{if } y^R > y, \end{cases} \tag{3}$$

where $\lambda \geq 0$ captures the weight the doer places on psychological losses or the extent of the doer's loss aversion.

Given the output environment is stochastic here, both true output, y, and reference output, y^R, may be drawn from a distribution. In such situations, the doer calculates her psychological utility using a *stochastic reference point*, as in Kőszegi and Rabin (2006). Specifically, if true output, y, is distributed according to G and expected output, y^R, is distributed according to H, then the doer experience psychological utility equal to $\int_y \int_{y^R} \mu(y|y^R) dH(y^R) dG(y)$.

Goals: The reference distribution of output for the doer is determined by a goal set by the planner. Formally, the planner selects a level of effort, $g \in [0,1]$, which can be interpreted as a level of effort the planner expects the doer to exert. The doer uses this expected level of effort to form her expectations over output. More precisely, given goal g, the doer's reference distribution over output, H, is such that $H(y^R) \equiv F(y^R|g)$.[10] Instead, when the doer selects an effort level e, she creates an *actual* distribution over output of $G(y) \equiv F(y|e)$.

Doer's Decision Problem: Given goal $g \in [0,1]$, the doer chooses an effort level $e \in [0,1]$ that maximizes the sum of her material and psychological utility. That is, the doer selects $e \in [0,1]$ to solve

$$\max_{e \in [0,1]} U_D(e|g) \equiv ev - \frac{C(e)}{\beta} + \int_y \int_{y^R} \mu(y|y^R) dF(y^R|g) dF(y|e) \tag{4}$$

Let $e_D^*(g)$ denote a solution to (4). We characterize different types of goals by comparing $e_D^*(g)$ to g. We say that goal g is *exceeded* if $e_D^*(g) > g$. Instead, we say that goal g is *aspirational* if $e_D^*(g) < g$ and *satisfies rational expectations* if $e_D^*(g) = g$. Finally, we say that goal g is *achieved* if it is either exceeded or satisfies rational expectations.

Planner's Decision Problem: The objective of the planner is to choose a goal that maximizes her material utility given in (1), taking into account how the doer responds to goals, as determined by the problem in (4). Formally, the optimization problem of the doer is to select g to solve

$$\max_{g \in [0,1]} e_D^*(g)v - C(e_D^*(g)). \tag{5}$$

[10] This implies that, while a goal is a level of effort (i.e. an input to the problem), the psychological utility of the doer is determined by the implication of this input for output. As such, we are taking a consequentialist perspective on how psychological utility is determined. One may also think that it is reasonable to hold reference-dependent preferences over the input of effort directly. We explore this further in Section A.3 of the Online Appendix.

In the benchmark version of the model, we assume that the planner does not account for psychological (dis)utility in their objective function. This assumption is most reasonable in situations in which the planner and the doer are separate entities (such as a manager and an employer, or a teacher and a student). It is less suitable when the planner is setting a goal for their future self. For completeness, we explore a relaxation of this assumption in Section 4.2.

3.2. Discussion of key assumptions

The model provided is purposefully simple to highlight the important aspects necessary for our result that optimal goals can be aspirational (and, hence, not achieved) in equilibrium. In this section, we motivate these assumptions more thoroughly and discuss the extent to which the model's results are robust to their relaxation.

3.2.1. No psychological gain utility

In the model we assume that the doer experiences psychological losses when output falls below their expectations but, asymmetrically, does not experience psychological gains when output realizations exceed their expectations. This is because the key ingredient for effective goal setting is that the disutility of the doer from losses outweigh the utility from equally-sized gains. Since only this asymmetry is required, we make the simplifying assumption to shut down psychological gain utility. This assumption has also been made in other contexts such as markets for advance purchases (Lin, 2023) and monopoly pricing (Hancart, 2021).

By considering psychological losses only, we focus mainly on the behavioral implications of loss aversion for the doer and less on its utility consequences. As such, one could interpret λ as a *behavioral* or *de facto* coefficient of loss aversion (Dreyfuss et al., 2022). There are some situations, however, where utilizing only a behavioral notion of loss aversion may be overly restrictive. For example, if the planner incorporates psychological utility into their objective, then whether psychological gains are considered or not has utility consequences and, thus, might affect optimal goal-setting. Moreover, psychological losses are usually measured relative to psychological gains rather than measured relative to material payoffs, which implies it lacks axiomatic foundation. In Section A.1 of the Online Appendix, however, we establish that incorporating psychological gains has no significant impact on the model's predictions regarding goal deviation in equilibrium.

3.2.2. Linear psychological utility

In the model we assume psychological losses experienced by the doer are linearly increasing in the distance between actual output and reference output. Thus, we do not allow for diminishing sensitivity as proposed by Kahneman and Tversky (1979): the idea that utility should be concave in the domain of gains and convex in the domain of losses. Diminishing sensitivity is an important aspect of reference-dependent preferences in many contexts (Wakker, 2010) and has been shown to be important for goal-setting, especially in situations in which output is a deterministic function of effort (Wu et al., 2008; Corgnet et al., 2015; Dalton et al., 2016; Brookins et al., 2017).

Our model does not require diminishing sensitivity in order to generate goal deviation in equilibrium. It is important, however, to establish that the absence of diminishing sensitivity is not the fundamental reason for this prediction. As such, we extend the model to allow for a particular parameterization of diminishing sensitivity in Section A.1 of the Online Appendix. It is shown there that incorporating diminishing sensitivity has no impact on the qualitatively important predictions of the model.

3.2.3. Reference-dependent preferences over stochastic output

The two most important assumptions of the paper for the prediction of goal deviation in equilibrium are that (1) output is a stochastic function of effort and (2) the doer has reference-dependent preferences over stochastic output (and not effort itself). If output is a deterministic function of effort, then there always exists an optimal goal that satisfies rational expectations (see Section 4.3). Similarly, if the doer holds reference-dependent preferences over effort only then, as in the case of deterministic output, they would have complete control over whether psychological losses are experienced since achieving any particular effort level is entirely under their control. Consequently, there would also always exist an optimal goal that satisfies rational expectations in this setting. Nonetheless, input or effort goals have been shown empirically to be more effective than output or performance goals and one reason for this could be that individuals also experience psychological utility over effort directly, rather than over only the consequence of effort in the form of output (Clark et al., 2020). As such, it is interesting to ask what happens if the doer holds reference-dependent preferences over both stochastic output and effort simultaneously. If this is the case optimal goals more often satisfy rational expectations but the propensity towards aspirational goal-setting is not eliminated (see Section A.3 of the Online Appendix).

We have purposefully chosen an extremely simple version of the mapping between effort and output in order to illustrate the model's predictions. The binary output model, however, may not be realistic in all contexts. As such, we extend the model in which output is drawn from a continuous distribution as a function of effort, which does not impact on the model's key findings (see Section A.5 of the Online Appendix). Moreover, as in Kőszegi and Rabin (2006), reference-dependent preferences are often defined over *all* relevant dimensions of consumption, which would include effort costs in the context of this model. The model's findings, however, are only minimally impacted by allowing the doer to hold reference-dependent preferences over effort costs in conjunction with reference-dependent preferences over output (see Section A.4 of the Online Appendix).

3.2.4. Common information between planner and doer

The benchmark version of the model is a common information environment. That is, the planner possesses the same knowledge of the environment that the doer has. We made this assumption to highlight our finding in the starkest possible environment: even

in the absence of other sources of uncertainty outside of the mapping from effort to output, we predict goal deviation in equilibrium. In many instances, however, the doer may possess an information advantage over the planner regarding aspects of the environment that both deem relevant for utility. For example, the doer may have more information about the level of resources utilized with effort exertion or the likelihood that the project will succeed with effort. This creates uncertainty in the planner's desired effort level and the doer's effort as a function of a given goal. This can lead to goal deviation if the doer's effort choice responds to their private information but the planner's optimal goal can not (Corgnet et al., 2015).

In Section A.2 of the Online Appendix, we extend the model to allow the doer to hold private information over effort costs. Doing so implies that, in some situations, the optimal goal of the planner is sometimes achieved and sometimes fallen short of, depending on the realized cost state. The main finding of our paper, however, continues to hold *on average*; that is, if one averages over all cost realizations.

3.2.5. Comparison to other theories of reference point formation

In our theory, the doer's reference distribution of output is formed by the goal set by the planner (i.e., is given by $F(y^R|g)$) while their effort choice is simply a best response to this reference distribution. Consequently, the realized distribution of output (as determined by effort) may not coincide with the reference output-distribution (which is determined by the goal). As such, our model differs from theories in which the reference distribution is formed via rational expectations such as Unacclimating Personal Equilibrium (UPE) introduced in Kőszegi and Rabin (2006) and Choice Acclimating Personal Equilibrium (CPE) introduced in Kőszegi and Rabin (2007).

If the model were to utilize the solution concept of UPE, then the planner would be restricted to set goals such that $g \in \arg\max_{e\in[0,1]} U_D(e|g)$. That is, the planner would be restricted only to goals that the doer best responded to by setting effort equal to that goal. The planner can then choose the optimal goal that constitutes a UPE, which is sometimes called their Preferred Personal Equilibrium (or PPE). This solution concept has been commonly applied in the goal-setting literature (Suvorov and Van de Ven, 2008; Hsiaw, 2013; Koch et al., 2014; Koch and Nafziger, 2016, 2020). In a CPE, instead, the reference-point of the doer acclimates to their effort choice. Consequently, the doer's optimal effort choice should solve $\max_{e\in[0,1]} U_D(e|e)$, greatly limiting the extent to which the planner can use goals to shape the doer's behavior. In both solution concepts, the reference point of the doer coincides with their actual behavior (i.e., satisfies rational expectations) and, consequently, neither can be used to model systematic goal deviation. We relax the requirement of rational expectations in order to investigate the propensity for goal deviation to occur in equilibrium.

The main aspect that our model has in common with these alternative frameworks is that we assume that the psychological utility of the doer is formulated over the *entire* distribution of output implied by the goal. Thus, the doer has a reference-point rule that includes all possible outcomes (Delquié and Cillo, 2006) or, equivalently, uses a stochastic reference point (Kőszegi and Rabin, 2006, 2007). We utilize this formulation of the reference point because there is experimental evidence for stochastic reference points (Sprenger, 2015) and because it is easily applied regardless of the stochastic map from effort to output. Prior research that utilizes stochastic reference points, however, generally also applies a rational expectations solution concept on their formation. Thus, a theoretical contribution of our paper is to utilize the stochastic reference-point formulation without simultaneously imposing the requirement of rational expectations.

4. Main results

4.1. Solving the benchmark model

We now derive the planner's optimal goal and the extent of goal deviation in this baseline version of the model. First, we derive the doer's best response to an arbitrary goal g, $e_D^*(g)$; that is, we solve the problem described in (4). The following proposition provides the solution to this problem.

Proposition 1. *Given a goal $g \in [0,1]$, the doer chooses effort $e_D^*(g) \in (0,1)$ defined implicitly by $C'(e_D^*(g)) = \beta v(1+\lambda g)$. As such, $e_D^*(g)$ is strictly increasing in g.*

Proposition 1 establishes that the doer exerts more effort the higher is the goal set by the planner. Specifically, when $g = 0$, the doer exerts their own desired effort level (i.e., $e_D^*(0) = e_D^*$) so that, in the benchmark model, it is as if the planner sets no goal when choosing $g = 0$. Then, as the planner increases goal g, the doer responds by exerting more and more effort. The intuition for this is simple. First, the higher is the goal, the higher is the doer's expectation that they will successfully complete the task and, as such, the higher is the likelihood they experience a psychological loss (when realized output is zero). As a response to this, the doer chooses higher effort to decrease the likelihood that this state occurs.

This fits with the empirical evidence provided in Locke and Latham (2002) that suggests that there is a positive association between goals and task performance. Moreover, most models of goal-setting in the literature make this prediction (for example, the papers referenced in the literature review generally draw this conclusion). We contribute to this literature by showing that this result continues to hold in our setting in which effort stochastically generates output and goals induce stochastic reference points.

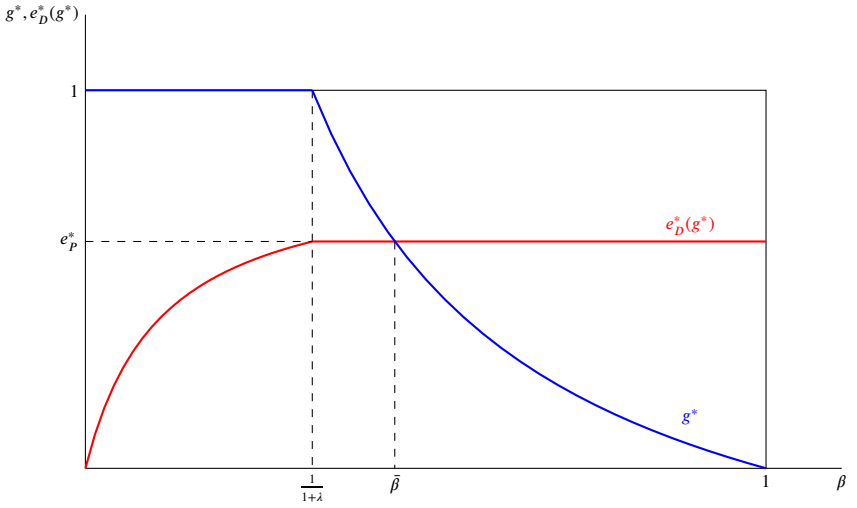

Fig. 2. Optimal goal g^* and the doer's equilibrium effort level, $e_D^*(g^*)$ as a function of β.

Note that there always exists a goal that satisfies rational expectations; that is, there exists a goal $g^{RE} \in (0,1)$ such that $g^{RE} = e_D^*(g^{RE})$.[11] This implies that the model does not preclude the planner from setting a goal that is precisely achieved by the doer. Thus, any goal deviation is the result of an active decision by the planner to deviate from the rational expectations paradigm. The next proposition provides the magnitude and direction of goal deviation when the planner selects the optimal goal, g^*.

Proposition 2. *There exists a threshold $\bar{\beta} \in (0,1)$ such that the optimal goal*

(a) is aspirational (i.e. $e_D^*(g^*) < g^*$) for $\beta < \bar{\beta}$;
(b) is exceeded (i.e. $e_D^*(g^*) > g^*$) for $\beta > \bar{\beta}$; and
(c) satisfies rational expectations (i.e. $e_D^*(g^*) = g^*$) for $\beta = \bar{\beta}$.

Proposition 2 establishes that all types of goals are optimal (aspirational, exceeded, and rational expectations), depending on the extent of preference disagreement between the planner and the doer, β. Fig. 2 illustrates both the optimal goal, g^*, and the doer's equilibrium effort, $e_D^*(g^*)$, as a function of β. When preference disagreement is large (i.e. β is sufficiently small), the planner needs to set high goals to maximally motivate the doer. These goals, however, serve only as an aspiration as the doer falls short of them. Instead, when the degree of preference disagreement is small (i.e. β is sufficiently large), the planner needs to motivate the doer to less of an extent. Hence, the planner sets low goals in order to ensure that they do not motivate the doer too strongly. In the limit when there is no preference disagreement (i.e. $\beta = 1$), the planner sets the goal $g^* = 0$ to effectively shut off the motivating effect of goals.

There is some suggestive evidence in the literature for this finding. In an educational setting, Clark et al. (2020) find that, even though individuals have full control over the achievement of input goals, there is still significant goal deviation from such goals (only 53% of participants achieve their input goal). Our model predicts such deviation from input goals if the degree of preference disagreement is sufficiently large. Moreover, they find that men are more likely to fall short of input goals (50% achievement versus 55% achievement for women). This is consistent with our model as men have been shown to display greater self-control problems in educational settings relative to women (Duckworth et al., 2015). Thus, β should be smaller for men and, consequently, my model predicts greater goal deviation for men.

There are two important implications of Proposition 2. First is related to the commonly proposed self-help method of setting S.M.A.R.T. goals.[12] Under S.M.A.R.T. goal-setting, (A)chievability is considered an important component of an effective goal. Proposition 2 shows that achievability is a suitable normative criterion only when the extent of preference disagreement between the planner and the doer is small. If this is not the case, goals *should be* aspirational and *should* systematically be fallen short of. Achievability of goals for the sake of achievability is not necessarily desirable. Rather, it depends on the context within which goals are being used.

The second point is related to the source of reference points. Kőszegi and Rabin (2006) propose a theory in which rational expectations serve as the source of reference points. Rational expectations are certainly an extremely valid source of reference points and have been utilized previously in the context of goal setting. Proposition 2 suggests, however, that if goals are a source of reference

[11] To see this, note that $e_D^*(0) > 0$ (since $C'(0) = 0$) and $e_D^*(1) < 1$ (since $\lim_{e \to +\infty} C'(e) = +\infty$). Since $e_D^*(\cdot)$ is continuous, by the intermediate value theorem there exists a goal $g^{RE} \in (0,1)$ such that $e_D^*(g^{RE}) = g^{RE}$.

[12] The full expansion of the acronym is S(mart).M(easurable).A(chievable).R(elevant).T(ime bound).

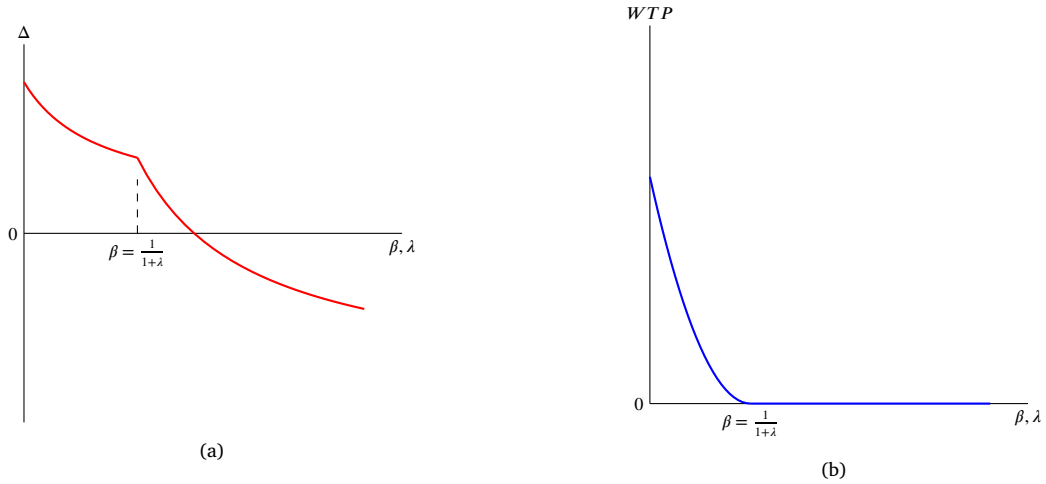

Fig. 3. Figures (a) and (b) illustrate the relationships between both β and λ and goal deviation and the willingness-to-pay for full commitment respectively. The kink in Figure (a) is due to a regime change in terms of the planner's optimal goal. Specifically, the planner moves from setting a maximal goal (i.e., $g^* = 1$) for $\beta < 1/(1+\lambda)$ to the goal $g^* = (1-\beta)/(\beta\lambda)$ which implements their desired effort level, e_P^*, for $\beta \geq 1/(1+\lambda)$.

points and are set to manage a conflict of interest between a planner and a doer, then it may be overly restrictive to simply assume they also rational expectations. Rather, a rational-expectations goal is an exception in our case, rather than the norm. We will see that the stochastic map from effort to output plays an important role in this finding in the context of our model. Indeed, in Section 4.3, we show that it is essentially without loss of generality to restrict attention to rational-expectations goals when output is a deterministic function of effort.

To this point, we have characterized when optimal goals are achievable and when, instead, they are aspirational. We now explore other comparative-static predictions of our model. Our main variables of interest for this analysis are (i) a measure of the magnitude of goal deviation and (ii) a measure of the planner's willingness to pay to impose commitment power on the doer. Our measure of the magnitude of goal deviation is given by $\Delta \equiv g^* - e_D^*(g^*)$. Instead, our measure of the planner's willingness to pay for full commitment power is given by $WTP \equiv e_P^* v - C(e_P^*) - (e_D^*(g^*)v - C(e_D^*(g^*)))$. That is, WTP is equal to the utility difference between the planner's first-best payoff and the utility achieved through optimal goal-setting. The following proposition displays how these measures vary with the model's primitives.

Proposition 3. *Suppose that the planner sets an optimal goal. Then,*

(a) goal deviation, Δ, is decreasing in β and λ.
(b) willingness-to-pay for full commitment, WTP, is

(i) equal to zero for $\beta \geq \frac{1}{1+\lambda}$; and
(ii) decreasing in β and λ for $\beta < \frac{1}{1+\lambda}$.

Part (a) of Proposition 3 states that the magnitude of goal deviation is decreasing as preferences become more aligned (i.e. β increases) or the doer becomes more loss averse (i.e. λ increases). Part (b) of the proposition states that the planner's equilibrium utility attainable with goals moves closer to its first-best level as both β and λ increase. Indeed, when β is sufficiently large, goal-setting is sufficient to achieve the planner's first-best so that no additional commitment power is valued. Fig. 3 provides a graphical illustration of these comparative statics.

Fig. 2 illustrates precisely why this result holds. When β is small (i.e., $\beta < 1/(1+\lambda)$), there is a large wedge between the level of effort desired by the planner and by the doer. Moreover, if λ is small, the planner is not able to effectively motivate the doer through the use of goals alone. In this case, the planner sets as high a goal as possible ($g^* = 1$), the doer falls short of this goal, and there is a positive value to full commitment power to the planner. As either β or λ increases locally, however, $e_D^*(g)$ increases (so that goal deviation decreases) and moves closer to the planner's desired effort level (so that WTP also decreases). When either β or λ are sufficiently large (i.e., $\beta \geq 1/(1+\lambda)$), the planner can implement their desired effort level, e_P^*, through goal-setting. Moreover, the optimal goal that achieves this is decreasing in both β (since preferences are more closely aligned) and λ (since a given goal is a more effective motivator). Hence, equilibrium goal deviation is still decreasing in this region. The planner implements their desired level of effort, however, and, consequently, the planner does not exhibit demand for imposing commitment power on the doer when the preference wedge is sufficiently small.

The main implication of Proposition 3 is that the magnitude of goal deviation should decrease the lower is the degree of preference disagreement. Clark et al. (2020) also find suggestive evidence of this. Indeed, the authors find that, on average, men exhibit a greater magnitude of goal deviation relative to women (the equivalent of 1 task for men versus 0.84 tasks for women). As was the case with

results concerning the likelihood of goal deviation from Proposition 2, this is consistent with my model since men exhibit greater self-control problems (i.e., a larger preference disagreement) in educational settings relative to women (Duckworth et al., 2015).

Proposition 3 also suggests that preference agreement and loss aversion are substitutes for each other. This is because both primitives have the same directional impact on both equilibrium goal deviation and the equilibrium willingness-to-pay for full commitment. As such, a larger preference disagreement can always be ameliorated by a higher degree of loss aversion. This implies that what we observe cannot be understood without taking into consideration both β and λ. For example, if we were to observe a doer taking an action close to that which is desired by some planner, this could either be because the extent of preference disagreement is small, or that the doer is very loss averse and, as such, goals are extremely effective motivational tools.

4.2. Psychological utility in the Planner's objective

In the baseline version of the model, the planner cares only about the material consequences of effort. Specifically, we assumed that the planner does not take psychological utility into consideration when setting an optimal goal. While this may be a reasonable assumption in situations in which the planner and the doer are separate entities, it may be less suitable when goals are self set. For completeness, we now assume that the planner incorporates psychological utility into their objective function.

Formally, we assume that the planner has the same psychological utility as the doer, $\mu(y|y^R)$, as provided in equation (3). Correspondingly, if true output, y, is distributed according to G and reference output, y^R, is distributed according to H, then the planner experiences psychological utility (in addition to material utility) of $\int_y \int_{y^R} \mu(y|y^R) dH(y^R) dG(y)$.

The main difference between the planner and the doer is how the reference distribution over output is determined. The doer uses the goal set by the planner, g, to form her expectations over output; that is, $H(y^R) = F(y^R|g)$. Instead, since the planner anticipates the effort level that the doer will take for a given goal, we assume that the planner's psychological utility is formed via rational expectations (i.e. the effort the doer will *actually* take for a given goal). As such, the planner chooses a goal to maximize

$$\max_{g\in[0,1]} e_D^*(g)v - C(e_D^*(g)) + \int_y \int_{y^R} \mu(y|y^R) dF(y^R|e_D^*(g)) dF(y|e_D^*(g)), \tag{6}$$

where $e_D^*(g)$ is the doer's best response to goal g as described in Proposition 1.[13] We assume that $C''(e) > 2\lambda v$ for all e, which ensures the planner's problem is strictly concave. The following proposition establishes that having the planner incorporate psychological utility has limited impact on the model's predictions.

Proposition 4. *If $\lambda < 1$, then, there exists a threshold $\beta^{PU} \in (0,1)$ such that:*

(a) the optimal goal is aspirational for $\beta < \beta^{PU}$;
(b) the optimal goal is exceeded for $\beta > \beta^{PU}$; and
(c) the optimal goal satisfies rational expectations for $\beta = \beta^{PU}$.

Instead, if $\lambda \geq 1$, the optimal goal is always exceeded.

Proposition 4 establishes that incorporating psychological utility into the planner's objective has no qualitative impact on the model's predictions as long as the feeling of loss aversion is not sufficiently strong. Specifically, for λ sufficiently small, optimal goal-setting is aspirational when the degree of preference disagreement is large, and is achievable when the degree of preference disagreement is small. Instead, if λ is too large, managing the disappointment that arises from not successfully completing the task becomes the dominant focus of the planner. This leads the planner to desire the doer to implement as little effort as possible, which is achieved by setting no goal (i.e. $g = 0$) in order to minimize the goal's motivational power. As such, goals are always exceeded in equilibrium if feelings of loss aversion are too strong.[14]

[13] Alternatively, the planner's problem can be written in the effort-space as

$$\max_{e\in[e_D^*(0),e_D^*(1)]} ev - C(e) + \int_y \int_{y^R} \mu(y|y^R) dF(y^R|e) dF(y|e).$$

As such, the planner evaluates psychological utility using a rational-expectations reference point and this reference-point adapts (or acclimates) to the doer's effort choice. Thus, our planner that incorporates psychological utility exhibits behavior that is consistent with a *constrained* version of Choice-Acclimating Personal Equilibrium (or CPE) as in Kőszegi and Rabin (2007). The CPE-like behavior is constrained as it is restricted to effort levels that the planner can implement through goals (i.e., the interval $[e_D^*(0), e_D^*(1)]$) rather than all feasible effort levels (i.e., the interval $[0,1]$).

[14] Assuming that $\lambda < 1$ is similar to the assumption of *no dominance of gain-loss utility*, as introduced in Herweg et al. (2010). As Kőszegi and Rabin (2007) and Masatlioglu and Raymond (2016) discuss, individuals with gain-loss utility over stochastic reference points may avoid even miniscule risks to avoid disappointment if gains are not realized, resulting in extreme violations of first-order stochastic dominance. A no dominance of gain-loss utility condition ensures that material utility is weighted more heavily than gain-loss utility in order to avoid such situations. In our case, $\lambda < 1$ ensures that the model's predictions are qualitatively the same as those found when the planner ignores psychological utility, as in the benchmark model.

4.3. Deterministic output

In the baseline version of the model we assumed that output was a stochastic function of effort. Now, instead, we suppose that output is a deterministic function of effort. Specifically, we assume that, given effort level $e \in [0,1]$, output is given by $y = ve$. All other assumptions remain the same.

Notice that formulating output in this way has no effect on the material preferences of either the planner (represented in (1)) or the doer (represented in (2)). As such, the desired effort levels of both the planner and the doer in the deterministic model coincide with those in the stochastic model (e_P^* and e_D^* respectively). The main difference is that the doer has greater control over her psychological utility when output is deterministic. Indeed, the doer's utility from choosing $e \in [0,1]$ given goal $g \in [0,1]$ is

$$U_D^{do}(e|g) \equiv \begin{cases} \beta ve - C(e)/\beta & \text{if } e \geq g \\ \beta ve - C(e)/\beta - \lambda v(g-e) & \text{if } e < g. \end{cases} \tag{7}$$

As such, the doer can avoid psychological losses in the model with deterministic output as long as she achieves the goal set by the planner. This is in contrast to the model with stochastic output, where the doer could only be sure to achieve her goal if either (i) there was no goal ($g = 0$), or (ii) she exerted effort to guarantee success ($e = 1$).

Let $e_D^{do}(g)$ denote the best response of the doer to an arbitrary goal g set by the planner; that is, $e_D^{do}(g)$ is the maximizer of (7). Then, the planner chooses a goal $g \in [0,1]$ to maximize $ve_D^{do}(g) - C(e_D^{do}(g))$. Note that the planner's objective coincides with that in the benchmark model given in (5).[15] Thus, the level of effort the planner wants the doer to exert is the same in both contexts. However, the types of goals that are utilized vary significantly across the two frameworks. The following proposition derives the optimal goal of the planner with deterministic output.

Proposition 5.

(a) The set of implementable effort levels coincide between the stochastic and deterministic models: the planner can incentivize the doer to undertake effort in $[e_D^*(0), e_D^*(1)]$ *only.*

(b) In the deterministic-output model, the optimal goal of the planner is:

(i) Any $g \in [e_D^*(1), 1]$ *which implements effort* $e_D^*(1)$ *for* $\beta < \frac{1}{1+\lambda}$*;*

(ii) $g = e_P^*$ *which implements effort* e_P^* *for* $\beta \in \left(\frac{1}{1+\lambda}, 1\right)$*; and*

(iii) Any $g \leq e_D^*(0)$ *for* $\beta = 1$.

Part (a) of Proposition 5 states that the feasible set of effort levels that the planner can induce the doer to exert is the same regardless of whether output is stochastic or deterministic. Given that the objective function of the planner is also the same in both frameworks, it follows that she will always choose a goal that induces the same level of effort across the two models. The main difference, however, is the exact goal that achieves this objective. In the stochastic-output model, this goal, generically, did not satisfy rational expectations. Instead, part (b) of Proposition 5 establishes that this optimal level of effort can *always* be implemented via use of a rational expectations goal when output is a deterministic function of effort.[16]

Overall, our model suggests that there should be less goal deviation when the mapping from effort to output is more deterministic (Proposition 5) and the prevalence of goal deviation should increase as this mapping becomes stochastic (Proposition 2). There is also empirical evidence that is consistent with this. Specifically, Brookins et al. (2017) find that there is less goal deviation for tasks that exhibit low complexity (i.e., effort is likely to translate into success, as in Proposition 5) relative to tasks that are more complex (i.e., the mapping between effort and output is much noisier as in Proposition 2).

There are two important implications of this finding. First, our deterministic-output model cannot be used to explain systematic goal deviation. This highlights the important role that stochasticity of output plays in rationalizing the tendency for individuals to deviate from their goals. Second, it supports the idea of goals being restricted to satisfy rational expectations in deterministic settings, as suggested by Kőszegi and Rabin (2006). Instead, in stochastic settings it may not be without loss of generality to restrict attention to rational-expectations goals.

[15] One difference with the benchmark model is in the case in which the planner incorporates psychological utility as in Section 4.2. In the stochastic output model, introducing psychological utility changed the planner's objective to that in (6). Instead, in the deterministic output model, the planner's objective remains maximize $ve_D^{do}(g) - C(e_D^{do}(g))$. This is because, if output is deterministic and the planner has rational expectations regarding the behavior of the doer, then their output expectations are always met and they do not experience psychological utility. This implies that the results in this section hold regardless of whether the planner incorporates psychological utility as in Section 4.2 or not.

[16] Note that there is some indeterminacy in the planner's optimal goal, which is particularly important in the case in which $\beta < 1/(1+\lambda)$. In this case any $g \geq e_D^*(1)$ is optimal and, as such, the planner might be willing to set a goal the doer falls short of. If the planner cares about the doer's experienced psychological utility as a secondary concern, however, then they would never set any $g > e_D^*(1)$, as this only negatively impacts on the doer without affecting material outcomes. Thus, it would be difficult to rationalize systematic aspirational goal setting in the deterministic output model unless one believes that the planner would seek to actively cause the doer psychological harm.

5. Conclusion

This paper has provided a model in which a planner sets goals for a doer to manage a preference disagreement over effort provision. Goals motivate the doer by forming a reference point which, due to loss aversion, incentivizes her to exert higher effort than she would otherwise.

We have shown that the optimal goal, generically, does not satisfy rational expectations. That is, the doer deviates from the goal that the planner sets in equilibrium. In particular, when the extent of preference disagreement is large, the planner optimally sets a goal that the doer systematically falls short of. Rather, only when the extent of preference disagreement is small do optimal goals satisfy the criterion of achievability. We have shown that this result is robust to a number of alternative formulations of the model. Moreover, we have established that it relies crucially on the assumption that output is a stochastic function of effort: when output is a deterministic function of effort, any optimal effort level can be implemented using a rational-expectations goal.

We have discussed a number of empirical findings in the goal-setting literature that are consistent with our model's predictions. Specifically, our prediction that there should be more goal deviation the larger the degree of preference disagreement is consistent with the findings of Clark et al. (2020) and the prediction that there should more goal deviation the noisier is the mapping from effort to output is consistent with the findings of Brookins et al. (2017). These papers, however, did not explicitly set out to investigate the determinants of goal deviation. We hope that our work will spur new empirical research to further validate the mechanism underlying goal deviation that we have proposed.

We conclude by remarking on the fact that it may not always be suitable to assume that reference points are derived from rational expectations in every economic setting. This is especially true in situations in which reference points can be manipulated, as they can here via the setting of goals. Moreover, observing goal deviation does not necessarily imply that individuals are engaged in sub-optimal goal-setting. Instead, we have shown that aspirational goals can serve as valuable motivational tools that generate more desirable outcomes than can be reached by restricting attention to those that are achievable. As such, motivating people to set realistic goals under all circumstances is not unambiguously desirable.

6. Appendix

Proof of Proposition 1. Given goal $g \in [0,1]$, the optimization problem of the doer is

$$\max_{e\in[0,1]} e\beta v - \frac{C(e)}{\beta} + \int_y \int_{y^R} \mu(y|y^R) dF(y^R|g) dF(y|e) = e\beta v - \frac{C(e)}{\beta} - (1-e)\lambda\beta v g.$$

Note that this objective is strictly concave in e (since C is strictly convex). It follows that the solution to this problem is the unique $e_D^*(g) \in (0,1)$ such that $C'(e_D^*(g)) = \beta v(1+\lambda g)$. Clearly, $e_D^*(g)$ is strictly increasing in g as C is strictly convex and $\beta v(1+\lambda g)$ strictly increases in g. □

Proof of Proposition 2. Recall that the planner's desired level of effort is implicitly defined by $C'(e_P^*) = v$. From Proposition 1, the set of effort levels that the planner can implement is given by $[e_D^*(0), e_D^*(1)]$, where $e_D^*(g)$ is implicitly defined by $C'(e_D^*(g)) = \beta v(1+\lambda g)$. We have that $e_D^*(0) = e_D^* \leq e_P^*$, with strict inequality for $\beta \in (0,1)$. Moreover, $e_D^*(1) \geq e_P^*$ if and only if

$$\beta v(1+\lambda) \geq v \quad \Leftrightarrow \quad \beta \geq \frac{1}{1+\lambda}.$$

Hence, e_P^* is implementable if $\beta \in [1/(1+\lambda), 1]$, which is achieved with goal $g^* = (1-\beta)/(\beta\lambda)$. Instead, for $\beta < 1/(1+\lambda)$, $e_D^*(g) < e_P^*$ for all $g \in [0,1]$. Given that the objective of the planner is strictly concave in e, it follows that implementing as high effort as possible is optimal in this region. This is done by choosing $g^* = 1$.

Finally, we describe goal deviation as a function of β at g^*. For $\beta < 1/(1+\lambda)$, the doer falls short of her goal as $g^* = 1 > e_P^* > e_D^*(1)$. Instead, for $\beta \in [(1/(1+\lambda), 1]$, we have that g^* decreases continuously in β (from $g^* = 1$ when $\beta = 1/(1+\lambda)$ to $g^* = 0$ when $\beta = 1$). Instead, $e_D^*(g^*) = e_P^*$ which is independent of β. Hence, there exists a unique $\bar{\beta} \in (1/(1+\lambda), 1)$ such that $\frac{1-\bar{\beta}}{\bar{\beta}\lambda} = e_P^*$. If $\beta < \bar{\beta}$, the doer falls short of the optimal goal. Instead, if $\beta \in (\bar{\beta}, 1]$, the doer exceeds the optimal goal. Finally, the optimal goal satisfies rational expectations only at the point $\beta = \bar{\beta}$. □

Proof of Proposition 3. To show part (a) of the proposition, note that equilibrium goal deviation, Δ, is given by

$$\Delta = \begin{cases} 1 - e_D^*(1) & \text{if } \beta < \frac{1}{1+\lambda} \\ \frac{1-\beta}{\beta\lambda} - e_P^* & \text{if } \beta \in \left[\frac{1}{1+\lambda}, 1\right]. \end{cases}$$

Note that $e_D^*(1)$ strictly increases in β and λ. This is because $C'(e_D^*(1)) = \beta v(1+\lambda)$ and C is strictly convex. Since Δ is continuous in β and λ, it is also strictly decreasing in β and λ.

To show part (b) of the proposition, we have that equilibrium willingness-to-pay for full commitment is given by

$$WTP = \begin{cases} ve_P^* - C(e_P^*) - \left[ve_D^*(1) - C(e_D^*(1))\right] & \text{if } \beta < \frac{1}{1+\lambda} \\ 0 & \text{if } \beta \in \left[\frac{1}{1+\lambda}, 1\right]. \end{cases}$$

Suppose that $\beta < 1/(1+\lambda)$. Then, as either β or λ increase, $e_D^*(1)$ increases towards e_P^* and, as such, WTP decreases. Instead, $WTP = 0$ for $\beta \in [1/(1+\lambda), 1]$ as the planner achieves first-best in this region. □

Proof of Proposition 4. The optimal effort level from the planner's perspective solves

$$\max_{e\in[0,1]} \pi(e) \equiv ev - C(e) - \lambda v e(1-e).$$

The first derivative of this objective function is given by $\pi'(e) = v - C'(e) - \lambda v(1-2e)$, which is strictly decreasing by the assumption that $C''(e) > 2\lambda v$ for all $e \in [0,1]$. If $\lambda \geq 1$, then $\pi'(0) \leq 0$ and so the planner desires that the doer exerts zero effort. As such, the optimal goal is that which induces the doer to take as little effort as possible, which is $g = 0$. Since $e_D^*(0) > 0$, it follows that the optimal goal is always exceeded in this case.

Instead, if $\lambda < 1$, the planner's desired effort level is the unique $e_P^* \in (0,1)$ such that $\pi'(e_P^*) = 0$ or $C'(e_P^*) = v\left[1 - \lambda(1-2e_P^*)\right]$. In order to derive the optimal goal when $\lambda < 1$, there are at most three cases to consider:

Case 1 - $\beta < \frac{1-\lambda+2\lambda e_P^*}{1+\lambda}$**:** In this case, the planner cannot induce the doer to exert sufficient effort through the use of goals as

$$e_P^* > e_D^*(1) \Leftrightarrow v[1-\lambda(1-2e_P^*)] > \beta v[1+\lambda] \Leftrightarrow \frac{1-\lambda+2e_P^*}{1+\lambda} > \beta.$$

As such, the optimal goal is $g^* = 1$ (to maximize the doer's effort provision), which the doer always falls short of as $e_D^*(1) < 1$.

Case 2 - $\beta \in \left[\frac{1-\lambda+2\lambda e_P^*}{1+\lambda}, 1-\lambda+2\lambda e_P^*\right]$**:** In this case, there exists a goal that induces the doer to exert the planner's desired effort level, e_P^*. Specifically, if the planner sets the goal $g^* = (1-\beta-\lambda(1-2e_P^*))/(\beta\lambda)$, then the doer chooses effort e_P^* as

$$C'(e_D^*(g^*)) = \beta v\left[1 + \lambda\frac{1-\beta-\lambda(1-2e_P^*)}{\beta\lambda}\right] = v[1-\lambda(1-2e_P^*)] = C'(e_P^*).$$

Note that $g^* \in [0,1]$ if and only if $\beta \in \left[\frac{1-\lambda+2\lambda e_P^*}{1+\lambda}, 1-\lambda+2\lambda e_P^*\right]$. Moreover, this case is exhaustive if $1-\lambda+2\lambda e_P^* \geq 1$ or $e_P^* \geq 1/2$.

Case 3 - $\beta > 1-\lambda+2\lambda e_P^*$**:** Note that this case is relevant only if $e_P^* < 1/2$. Here, the amount of effort that the doer exerts strictly exceeds the planner's desired effort level, regardless of the goal. This is because

$$e_D^*(0) > e_P^* \Leftrightarrow \beta v > v[1-\lambda(1-2e_P^*)] \Leftrightarrow \beta > 1-\lambda+2\lambda e_P^*.$$

As such, the planner optimally induces the doer to implement as little effort as possible by choosing the goal $g^* = 0$. In this case, the optimal goal is always exceeded as $e_D^*(0) > 0$.

We now show that there exists a threshold $\beta^{PU} \in (0,1)$ such that the optimal goal is aspirational for $\beta < \beta^{PU}$ and is achieved for $\beta \geq \beta^{PU}$. Note that, in Case 1 the optimal goal is aspirational and in Case 3 the optimal goal is exceeded. In Case 2, the doer chooses the planner's desired effort level, e_P^*, which is independent of β. As such, the optimal goal is aspirational if and only if

$$g^* = \frac{1-\beta-\lambda(1-2e_P^*)}{\beta\lambda} > e_P^* \Leftrightarrow \beta < \frac{1-\lambda+2\lambda e_P^*}{1+\lambda e_P^*} \equiv \beta^{PU},$$

where $\beta^{PU} \in \left(\frac{1-\lambda+2\lambda e_P^*}{1+\lambda}, 1-\lambda+2\lambda e_P^*\right)$ since $e_P^* \in (0,1)$. It follows that for $\beta > \beta^{PU}$ the optimal goal is exceeded, while for $\beta = \beta^{PU}$, the optimal goal satisfies rational expectations. □

Proof of Proposition 5. Suppose that the planner selects an arbitrary goal $g \in [0,1]$. First, consider the case where the doer chooses effort $e \geq g$. Then, she receives utility $ve - C(e)/\beta$, which has unconstrained maximizer $\underline{e}$ that is implicitly defined by $C'(\underline{e}) = \beta v$. Hence, the doer's optimal choice of $e \geq g$ is given by $\max\{\underline{e}, g\}$.

Instead, suppose that the doer chooses effort level $e \leq g$. Then, she receives utility $ve - C(e)/\beta - \lambda v(g-e)$, which has unconstrained maximizer $\bar{e}$ that is implicitly defined by $C'(\bar{e}) = \beta v(1+\lambda)$. Hence, the doer's optimal choice of $e \leq g$ is given by $\min\{\bar{e}, g\}$.

Since $\bar{e}(g) > \underline{e}(g)$ for $\lambda > 0$, it follows that the best response of the doer to an arbitrary goal $g \in [0,1]$ is

$$e_D'(g) \equiv \begin{cases} \bar{e} & \text{if } g > \bar{e} \\ g & \text{if } g \in [\underline{e}, \bar{e}] \\ \underline{e} & \text{if } g < \underline{e}. \end{cases} \tag{8}$$

Using (8), it follows that the set of implementable efforts is given by $[\underline{e}, \bar{e}]$. In comparison, the set of implementable efforts in the stochastic model is given by $[e_D^*(0), e_D^*(1)]$, where $C'(e_D^*(g)) = \beta v(1+\lambda g)$. As such, $e_D^*(0) = \underline{e}$ and $e_D^*(1) = \bar{e}$, so that the set of implementable efforts in both the stochastic and deterministic models coincide. This proves part (a) of the proposition.

To show part (b), note from the proof of Proposition 2, $e_P^* \in [e_D^*(0), e_D^*(1)]$ if and only if $\beta \geq 1/(1+\lambda)$. Instead, for $\beta < 1/(1+\lambda)$, $e_D^*(1) < e_P^*$ and the optimal goal is any goal that implements maximal effort; that is, any $g \in [e_D^*(1), 1]$. Instead, for $\beta \in \left(\frac{1}{1+\lambda}, 1\right)$, e_P^* is implementable via only the rational-expectations goal $g = e_P^*$. Finally, for $\beta = 1$, $e_D^*(0) = e_P^*$ and, consequently, any goal that implements $e_D^*(0)$ is optimal; that is any $g \leq e_D^*(0)$. □

Declaration of competing interest

None.

Data availability

No data was used for the research described in the article.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jebo.2024.01.003.

References

Abeler, J., Falk, A., Goette, L., Huffman, D., 2011. Reference points and effort provision. Am. Econ. Rev. 101 (2), 470–492.
Ali, S.N., 2011. Learning self-control. Q. J. Econ. 126 (2), 857–893.
Allen, E.J., Dechow, P.M., Pope, D.G., Wu, G., 2016. Reference-dependent preferences: evidence from marathon runners. Manag. Sci.
Ashraf, N., Karlan, D., Yin, W., 2006. Tying Odysseus to the mast: evidence from a commitment savings product in the Philippines. Q. J. Econ., 635–672.
Brookins, P., Goerg, S., Kube, S., 2017. Self-chosen goals, incentives, and effort. Unpublished manuscript.
Carrillo, J.D., Dewatripont, M., 2008. Promises, promises? Econ. J. 118 (531), 1453–1473.
Clark, D., Gill, D., Prowse, V., Rush, M., 2020. Using goals to motivate college students: theory and evidence from field experiments. Rev. Econ. Stat. 102 (4), 648–663.
Corgnet, B., Gómez-Miñambres, J., Hernán-Gonzalez, R., 2015. Goal setting and monetary incentives: when large stakes are not enough. Manag. Sci. 61 (12), 2926–2944.
Dalton, P.S., Ghosal, S., Mani, A., 2016. Poverty and aspirations failure. Econ. J. 126 (590), 165–188.
Della Vigna, S., Malmendier, U., 2006. Paying not to go to the gym. Am. Econ. Rev. 96 (3), 694–719.
DellaVigna, S., Malmendier, U., 2004. Contract design and self-control: theory and evidence. Q. J. Econ. 119 (2), 353–402.
Delquié, P., Cillo, A., 2006. Disappointment without prior expectation: a unifying perspective on decision under risk. J. Risk Uncertain. 33, 197–215.
Dreyfuss, B., Heffetz, O., Rabin, M., 2022. Expectations-based loss aversion may help explain seemingly dominated choices in strategy-proof mechanisms. Am. Econ. J. Microecon. 14 (4), 515–555.
Duckworth, A.L., Shulman, E.P., Mastronarde, A.J., Patrick, S.D., Zhang, J., Druckman, J., 2015. Will not want: self-control rather than motivation explains the female advantage in report card grades. Learn. Individ. Differ. 39, 13–23.
Hancart, N., 2021. Managing the expectations of buyers with reference-dependent preferences.
Harding, M., Hsiaw, A., 2014. Goal setting and energy conservation. J. Econ. Behav. Organ. 107, 209–227.
Heath, C., Larrick, R.P., Wu, G., 1999. Goals as reference points. Cogn. Psychol. 38 (1), 79–109.
Herweg, F., Müller, D., Weinschenk, P., 2010. Binary payment schemes: moral hazard and loss aversion. Am. Econ. Rev. 100 (5), 2451–2477.
Hsiaw, A., 2013. Goal-setting and self-control. J. Econ. Theory 148 (2), 601–626.
Hsiaw, A., 2018. Goal bracketing and self-control. Games Econ. Behav. 111, 100–121.
Jain, S., 2009. Self-control and optimal goals: a theoretical analysis. Mark. Sci. 28 (6), 1027–1045.
Kahneman, D., Tversky, A., 1979. Prospect theory: an analysis of decision under risk. Econometrica 47 (2), 263–292.
Kaiser, J.P., Koch, A.K., Nafziger, J., 2021. Self-set goals are effective self-regulation tools–despite goal revision. CEPR Discussion Paper No. DP15716. Available at SSRN: https://ssrn.com/abstract=3783942.
Karle, H., Kirchsteiger, G., Peitz, M., 2015. Loss aversion and consumption choice: theory and experimental evidence. Am. Econ. J. Microecon. 7 (2), 101–120.
Kaur, S., Kremer, M., Mullainathan, S., 2015. Self-control at work. J. Polit. Econ. 123 (6), 1227–1277.
Koch, A.K., Nafziger, J., 2011. Self-regulation through goal setting. Scand. J. Econ. 113 (1), 212–227.
Koch, A.K., Nafziger, J., 2016. Goals and bracketing under mental accounting. J. Econ. Theory 162, 305–351.
Koch, A.K., Nafziger, J., 2020. Motivational goal bracketing: an experiment. J. Econ. Theory 185, 104949.
Koch, A.K., Nafziger, J., 2021. Motivational goal bracketing with non-rational goals. J. Behav. Exp. Econ. 94, 101740.
Koch, A.K., Nafziger, J., Suvorov, A., van de Ven, J., 2014. Self-rewards and personal motivation. Eur. Econ. Rev. 68, 151–167.
Kőszegi, B., Rabin, M., 2006. A model of reference-dependent preferences. Q. J. Econ., 1133–1165.
Kőszegi, B., Rabin, M., 2007. Reference-dependent risk attitudes. Am. Econ. Rev. 97 (4), 1047–1073.
Laibson, D., 1997. Golden eggs and hyperbolic discounting. Q. J. Econ., 443–477.
Lin, S., 2023. Buy it now, or later, or not: Loss aversion in advance purchasing.
Locke, E.A., Latham, G.P., 2002. Building a practically useful theory of goal setting and task motivation: a 35-year odyssey. Am. Psychol. 57 (9), 705.
Markle, A., Wu, G., White, R.J., Sackett, A.M., 2015. Goals as reference points in marathon running: a novel test of reference dependence. Research Paper, 2523510. Fordham University Schools of Business.
Masatlioglu, Y., Raymond, C., 2016. A behavioral analysis of stochastic reference dependence. Am. Econ. Rev. 106 (9), 2760–2782.
Phelps, E.S., Pollak, R.A., 1968. On second-best national saving and game-equilibrium growth. Rev. Econ. Stud. 35 (2), 185–199.
Sprenger, C., 2015. An endowment effect for risk: experimental tests of stochastic reference points. J. Polit. Econ. 123 (6), 1456–1499.
Suvorov, A., Van de Ven, J., 2008. Goal setting as a self-regulation mechanism. SSRN Working Paper 1286029.
Toussaert, S., 2018. Eliciting temptation and self-control through menu choices: a lab experiment. Econometrica 86 (3), 859–889.
Wakker, P.P., 2010. Prospect Theory: For Risk and Ambiguity. Cambridge University Press.
Wu, G., Heath, C., Larrick, R., 2008. A prospect theory model of goal behavior. Unpublished manuscript.