NBER WORKING PAPER SERIES

PROCRASTINATION IN TEAMS

Joshua S. Gans
Peter Landry

Working Paper 21891
http://www.nber.org/papers/w21891

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2016

Thanks to seminar participants at the University of Toronto for helpful comments. Responsibility for all errors remains our own. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed a financial relationship of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w21891.ack

Procrastination in Teams
Joshua S. Gans and Peter Landry
NBER Working Paper No. 21891
January 2016
JEL No. C72,D03,M11

ABSTRACT

Naively present-biased agents are known to be severe procrastinators. In team settings, procrastination can represent a form of free-riding that, in excess, can jeopardize a team's ability to meet a deadline. Here we show how naivete and present bias, despite their reputations, can be desirable traits in a teammate, enabling a team to optimize its performance while eliminating inefficient free-riding. These benefits emerge only from a more flexible specification (in comparison to existing models) as to how naive players reassess prior beliefs upon confronting present bias. By allowing the 'depth' and 'direction' of such reassessments to vary, our model links present-biased discounting theories to the recently-revived interest in modeling non-Bayesian reactions to null events, while offering a distinct approach reminiscent of level-k reasoning. Key themes from our results include the value of behavioral diversity, the opposite effects of 'introspection' and 'extrospection' on motivation, and that under- and over-thinking can both undermine efficiency.

Joshua S. Gans
Rotman School of Management
University of Toronto
105 St. George Street
Toronto ON M5S 3E6
and NBER
joshua.gans@gmail.com

Peter Landry
University of Toronto, Dept of Management
Mississauga ON
Canada
Peter.Landry@rotman.utoronto.ca

1 Introduction

Just because a project is known to be worth the effort in the long-run does not guarantee it will be completed in a timely fashion. According to standard economic treatments, *team* projects are particularly prone to inefficiencies of this sort. Holmstrom (1982) showed it is largely impossible to use the value created by a team to motivate its members due to moral hazard concerns.[1] A literature on voluntary public goods contributions has demonstrated similar difficulties in overcoming free-riding (Bagnoli and Lipman, 1989).[2] In many cases, these incentive issues can be further compounded by the possibility of coordination failure, as it can lead to inefficient under-provision (or even over-provision) of effort on team projects (Wittenbaum et al., 1998; Brandts and Cooper, 2006).

In a separate line of research, behavioral economic theories have shown that even worthwhile *individual* projects can be derailed, as a consequence of *present bias*. Present-biased agents discount future utility at t using the time-*in*consistent quasi-hyperbolic discount function, $\beta\delta^t$, with "present bias factor" $\beta < 1$ (Laibson, 1997; O'Donoghue, 1999, 2001). By reducing the effective weight of future payoffs in decision-making (relative to a time-consistent evaluation), present bias can, in and of itself, tempt an agent to put off investing effort on a worthy project. However, this problem can be severely exacerbated by *naive* present bias — a situation in which the agent is oblivious to their present bias until that present arrives. Compared to their *sophisticated* counterparts (who correctly anticipate their own time-inconsistency), naively present-biased agents can better justify putting off a task today based on a false expectation that they will be motivated to do it tomorrow.[3]

Given the many obstacles to efficient team production already known to exist with standard time-consistent agents, it would be natural to expect even worse outcomes if present bias — especially with naivete — was brought into a team framework. To look into this further, we develop a model of team collaboration with present-biased agents who, building on familiar notions of naivete and sophistication from single-agent models, may be naive or sophisticated towards their partners' present bias

[1] For a recent treatment, see Bonatti and Horner (2011).

[2] Of note, this literature has shown that designing outcomes to ensure a particular agent is pivotal (that is, has default control over the outcome) can generate incentives that mitigate inefficiency due to free-riding.

[3] While Akerlof (1991) pointed to the role of naivete in a more general sense, the link between procrastination and naivete as it pertains to present bias was formally worked out by O'Donoghue and Rabin (1999). O'Donoghue and Rabin (2001, 2008) later showed (respectively) that the vulnerability to severe procrastination extends to present-biased agents who exhibit a form of partial naivete and that (full) naivete can lead to endless procrastination on longer-term projects.

in addition to their own. To fully translate naive present bias into a team setting, however, a trickier modeling question arises: how do such agents update their other beliefs (including higher-order beliefs) when they inevitably discover their own present bias?

To understand this technical challenge in our setting, first consider a familiar situation in which two co-authors — Alice and Bob — are collaborating on a critical and time-sensitive research project. At present, Alice and Bob are each deciding whether or not to exert effort on a task that will bring the project closer to completion. Since effort is costly, both co-authors have an incentive to minimize their contribution to the project — provided it still gets done — creating a potential for free-riding. For instance, if Bob chooses to put off doing a task, Alice will be forced to undertake additional tasks to ensure the project is completed on time. Thus, there can be a strategic advantage to procrastination in that it can allow an agent to economize on their task share, but procrastination by both co-authors can prevent the team from being able to meet its deadline.

Next, suppose Alice is fully naive in that, prior to facing her present decision, she believed that both she and Bob would be time-consistent (and that this was common knowledge). Now that the decision has arrived, however, Alice discovers that she is in fact present-biased. While re-computing her optimal strategy in light of this unexpected discovery, Alice may revisit her other prior assumptions. For instance, she may ask herself, *if I am present-biased, does this mean Bob is too?* Alice may also wonder, *does Bob know that I am present-biased?* If Alice concludes that the answer to either (or both) of these questions is 'yes,' we say that she has *reassessed* (i.e. reversed) her associated prior belief — and all of Alice's infinite higher-order prior beliefs are also subject to reassessment in this sense. To complicate matters significantly, Alice cannot turn to Bayes' Rule to resolve such questions. That is, Alice was *certain* that she would not be present-biased, but Bayes' Rule does not say anything about how beliefs should be updated in response to 'null' events such as this, to which the agent previously assigned zero-probability.

To address these issues, we formalize a tractable yet flexible "sequential reassessment" rule in which naive agents, upon confronting their own present bias, reassess their prior beliefs about each players' time-(in)consistency up to — but not beyond — a given *depth* (i.e. order) between zero and infinity. Alice's posterior, decision-relevant beliefs can therefore be summarized by two numbers: the depth of her 'inward' reassessment regarding her own present bias and the depth of her 'outward' reassessment concerning Bob's present bias. In turn, these numbers determine the overall *direction* of Alice's reassessment. That is, Alice is either considered more 'introspective' or 'extrospective,' on balance, depending on whether her inward or outward reassessment is deeper.[4]

A main implication of our model is that, contrary to what one might expect, introducing naively present-biased players into a team setting can allow a team to optimize its performance while overcoming inefficient free-riding (and related coordination issues) that would otherwise exist. In establishing this and other results, the model demonstrates the importance of its unique behavioral elements arising from our novel sequential reassessment approach to confronting naive present bias, as the depths and directions of such reassessments prove to be critical determinants of both individual behavior and team efficiency. For example, we find that 'extrospection' (in the sense described above) will motivate a naively present-biased agent to overcome their infamous tendency to procrastinate, while 'introspection' has the opposite effect. Moreover, a team with two naively present-biased agents will achieve full efficiency in our setting if and only if players are diverse in this regard. This prediction speaks to a recurring theme in our model: that the ability of a team to work together often depends on its precise composition. The result also offers a contribution to the theoretical literature on the value of diversity, as it highlights a novel channel — variation in the directions of naive agents' reassessments — that can improve team functionality in the absence of more familiar channels, such as diversity in skillsets or technical knowledge.[5]

There have been a few models to date that have considered what happens when (possibly naive) present-biased agents interact. However, these models make non-obvious (in our view) simplifying assumptions that sidestep thorny questions about agents' higher-order beliefs, and in doing so, preclude the forms of behavioral variation that prove important in our model.[6] In the language of our earlier example, these

[4]Yes, this is actually a word. See http://dictionary.reference.com/browse/extrospective.

[5]The composition of well-functioning teams has been at the heart of economics since Adam Smith's 'division of labor.' Based on productive efficiency alone, there are gains from having a diverse set of skills among team members. (See, for example, Becker and Murphy, 1992). Indeed, much of the research outside of economics on diversity and team functionality has been based on the theoretical proposition that diversity can provide a team with different information and insights, as well as the appropriate compositions of skills for various tasks that need to be accomplished. (See for instance, the comprehensive review of that literature by Mannix and Neale, 2005).

[6]In an unpublished paper, Sarafidis (2006) examines the interaction of naive and sophisticated agents and highlights when predictions are likely to differ from games with time-consistent agents. Akin (2009) develops this further, exploring when such agents may generate inefficient outcomes in alternating-offer bargaining games. Finally, Haan and Hauck (2014) highlight the need to consider present-biased agents' assumptions about the time-consistency of agents they interact with. Although they focus on sophisticated

models effectively assume that Alice would reassess *all* of her prior beliefs, concluding that both she and Bob are time-consistent (and that this is common knowledge) — which is precisely the opposite of what she believed prior to confronting her present bias.[7] In our model, however, we take seriously the notion that naive agents may revise their prior beliefs in other ways.

This paper proceeds as follows. In the next section, we introduce a simple, two person, multiple task environment. Our goal here is to keep the underlying environment simple so as to place our analytical attention on the belief structure (on this dimension, we aim for generality). Here we also describe our notion of perception-perfect equilibrium in the underlying dynamic game that itself is a natural extension of perception-perfect strategies considered by O'Donoghue and Rabin (1999, 2001).[8]

Section 3 characterizes the equilibrium strategies and outcomes that arise in teams with fully naive agents who do not anticipate either their own or their partner's present bias. It is there that we introduce our key concept — sequential reassessment — while showing how the team can achieve efficiency if its members vary in the direction of their reassessments. Section 4 then incorporates various forms of sophistication into the framework. In particular, 'self-sophisticated' agents hold correct beliefs (including higher-order beliefs) about their own present bias and 'other-sophisticated' agents hold correct beliefs about their partner's present bias. Sophistication can both help and hurt the performance of a team depending on the precise nature of players' sophistication as well as the composition of the team. While it is difficult to succinctly summarize these relationships, the key to assembling an efficient team is to find agents who agree on their 'differences' at some level even if their exact beliefs are incompatible — as was the case in Section 3.

As discussed at greater length in Section 5, our model provides a bridge between established theories of present-biased discounting and the recently-revived decision-theoretic interest in non-Bayesian updating in response to null events. Besides our new application to present bias, sequential reassessment represents a distinct approach

agents, Brocas and Carrillo (2001) provide an earlier model of present bias in the context of both competitive and cooperative projects, showing how competition can alleviate inefficiency stemming from a tendency to procrastinate and also from a tendency to rush.

[7] Technically, these papers generally achieve this by assuming that, when the present arrives, naive agents' present biases are publicly observable or otherwise common knowledge. Why this would be so is not discussed. Unlike conventional assumptions of common knowledge regarding players' preference parameters, common knowledge in this context entails a sudden reversal of beliefs, and hence learning, which requires an observable signal of some sort (but it is not clear what this signal would look like or where it comes from). Nonetheless, this has the effect of making it common knowledge that agents are present-biased and so there are no questions regarding higher-order beliefs.

[8] See also, Haan and Hauck (2014).

to this problem, which we compare and contrast to the treatments of Ortoleva (2012) and Karni and Viero (2013). As formalized in our model, sequential reassessment could also be cast as an extension of level-k reasoning models into the domains of present bias and non-Bayesian updating in response to null events. Level-k models posit that players' beliefs (and strategies) are anchored to some naive prior held by a "level-0" player such that a level-1 player best-responds to the strategy of a level-0 player, a level-2 player best-responds to a level-1 player, and so on.[9] Under (finite) sequential reassessment, naively present-biased agents similarly believe they are one step ahead of their partner and best-respond accordingly. Unlike standard applications of level-k reasoning, however, we consider players' 'thought processes' in two opposing dimensions — 'introspection' and 'extrospection,' as we have labeled them — while demonstrating the behavioral relevance of this distinction.

Although we believe naive present bias to be the most natural and well-grounded motivation, sequential reassessment could in principle be applied by players who confront false preconceptions regarding any utility parameter. Building on this observation, in Section 6 we show how our framework can be collapsed to a strategically-equivalent static game with the payoff structure of a two-player volunteer's dilemma. We then entertain one potential interpretation of "payoff naivete," as it arises in this context, that relates to biases in the maintenance of a positive self-image. By viewing the model through this lens, we can relate our approach to established psychological concepts and also to some economic theories of overconfident (or otherwise overoptimistic) players in team-like settings.

2 Model

We study a dynamic game of collaboration on a multi-task project. To focus on the simplest nontrivial case of the general game we have in mind, we model a two-player team that has two periods to complete a project that consists of three identical tasks. In each period, $t = 1, 2$, players simultaneously choose whether or not to complete a single task. Any such task requires an effort cost $c > 0$ and can be completed by either player, although each player can only complete one task per period, while

[9] See the models of Nagel (1995), Stahl and Wilson (1995), Costa-Gomes et al. (2001), Costa-Gomes and Crawford (2006), and the closely-related 'Cognitive Hierarchy' model of Camerer et al. (2004). Real-life players in normal form games do appear to exhibit a limited depth of reasoning in the manner suggested by these models, as many empirical and experimental studies show that level-k reasoning often out-predicts models based on standard equilibrium concepts (for a review, see Crawford et al., 2013).

each player receives a payoff equal to 1 if and when the project is completed.[10] It is important to note that it is always possible to complete a task — even in excess of the number required for the project. Thus, if both players choose to exert effort in both periods, there is an inefficient overprovision of effort as four tasks are completed for a project that only requires three (certainly, there can be underprovision too, as would be the case if the team collectively completes less than three tasks by the end of the final period).

We assume that both players have time-inconsistent preferences whereby future utility is discounted by a common present-bias factor $\beta < 1$.[11] For simplicity, future utility is not discounted other than through this present bias.[12] We want to consider projects that are worth the effort to complete, even from the present-biased perspective of a player who ultimately exerts effort in both periods. Therefore, we assume $c < \dfrac{\beta}{1+\beta}$, which ensures that the player's discounted payoff (β) exceeds the discounted sum of the effort costs ($c + \beta c$) in this full-effort scenario.[13]

A distinction that is commonly made — and has proven to be behaviorally important — in single-agent models with present-biased discounting is whether the individual is a "sophisticate" who correctly believes that they will be present-biased in future periods, or a "naif" who is oblivious of their own future present bias until that future "present" arrives. Translating these standard notions of naivete and sophistication into our multi-agent setting (where each player's beliefs and higher-order beliefs regarding *both* players' present biases need to be specified) represents a whole and messier-than-expected can of worms. In subsequent sections, we will do our best to slowly and carefully peel this can open, and in doing so, we will characterize be-

[10]By modeling a three-task project in this timeframe (and bearing in mind the implied labor capacity constraint), we can consider inefficiencies due to free-riding and coordination failure while preserving the notion that a team can actually create value. For a four-task project, there would be no incentive to procrastinate. With two tasks, a team would serve no purpose since any one player would be able to complete the task without the help of a teammate (and, in doing so, the threats to efficiency that exist in team settings would be negated).

[11]As we will see, solving our model with two present-biased players will require us to first characterize the equilibria for the cases in which one or both players are time-consistent (i.e. $\beta = 1$). Therefore, even with the assumption that both players are present-biased, we will be able to derive predictions involving (hypothetical) time-consistent players along the way.

[12]Since we are only considering a two-period model, the only relevant discount is the weight of β on utility at $t = 2$, from the perspective at $t = 1$. In and of itself, this present bias would have the same effect if it was conceived as a standard, exponential discount factor. While we could, in principle, consider naive beliefs with respect to a standard discount factor, we do not adopt this interpretation because it is only conventional to consider such naivete with respect to present-biased discounting.

[13]Alternatively, we can express this condition as $\beta > \frac{c}{1-c}$, and interpret it as an assumption that the present bias is mild enough to ensure that a time-inconsistent player prefers to have the project completed even if it requires effort in both periods.

havior over a large range of time-inconsistent "types." As will later become clear, we will be able to formally express all of this diversity as variation in players' beliefs (importantly including higher-order beliefs) at the time choices are made, regarding both player's present biases. While we will put more structure on these beliefs later (informed by established notions of naivete and sophistication), for now we formalize them as follows.

2.1 General Formulation of Beliefs

For a given player, let $\mu = (\mu_O; \mu_I)$ denote their full set of beliefs regarding the present biases (or lack thereof) of both players. In particular, $\mu_O = (\mu_O(1), \mu_O(2), \dots)$ denotes *outward* beliefs pertaining to the other player's present bias factor and $\mu_I = (\mu_I(1), \mu_I(2), \dots)$ denotes *inward* beliefs pertaining to one's own present bias factor, where the argument n in $\mu_O(n)$ or in $\mu_I(n)$ connotes the degree of the associated belief. Formally, these first-degree beliefs are defined as:

$$\begin{aligned} \mu_O(1) = B \quad &\Leftrightarrow \quad \text{I believe my partner's present-bias factor is } B, \\ \mu_I(1) = B \quad &\Leftrightarrow \quad \text{I believe my partner believes my present-bias factor is } B, \end{aligned} \tag{1}$$

where, for reasons that will soon be evident, we will restrict $B \in \{\beta, 1\}$. For each $z \in \{O, I\}$, higher-degree beliefs are then defined recursively by:

$$\mu_z(n+1) = B \quad \Leftrightarrow \quad \text{I believe my partner believes } \mu_z(n) = B, \quad n = 1, 2, \dots$$

Thus, if $\mu_O(1) = \beta$ but $\mu_O(2) = 1$, for example, the player whose beliefs are given by μ believes that their partner is present-biased, but also believes that their partner doesn't know they knows this.[14]

Finally, we can also let $\mu_I(0) = \beta$ denote the player's belief regarding their own present bias where, unlike other beliefs defined above, it is necessarily correct because a player must be aware (even if it is a brand new discovery) of their own present bias at the time of their choice. That said, a player may believe their partner is time-consistent (i.e. if $\mu_O(1) = 1$). Thus, although we assume players are present-biased, we will need to characterize the equilibrium strategies of hypothetical time-consistent players for whom $\mu_I(0) = 1$. Note, in either case, $\mu_I(0)$ will be interpreted as the (actual or hypothetical) player's true present-bias factor and is therefore excluded

[14]Translating high-enough order beliefs into easy-to-grasp (and keep track of) language can be hard, so we will keep our illustrative examples as simple as possible.

from the inward-belief vector, μ_I.

For the player with beliefs μ, let $\tilde{\mu}' = (\tilde{\mu}'_O, \tilde{\mu}'_I)$ denote their beliefs regarding their partner's beliefs, and let $\mu' = (\mu'_O, \mu'_I)$ denote their partner's true beliefs. The (original) player's beliefs regarding their partner's beliefs, $\tilde{\mu}'$, are uniquely determined from μ using:

$$\tilde{\mu}'_O(n) = \mu_I(n), \quad \text{and} \quad \tilde{\mu}'_I(n) = \mu_O(n+1), \qquad \text{for all } \ n = 1, 2, \ldots \tag{2}$$

To see why $\tilde{\mu}'_O(n) = \mu_I(n)$ holds for the simplest case with $n = 1$, first recall if Alice is the player whose beliefs are given by μ and Bob is her partner, then from equation (1) we see that $\mu_I(1) = B$ means that Alice believes that Bob believes that her present bias factor is B. Now $\tilde{\mu}'_O(1) = B$ means that Alice believes $\mu'_O(1) = B$. Since $\mu'_O(1)$ means that Bob believes that Alice's present bias factor is B, it follows that if Alice believes this, i.e. if $\tilde{\mu}'_O(1) = B$, this means exactly the same thing as $\mu_I(1) = B$. The validity of the second expression in equation (2), $\tilde{\mu}'_I(n) = \mu_O(n+1)$, can be verified in a similar fashion. Lastly note, we can also define $\tilde{\mu}'_I(0) = \mu_O(1)$, which represents (in this case) Alice's belief regarding Bob's present-bias factor.

To simplify our analysis while isolating the effect of μ on behavior, we rule out all possible bases for heterogeneity in behavior besides those attributable to differences in μ. Put differently, two players with the same beliefs μ and in the same (sub)game will employ the same strategy.[15] Moreover, since players' present biases and associated beliefs are irrelevant at $t = 2$ — recall, both the effort cost and potential payoff are realized immediately — we presume all players have identical strategies at $t = 2$, which will allow us to focus predominantly on how beliefs matter for behavior at $t = 1$.[16]

2.2 Equilibrium Behavior

We assume players' strategies are *perception-perfect* in that a player chooses the optimal action given their current preferences, their perceptions of what their partner's current action will be, and their perceptions of them and their partner's future ac-

[15]As seen in Section 3 and 4, we will allow players to differ with respect to their statuses as naive or sophisticated (including forms of partial naivete/sophistication), but individual variation stemming from these differences will be entirely captured by variation between μ and μ'. That said, when expanding our analysis to include hypothetical time-consistent players, whether or not a player is present-biased will surely also be a permissible basis for differences in behavior.

[16]The simplifications also allow us to abstract from potentially thorny questions related to learning that are outside the scope of this paper. For instance, if a player's beliefs at $t = 1$ are contradicted by their partner's chosen action, it is not obvious how or whether such beliefs at $t = 2$ might change.

tions.[17] A player's "perceptions" in this sense, while not necessarily correct, will just be a straightforward extrapolation of what would happen given their beliefs μ (which are also not necessarily correct).

In any subgame at $t = 2$ (where a subgame at $t = 2$ is uniquely defined by the number of tasks completed in $t = 1$), this perception-perfect solution concept effectively reduces to static Nash equilibrium since μ and μ' only matter at $t = 1$ (recall, all possible costs and payoffs associated with actions at $t = 2$ are experienced immediately and thus how players discount the future is of no consequence at this stage of the game). We can therefore fully characterize the equilibrium strategies at $t = 2$ in each possible subgame without worrying about players' "types." To start, if three tasks remain at $t = 2$, effort is futile, so both players optimally shirk. Since we are abstracting from variation in strategies besides those stemming from differences in μ (which only matter at $t = 1$), players play the symmetric equilibrium in the other potential subgames at $t = 2$. Consequently, if one task remains, both players mix, exerting effort with probability $1 - c$ — this is the unique symmetric equilibrium. If two tasks remain, however, there are multiple symmetric equilibria, so we assume players follow the Pareto-dominant such equilibrium in which both players exert effort, guaranteeing completion of the project.[18] Of note, if the project is still feasible in that either one or two tasks remain, then the per-person expected continuation payoff at $t = 2$ is $1 - c$.

As for the problem at $t = 1$, if a player believes their partner will exert effort with probability p, the perceived expected values of exerting effort and of shirking are $-c + \beta(1 - c)$ and $p \cdot \beta(1 - c)$, respectively. Following the discussion above, if a player perceives their partner as being the same type — i.e. if $\mu = \tilde{\mu}'$ and $\mu_I(0) = \tilde{\mu}'_I(0) = \beta$ — then they will exert effort at $t = 1$ with probability $1 - \frac{c}{\beta(1-c)}$, while believing their partner will do the same. Similarly, a hypothetical time-consistent player who believes their partner is also time-consistent and has identical beliefs will (hypothetically) mix at $t = 1$ with probability $1 - \frac{c}{1-c}$.

Besides isolating the effect of μ on behavior (i.e. only permitting heterogeneity in players' strategies when there is heterogeneity in players types), our focus on the unique, symmetric mixed-strategy equilibria in cases where players perceive their partners as the same type is motivated by additional considerations. One such moti-

[17] This is essentially O'Donoghue and Rabin's (1999, 2001) solution concept, extended to a multi-agent setting.

[18] This equilibrium selection rule can also be motivated by the fact that, for a player who believes their partner is going to shirk, exerting effort at $t = 1$ is only a best-response if both players exert effort in the subgame at $t = 2$ in which two tasks remain.

vation stems from a property of our game (to be discussed at greater length in Section 6) by which it can be collapsed to a strategically-equivalent static game featuring the payoff structure of a two-player "volunteers dilemma."[19] In the volunteers dilemma, players choose whether or not to make a costly sacrifice to provide a public good — such as witnesses to a crime-in-progress choosing whether or not to alert the police. A key aspect of this game is that only one volunteer is needed, so the Pareto-dominant Nash equilibrium involves free-riding from all but one player. Observations from the laboratory and the field, however, reveal a surprisingly high-incidence of outcomes in which no one volunteers to thwart the crime, which suggests that players often fail to select this efficient pure-strategy equilibrium, opting for a mixed-strategy equilibrium instead.[20]

Apart from the tractability- and empirically-based considerations mentioned above, our focus on the unique, symmetric mixed-strategy equilibria for homogeneous teams can also be justified on the grounds that the alternative is trivial and uninteresting. In particular, if we assume that players can reliably select the asymmetric, pure-strategy equilibrium — perhaps through some costless communication device — in which the project is always completed without any excess effort expenditure, then the team functions perfectly and there is no 'problem' in need of a solution. Furthermore, in our framework, teams would always select this equilibrium regardless of their composition. Thus, the alternate selection rule would prevent us from using the framework to consider the challenges that free-riding and procrastination — and of present bias more generally — can pose to team functionality, while precluding the potential for remedies through forms of behavioral diversity.[21]

2.3 Motivation and Efficiency

The first of two key questions the model will be used to answer is: how does a player's beliefs, μ, influence their behavior? Since all players are, in effect, identical at $t = 2$,

[19]Specifically, the $t = 2$ subgame with one task remaining is isomorphic to a static, two-player volunteer's dilemma and, given the unique and symmetric mixed-strategy equilibria is selected in this subgame, the payoff structure for the $t = 1$ subgame is likewise equivalent to that of a volunteer's dilemma.

[20]Darley and Latane (1968) provide early evidence of this well-known "bystander effect" leading to no-volunteer outcomes, and show that the effect is actually exacerbated as the number of players increases. For a review of such evidence and discussions of their theoretical implications, see Diekmann (1985, 1993) and Franzen (1999).

[21]The results of our model would be qualitatively robust to a generalization in which players select asymmetric pure-strategy equilibria and symmetric mixed-strategy equilibria with respective exogenous probabilities $1 - q$ and q for some $q \in (0, 1)$. Since any two teams would generate the same outcome with probability $1 - q$, the directions of our predictions would be preserved, while the expected magnitude of such effects would be weighted by q.

we will focus on how, at $t = 1$, μ determines a player's "motivation" in the following sense:

Definition A player is *motivated* (at $t = 1$) if their strategy is to exert effort. A player is *unmotivated* if their strategy is to shirk. A player is neither motivated nor unmotivated if their strategy is to mix.

Until we put more structure on μ, we will not be able to say much about a player's motivation. With that said, the following lemma holds universally, and will help us prove and provide intuition for later results. All proofs are in the appendix.

Lemma 1 *A player is motivated if they believes their partner is unmotivated, and unmotivated if they believes their partner is motivated.*

The second key question the model will be used to answer is: under what circumstances is a team efficient? Put differently, what types of players should be paired together to ensure the first-best outcome? Again, we can only say so much at present, but the following lemma will be helpful in proving and understanding later results characterizing a team's efficiency:

Lemma 2 *A team is efficient (in that it always achieves a first-best equilibrium) if and only if one player is motivated and the other player is unmotivated.*

Thus, an efficient team is necessarily "diverse" with respect to players' motivational dispositions, but what this actually means in terms of players' beliefs as captured by μ and μ' or in terms of the nature of their naivete and/or sophistication has yet to be determined. Note, in this first-best equilibrium, one out of two players completes a task at $t = 1$, and then both players complete a task at $t = 2$ so that the project is successfully completed without overprovision of effort. Since both players must behave differently at $t = 1$ to achieve this outcome, we can immediately establish the following important implication of Lemma 2:

Corollary 1 *Any team comprised of players with "symmetric" beliefs in that $\mu = \mu'$ is inefficient.*

Although we still have a long way to go in determining what it takes for a team to be efficient, Corollary 1 provides a helpful start by telling us that we can rule out any team in which both players are of the same type. That is, for a team to be efficient, players' must have *asymmetric* beliefs in that $\mu \neq \mu'$.[22]

[22]Note here, that "symmetric" beliefs are not necessarily the same in the sense that both players agree

3 Teams with (Fully) Naive Players

3.1 Reassessing Naive Preconceptions

The first category of players we consider are those who are "fully naive" in that they do not anticipate their own present bias nor do they anticipate their partner's present bias. To distinguish them from other types of time-inconsistent types considered later, each of these fully-naive players will simply be referred to as a *naif*, where we will proceed for now under the assumption that both players are naifs in this sense.

The language used in our motivating definition of "naivete" above reflects a very standard concept from existing models of present-biased discounting, but it will need to be further refined in the current framework in which a player's type will ultimately be defined by their beliefs, μ, on which they base their decision of whether or not to exert effort. Thus, to help formally capture what such naivete actually means in the current framework, we specify that a naif's *preconception* before each period is that neither they nor their partner are present-biased and that this is common knowledge. Put differently, prior beliefs — including all higher-order beliefs — regarding them and their partner's present bias factors are all equal to 1. However, at the beginning of a period, a naif unexpectedly learns of their own present bias, so that their updated, choice-relevant belief regarding their own present bias factor is $\mu_I(0) = \beta < 1$. At this point, the naif may then be compelled to reassess their other preconceptions, but Bayes' rule gives no guidance on how or whether their beliefs change because this discovery of their own naivete is an event to which they previously assigned zero probability.[23] At the lowest degree, for example, a naif may wonder, "does my partner know that I have a present bias" or "does my partner also have a present bias?" and their answers to these questions determine $\mu_I(1)$ and $\mu_O(1)$, respectively. Thus, for this case of a naive player, we can regard μ as their beliefs regarding both players' present biases (or lack thereof) following a possible *reassessment* of their preconceptions. Since a naif's preconceptions are that both players are time-consistent and that this was common knowledge, if a belief $\mu_z(n)$, $z \in \{O, I\}$ has *not*

on everything, while "asymmetric" beliefs are not necessarily different in the sense that players disagree on something. Instead, symmetry is used to describe the case in which a player's beliefs are a mirror-image of their partner's. That is, if you and I are partners, our beliefs are symmetric if I would describe my beliefs (including higher order beliefs) in exactly the same way as you would describe yours. For example, if I believe that you believe that *I* am present biased, then you too would say "I believe that you believe that I am present biased," except from my perspective this actually means that you believe that I believe that *you* are present biased.

[23] While our extension to time-inconsistent preferences is novel, Ortoleva (2012) and Karni and Viero (2013) axiomatically investigate non-Bayesian updating of this sort. See Section 5 for a more detailed discussion.

been reassessed, then $\mu_z(n) = 1$. If, however, the belief has been reassessed, we take this to mean that $\mu_z(n) = \beta$.[24,25]

Assumption 1 *If a belief is not reassessed, corresponding higher-order beliefs are also not reassessed:* $\mu_z(n) = 1$ *implies* $\mu_z(n+k) = 1$, *for any* $n, k \geq 1$ *and* $z \in \{O, I\}$.

Put differently, a higher-order belief can only be reassessed if its corresponding lower-order beliefs have been reassessed — a property we refer to as *sequential reassessment*. The idea here is that, if discovering one's own present bias doesn't prompt a naif to reassess a given preconceived belief, it is unlikely the naif would have revisited whether or not their partner believes they reassessed this preconception which wasn't actually reassessed. Although we believe this to be a natural and intuitively appealing restriction, Assumption 1 can also be relaxed significantly with the main results intact (see Appendix A.16). Also see Section 5 for a discussion on how sequential reassessment, as formalized here, can be motivated as an application (and generalization) of level-k reasoning theories.

Let $R_O, R_I \in \{0, 1, \ldots\}$ denote the *depths* of outward and inward reassessments, respectively. Formally, for $z \in \{O, I\}$,

$$R_z = \begin{cases} 0, & \text{if } \mu_z(n) = 1, \text{ for all } n = 1, 2, \ldots \\ n \in \{1, 2, \ldots\}, & \text{if } \mu_z(m) = \beta \text{ if and only if } m \leq n \\ +\infty, & \text{if } \mu_z(n) = \beta, \text{ for all } n = 1, 2, \ldots \end{cases} \tag{3}$$

Similar to our definitions of $\tilde{\mu}' = (\tilde{\mu}'_O, \tilde{\mu}'_I)$ and of $\mu' = (\mu'_O, \mu'_I)$, we also use $\tilde{R}'_O$ and $\tilde{R}'_I$ to denote the player's beliefs regarding the depths of their partner's outward and inward reassessments, respectively, and R'_O and R'_I to denote the true depths. We can also observe, from our definition of $\tilde{\mu}'$ along with equation (3), that $\tilde{R}'_O$ and $\tilde{R}'_I$ are uniquely determined from R_O and R_I by:

$$\tilde{R}'_O = R_I, \quad \text{and} \quad \tilde{R}'_I = \max\{R_O - 1, 0\}.$$

[24]The unexpected discovery of one's own present bias $\beta < 1$ can be regarded as a paradigm shift where the naif no longer implicitly assumes that everyone's present bias factor is 1 (i.e. that everyone is time-consistent), but is now confronted with the new knowledge that β is also a possible value for a present bias factor, so that if $\mu_z(n)$ is reassessed, this means they have abandoned their corresponding preconception and now assigns $\mu_z(n) = \beta$.

[25]This simplifying (and, in our view, intuitively reasonable) assumption can be relaxed significantly without compromising the main results. In particular, Appendix A.16 demonstrates that the model's predictions hold as long as the highest-degree reassessments in each "direction" (i.e. inward and outward) are on $[\beta, 1)$. For example, $\mu_O = (\beta, \beta, 1, 1, \ldots)$ has the same implications for behavior as $\mu_O = (\mu_O(1), \mu_O(2), 1, 1, \ldots)$, as long as $\beta \leq \mu_O(2) < 1$ (and where $\mu_O(1)$ could be any value).

We now introduce the following terminology to help us categorize naifs based on the overall extent to which they reassess their preconceptions:

Definition *Unlimited reassessment* means $R_I = R_O = +\infty$. *Limited reassessment* means either $R_I < +\infty$ or $R_O < +\infty$ (or both). *Zero reassessment*, a special case of limited reassessment, means $R_I = R_O = 0$.

Additionally, we can classify naifs for which reassessment is limited based on the overall "direction" of their reassessments:

Definition A naif's reassessment is strictly *inward* if $R_I > R_O$ and strictly *outward* if $R_O > R_I$. Accordingly, a naif's reassessment is weakly inward if it is limited with $R_I \geq R_O$, and weakly outward if it is limited with $R_O \geq R_I$. Colloquially, we may think of a naif with inward reassessment as more 'introspective,' on balance, and a naif with outward reassessment as more 'extrospective.'

3.2 Results for Naifs

The first result of this section establishes necessary and sufficient conditions for a naif to be motivated and to be unmotivated, based on the direction of their reassessments:

Proposition 1 *A naif is motivated if and only if their reassessment is strictly outward, and unmotivated if and only if their reassessment is weakly inward.*

This results highlights the polar effects of 'introspection' and 'extrospection' — taken here to mean inward and outward reassessment – on individual motivation. Namely, introspection in this sense can only serve to demotivate a naif, while extrospection can generate the motivation necessary to overcome procrastination.

Since a team is efficient if and only if one player is motivated and the other is unmotivated (Lemma 2), we can establish the following corollary to Proposition 1.

Corollary 2 *A team with two naifs is efficient if and only if one player's reassessment is weakly inward and the other player's reassessment is strictly outward.*

Thus, in a team with two naifs, efficiency requires a new and perhaps unusual form of diversity whereby players must differ with respect to the overall direction with which they reassess their naive preconceptions. Put differently, the team maximizes its performance if one player is introspective and the other is extrospective.

The next two corollaries take stock of the most extreme types of naifs — extreme in terms of the depths of their reassessments. In doing so, we will see how there can be both too little and too much collective reassessment of naive beliefs in a team with two naifs.

Corollary 3 *Nonzero reassessment is a necessary condition for a naif to be motivated. Therefore, in a team with two naifs, nonzero reassessment by at least one player is a necessary condition for efficiency.*

Corollary 4 *Under unlimited reassessment, a naif is neither motivated nor unmotivated. Therefore, any team is inefficient if it includes a naif with unlimited reassessment.*

Taken together, Corollaries 3 and 4 indicate that a team of naifs can only be efficient if some preconceptions are reassessed while other preconceptions are not reassessed. Put differently, collective 'under-thinking' in a team can prevent efficient outcomes — but so can collective 'over-thinking.'

3.3 Outcomes with (Hypothetical) Time-Consistent Teammates

Next we compare teams with two naifs to teams with one or more time-consistent players (TCs) who are not present-biased and also assumed to hold correct beliefs regarding both players' present biases (or lack thereof).[26] These comparisons are straightforward because, in computing naifs' perception-perfect equilibrium strategies, we first needed to characterize equilibrium behavior in hypothetical teams with one or more TCs (as evident in the proof of Proposition 1). For example, a naif with zero outward reassessment believes their partner is a TC and thus must ascertain how a TC would behave in order to determine their perception-perfect best-response.

Corollary 5 *A (hypothetical) team with one TC and one naif is efficient if and only if the naif's reassessment is limited. With two TCs, the team is inefficient.*

Hence, a team that lacks present bias is inherently inefficient, but this inefficiency can be overcome by replacing one TC with a naif, as long as the naif's reassessment is *not* unlimited. This reliance on limited reassessment is noteworthy in part because existing approaches to modeling naive present bias in multi-player games implicitly

[26]Relaxing this (standard) assumption may strengthen our main results. For example, suppose a team includes a TC who believes, perhaps incorrectly, that their partner is also time-consistent (and that this is common knowledge). Then the team must be inefficient because the TC will be neither motivated nor unmotivated (as would also be the case with two TCs). Thus, if TCs were 'unsophisticated,' efficiency would require two present-biased players, as opposed to the current formulation in which only one is needed (as will be seen in Corollary 5).

assume that naive reassessments are, in effect, unlimited — an assumption that would conceal a naif's potential value as a teammate in our setting. Also note, the capacity of the mixed team to attain the first-best outcome does not depend on the direction of the naif's (limited) reassessment, although the direction does determine (in the manner predicted by Proposition 1) which player gets to economize on their task share by procrastinating at $t = 1$ and which player is on the hook for exerting effort in both periods.[27]

4 Teams with Sophisticated Players (Partial and Full)

In this section, we relax our previous restriction that players are fully naive by allowing for various forms of sophistication. As a starting point, we can consider a standard, single-agent notion of sophistication as correctly anticipating one's own present bias prior to the present. However, for a variety of reasons, this notion will have to be refined and expanded in the current framework. First, we will now want to consider sophistication with respect to both one's own and one's partner's present bias. Second, we need to say something about higher-order beliefs to complete our definitions. Thus, just as we would naturally conceive sophistication as it applies to lowest-order beliefs, sophistication with respect to higher-order beliefs will mean that such beliefs are correctly anticipated. Lastly, if a player's prior beliefs are correct, there is no reason to believe such beliefs will be reassessed. Therefore, in contrast to our working conception of naivete, we do not need to separately consider a player's prior beliefs apart from their choice-relevant beliefs in defining sophistication. Consequently, sophistication will simply mean that the associated beliefs — as represented in μ — at the time choices are made are correct.

Formally, we define two forms of sophistication (and parallel forms of naivete) as follows:

Definition
(i-a) A player is *self-sophisticated* if all of their (choice-relevant) beliefs regarding their own present bias are correct: $\mu_I(n) = \tilde{\mu}'_O(n) = \mu'_O(n)$, for all $n = 1, 2, \ldots$.
(i-b) A player is *other-sophisticated* if all of their beliefs regarding their partner's

[27]For the case of a team with one TC and one fully-sophisticated present-biased player (we will elaborate on sophisticated players in the next section), our equilibrium selection rules do not pin down a unique equilibrium — both efficient and inefficient equilibria are possible. For a team with two fully-sophisticated time-inconsistent players for whom it is common knowledge that both players are present-biased the equilibrium is identical to the equilibrium for a team with two naifs who both have unlimited reassessment. Hence, in addition to present bias, some naivete is needed in a team to guarantee efficiency.

present bias are correct: $\mu_O(n) = \tilde{\mu}'_I(n-1) = \mu'_I(n-1), \quad \text{for all} \;\; n = 1, 2, \ldots.$

(ii-a) A player is *self-naive* if their preconception is that they are not present-biased and that this is common knowledge.

(ii-b) A player is *other-naive* if their preconception is that their partner is not present-biased and that this is common knowledge.

We can observe that the naifs we considered in Section 3 were both self-naive and other-naive (i.e. "fully naive," as we called them). Using our new terminology, we can now define different types of players, with some level of sophistication, as follows:

Definition

(i) A *self-sophisticate* is a player who is self-sophisticated, but other-naive.

(ii) An *other-sophisticate* is a player who is other-sophisticated, but self-naive.

(iii) A *full-sophisticate* is a player who is self-sophisticated and other-sophisticated.

At this point, it might be worthwhile to highlight some asymmetries inherent in both sets of definitions provided above. To start, in the first set of definitions (the adjectives), we see that the two forms of sophistication (i-a and i-b) refer to properties of a player's choice-relevant beliefs, $\mu = (\mu_O, \mu_I)$, while the two forms of naivete (ii-a and ii-b) only refer to a player's preconceptions while leaving open how their choice-relevant beliefs might look. This difference stems directly from the feature of our setup mentioned above that preconceptions are only relevant with regards to naive beliefs. Furthermore, in light of what actually causes reassessments of naive preconceptions, this difference also gives rise to other asymmetries pertaining to how the various types of players introduced in the second set of definitions (the nouns) arrive at their decisions. In particular, an other-sophisticate may reassess their naive (in this case, inward) preconceptions in the manner described in Section 3, yet a self-sophisticate will never reassess their naive (outward) preconceptions.

To understand why other-sophisticates — but not self-sophisticates — may reassess their naive preconceptions, recall from Section 3 that a specific event served as a necessary trigger for such reassessments: namely, the unexpected discovery of one's own present bias. Thus, since an other-sophisticate is self-naive, this trigger will still exist for such players. However, since a self-sophisticate correctly anticipates their own present bias, they are never confronted with any new information to challenge their other-naivete so that there will never be an impetus to reassess their naive outward beliefs.

Therefore, like the naif considered in Section 3, only the other-sophisticate finds herself unexpectedly questioning their preconceptions immediately prior to their effort choice at $t = 1$, while the self-sophisticate's subjective experience is quite different in that, like the full-sophisticate, they carry on without ever questioning (or having reason to question) what they previously thought to be true. As we will see, these different "experiences," so to speak, give rise to equilibrium behaviors that are not simple reflections of one another, and in our formal results, it will therefore often be practical to consider these two types of partial sophisticates separately (as we do below):

Proposition 2 *A self-sophisticate is unmotivated, regardless of their partner's type.*

Hence, a partially sophisticated player who is self-sophisticated but other-naive *always* procrastinates in the sense that they never exerts effort to complete a task before the final period. Furthermore, as will be fully apparent by the end of this section, when we categorize types solely by their status as naive or sophisticated with respect to both one's own and one's partner's present biases, the self-sophisticate is the only type of player for whom procrastination of this sort can always be expected. Put differently, naifs, other-sophisticates, and full-sophisticates can all be motivated in at least *some* circumstances — but self-sophisticates are never motivated.

We can understand the intuition for Proposition 2 by first noting that the self-sophisticate never reassesses their outward beliefs — i.e., $\mu_O = (1, 1, 1, \ldots)$ — so that their full set of beliefs will ultimately resemble that of a naif, as considered in Section 3, for whom reassessed preconceptions are necessarily inward (although possibly weakly). In turn, just as we saw in Proposition 1 for a naif with weakly inward reassessments, this self-sophisticate is always unmotivated.

This next result considers the more complicated case of an other-sophisticate:

Proposition 3
(i) Under finite inward reassessment ($R_I < \infty$), an other-sophisticate is motivated if and only if one of the following is true: (a) their partner is self-sophisticated (i.e. either a self-sophisticate or a full-sophisticate), or (b) their inward reassessment is weakly shallower than their partner's inward reassessment ($R_I \leq R'_I$).
(ii) Under infinite inward reassessment, an other-sophisticate is never motivated.

To understand why an other-sophisticate is motivated with a self-sophisticated partner (i-a), first note that in such cases both players have the same, correct beliefs regarding the partner's present bias, which implies that it truly is common knowledge that the partner is present biased: $\mu_O = \tilde{\mu}'_I = \mu'_I = (\beta, \beta, \beta, \ldots)$. Therefore, the other-sophisticate's outward beliefs end up looking like those of an other-naive player with infinite outward reassessment, so that with finite inward reassessment, the other-sophisticate's full set of beliefs resemble that of a strictly outward naif who is always motivated according to Proposition 1. The other-sophisticate is similarly motivated if their inward reassessment is weakly shallower than that of their partner (i-b) because their partner's inward reassessment determines the effective depth of their own sophisticated outward beliefs from $\mu_O(n) = \mu'_I(n-1)$, $n = 1, 2, \ldots$. As a result, the other-sophisticate's beliefs would again be outward on balance, making their motivated in the same vein as the strictly outward naif. In contrast, an other-sophisticate is never motivated under infinite reassessment for the same reasons that the naif in Proposition 1 was never motivated when $R_I = \infty$. In particular, under infinite reassessment, an other-sophisticate believes it is common knowledge that both players are present-biased and therefore mixes under the (possibly false) assumption that their partner has identical beliefs and will therefore mix as well.

Proposition 4 *A full-sophisticate is motivated if and only if their partner is unmotivated, and is unmotivated if and only if their partner is motivated.*

Hence, a full-sophisticate's motivational status is optimal in the sense that it will be the opposite of their partner's, so that their perception-perfect strategy is indeed the true best-response. This prediction is not too surprising since all of the full-sophisticate's beliefs are necessarily correct. However, as we will soon see, if the full-sophisticate's partner is neither motivated nor unmotivated (as would be the case if their partner was also a full-sophisticate), the situation is less promising from an efficiency perspective.

Proposition 5 *A team comprised of two sophisticates with asymmetric beliefs (i.e. $\mu \neq \mu'$) is efficient, unless one player is an other-sophisticate with infinite inward reassessment and the other player is a self-sophisticate.*

One key implication of Proposition 5 is that with (at least partially) sophisticated players, pairing different types is almost enough to guarantee efficiency, since there is only one special case in which efficiency is not guaranteed. The lesson here, it may seem, is that similarity promotes inefficient outcomes while dissimilarity promotes efficient outcomes. However, this generalization is problematic in light of our one exception to the rule. That is, a team will be inefficient if it is comprised of the two types of sophisticates who are (arguably) as dissimilar as possible. In particular, one player in this inefficient team is self-naive, other-sophisticated, and has infinite reassessment of their naive beliefs, while the other player is the opposite in each regard: other-naive, self-sophisticated, and by way of being self-sophisticated (and thus having no impetus to revisit their preconceptions), has zero reassessment of their naive beliefs.

Another apparent lesson of Proposition 5 is that sophistication is good because, along with asymmetric beliefs, sophisticated players almost always achieve the first-best outcome. However, such a generalization would also have a major hole because too much sophistication can be an obstacle to efficiency:

Proposition 6 *A team with a full-sophisticate is inefficient if and only if the other player is either a full-sophisticate as well or has reassessed beliefs that are identical to those of a full-sophisticate.*

Thus, pairing two full-sophisticates will never be efficient. Moreover, pairing a full-sophisticate with a naive (at least partially) player whose reassessed beliefs happen to coincide exactly with what a full-sophisticate would have believed all along will also be inefficient. With that said, pairing a full-sophisticate with anyone else — equivalently, any potential partner for whom $\mu' \neq \mu$ — does guarantee efficiency.

While the above result implicitly characterizes the efficiency of a team in which the (in this case, full) sophisticate's partner is either a naif or another type of sophisticate, we have yet to consider the efficiency of teams that consist of one partial sophisticate along with a naif. The next two results fill this gap:

Proposition 7 *A team with one self-sophisticate and one naif is efficient if and only if the naif's reassessment is strictly outward.*

As we've learned so far, self-sophisticates are always unmotivated while naifs are motivated only if their reassessments are strictly outward; thus, such strictly outward reassessment by the naif is necessary and sufficient for efficiency in a team comprised of one of each type.

Proposition 8 *Consider a team with an other-sophisticate and a naif.*
(i) If the naif's reassessment is weakly inward, the team is efficient if and only if the other-sophisticate's inward reassessment is weakly shallower than that of the naif.
(ii) If the naif's reassessment is strictly outward, the team is efficient if and only if the other-sophisticate's inward reassessment is strictly deeper than that of the naif.

This result is perhaps best understood if we interpret it while referring back to Proposition 3, which characterized the individual strategies that an other-sophisticate may employ. In particular, part (i-b) of Proposition 3 established that an other-sophisticate is motivated if their inward reassessment is weakly shallower than their partner's. Part (i) of the current result then follows because a weakly inward naif is unmotivated, so that the other-sophisticate must be motivated to ensure efficiency. Part (ii) similarly follows because a strictly outward naif is motivated, so that the other-sophisticate must be unmotivated for the team to be efficient.

5 Implications for Non-Bayesian Updating of Null Events

As mentioned earlier, our theory links established models of present-biased discounting to the recently-revived theoretical interest in non-Bayesian updating in response to zero-probability events (Ortoleva, 2012; Karni and Viero, 2013).[28] We feel it is quite natural to bridge this gap since the standard model of fully-naive present bias implicitly assumes that agents can assign zero-probability to events that may actually occur, thus providing a well-grounded point of entry into the realm of non-Bayesian updating.

Besides the new application to present bias, our model formalizes a novel "sequential reassessment" approach to Non-Bayesian updating in which an agent only updates their prior beliefs about some parameter up to some threshold order (between zero and infinity) so that higher-order beliefs above this threshold are not updated. In addition to its tractability, we believe sequential reassessment is appealing because it captures a realistic aspect of human cognition in that lower-order beliefs are more easily and naturally contemplated than higher-order beliefs.[29] For example, it is highly unlikely that you would stop to wonder whether or not I believe that you believe that I am ambidextrous without wondering whether I am ambidextrous in

[28]Also see Blume et al.'s (1991) and Myerson's (1986) earlier extensions of subjective expected utility theory to address updating in response to null events.

[29]The tractability of our approach is greatly facilitated by our simplifying assumption that players are absolutely certain (one way or another) regarding each of their beliefs. This assumption is carried over from canonical, single-agent models of naive present bias — and also exists in models of partially naive present bias, in which the agent anticipates a present bias but underestimate its severity (O'Donoghue and Rabin, 2001). Formally, partial naivete in this sense means the agent is 'positive' their future present bias will be $\hat{\beta} \in (\beta, 1)$. Sequential reassessment would also be amenable to the analysis of this type of player because confronting the true present bias, $\beta < \hat{\beta}$, still represents the realization of a zero-probability event. While probabilistic beliefs regarding one's present bias could be another reasonable approach to modeling partial naivete, it would not provide a natural entry point for sequential reassessment in response to null events (assuming the agent assigns a nonzero probability to the true present bias factor).

the first place. Moreover, by allowing unequal inward and outward reassessments, sequential reassessment can speak to individual variation in 'dispositions' of social cognition — some people may be more introspective and others more extrospective — while allowing for variation in the depths of reassessments fits with the idea that some of us may be more inclined to question our preconceived notions than others.

Sequential reassessment can also be regarded as a new application of *level-k reasoning* models to the domains of time-inconsistent preferences and of non-Bayesian updating in response to null events.[30] Level-k reasoning models are founded on the premise that, in many cases, standard equilibrium concepts rely on implausible assumptions about humans' reasoning capacities and that a more realistic characterization of how people actually think in strategic environments is needed. Formally, these models assume that players' beliefs (and strategies) are anchored to some naive prior held by a "level-0" player such that a level-1 player best-responds to the strategy of a level-0 player, a level-2 player best-responds to a level-1 player, and so on. Under finite sequential reassessment, naifs similarly believe they are one step ahead of this partner (in each direction) and best-respond accordingly. For example, suppose Alice is a naif with $R_O = R_I = 1$ (i.e. Alice believes that Bob knows she is present-biased and that Bob is present-biased too, but maintains all other preconceptions). Then Alice best-responds under the belief that Bob has only engaged in one degree of outward reassessment: $\tilde{R}'_O = 1$ and $\tilde{R}'_I = 0$, falling one degree short of what she would perceive as the correct beliefs from Bob's vantage point ($R'_O = 2$ and $R'_I = 1$). More generally, sequential reassessment can be mapped to level-k reasoning, where a naif's level is given by $k = \max\{2R_I + 1, 2R_O\}$, and where level-0 represents a hypothetical time-consistent player who believes it is common knowledge that neither player is present-biased.[31] Indeed, real-life players in normal form games do appear to exhibit a limited depth of reasoning, as many empirical and experimental studies show that level-k reasoning often out-predicts models based on standard equilibrium concepts (for a review, see Crawford et al., 2013).

[30] See Nagel (1995), Stahl and Wilson (1995), Costa-Gomes et al. (2001), Costa-Gomes and Crawford (2006), and the closely-related 'Cognitive Hierarchy' model of Camerer et al. (2004).

[31] In our setting, two players with different beliefs can be on the same level. For instance, suppose Alice Y's reassessment is as described above with $R_O = R_I = 1$ and Alice Z's reassessment depths are $R_O = 0$ and $R_I = 1$. Then both Alices are level-3 players best-responding to what they perceive as level-2 opponents (Bob Y and Bob Z) who themselves are under the false impression that they are best-responding to a level-1 opponent, who in turn is best-responding to the level-0 time-consistent type. The subtle distinction here is that Alice Y believes that Bob Y is present-biased with reassessment depths $R'_O = 1$ and $R'_I = 0$ while Alice Z believes that Bob Z is *not* present-biased yet, oddly enough, has the same nonzero outward reassessment $R'_O = 1$ that Alice Y perceives in Bob Y.

As we alluded to above, sequential reassessment represents a distinct approach for modeling non-Bayesian reactions to null events in relation to recently-proposed axiomatic decision theories that address this topic. In Ortoleva's (2012) "hypothesis testing" model, the realization of a null state prompts the agent to discard their priors in favor of a different set of priors. Unlike sequential reassessment, this approach relies on a pre-existing set of hypotheses (prior over priors) from which to draw. In Karni and Viero's (2013) "reverse Bayesianism" model, the discovery of a previously-unconceived possibility (i.e., a consequence, act, or consequence-act link) leads the agent to form posterior beliefs on those that were previously-conceived by proportionately reducing the associated priors while remaining agnostic to the magnitude of the posterior for the newly-discovered possibility. This approach rules out a key property of sequential reassessment by which it allows for, and derives important results from, the possibility that some priors are reassessed while others or not. If we were to limit our types to those satisfying proportional reassessment of non-contradicted priors, the naif in Section 3 would either engage in zero or unlimited reassessment, thus precluding, for example, the strictly outward reassessment that was necessary to motivate a naif and also to achieve efficiency in a team with two naifs (see Proposition 1 and Corollary 2).

We believe that updating in response to null events may follow a variety of approaches and that sequential reassessment may be more or less viable than the others depending on the situation. With that said, we also believe that sequential reassessment is the most reasonable for the situation we consider. To start, the other approaches are ideally-suited to situations in which agents respond to noisy signals about a 'state' (broadly speaking) by forming new beliefs regarding the likelihoods of other, mutually-exclusive elements of the state space. Agents in our setting, however, face a different problem: after learning the true value about a private parameter, they are left to form beliefs regarding the private value of their partner's parameter and all higher-order (and naturally sequenced) beliefs thereof.

Besides these technical considerations, the hypothesis testing and reverse-Bayesianism models are probably better suited for modeling more perceptive agents than those considered here. Namely, these approaches implicitly require a capacity for learning that naively present-biased agents are commonly understood to lack. A naif, as standardly conceived in single-agent models, unexpectedly confronts their present bias in every period, but carries on treating their inevitable future present bias as a zero-probability event. In the hypothesis testing and reverse-Bayesianism models, however, agents will not form posterior beliefs that have already been contradicted. Related to this, this naif holds (by definition) a false assumption, devoid of any doubt, regarding a parameter that would in principle be known if it weren't for a lack of sophistication. Sequential reassessment is uniquely in keeping with this self-assured concept of naivete as it renders posterior beliefs that are not necessarily correct, yet nonetheless held with certainty, while the hypothesis testing and reverse-Bayesianism models may better describe probabilistic beliefs held by less oblivious agents regarding invariably uncertain states.

6 Relation to a Static Volunteer's Dilemma with Self-Image Biases

Although we believe it to be the most natural and well-grounded motivation, our analysis of naive present bias in a dynamic collaboration game could be regarded as one of multiple candidate motivations for a less-specific approach in which players may confront false preconceptions regarding some generic payoff parameter. Moreover, by constructing a game in which we could effectively confine our analysis to the $t = 1$ decision, the framework can, in essence, be collapsed to a static game in which a generic 'payoff naivete' can be analyzed devoid of dynamic considerations.[32] To see this, consider the 'ex-ante' game with the payoff structure given at left below, for some $\bar{\phi} \in (0, 1)$. Here, player 1 is certain that this payoff structure is common knowledge, but when called to select an action, suddenly learns that their payoff in the bottom-left cell is higher than they initially expected, $\phi > \bar{\phi}$, giving rise to the payoff structure on the right:

	a	b
a	$(0, 0)$	$(0, \bar{\phi})$
b	$(\bar{\phi}, 0)$	$(-1, -1)$

Ex-Ante

$\longrightarrow$

	a	b
a	$(0, 0)$	$(0, \phi')$
b	$(\phi, 0)$	$(-1, -1)$

Action Selection

The parameter $\bar{\phi}$ in the top-right cell has also been replaced, in this case by ϕ', reflecting the possibility that learning about ϕ may compel player 1 to revisit their

[32]The collapsibility of our base model largely stems from our efforts to make it as simple as possible, avoiding any extraneous generalizations that aren't already in standard models of present bias. Accordingly, this feature would not be preserved in many natural extensions of the base model, thus necessitating a dynamic formulation with present bias. A few examples of extensions or modifications that might be worth considering include: eliminating deadlines, modeling non-Bayesian learning between periods in response to an inconceivable action by one's partner (see Footnote 16), exploring other dynamic games (such as a repeated Prisoner's Dilemma), considering teams of different sizes, and simultaneously varying the deadline length along with the total task requirements.

prior assumption regarding player 2's corresponding parameter, now labeled as ϕ'. If we presume player 1 only entertains two possibilities, $\phi' = \phi$ and $\phi' = \bar{\phi}$, and similarly restrict the space for all higher-order beliefs regarding ϕ and ϕ' to $\{\phi, \bar{\phi}\}$, then it is straightforward to show that, for $\bar{\phi} = \frac{c}{1-2c}$ and $\phi = \frac{c}{\beta(1-c)-c}$, their problem is strategically-equivalent to the $t = 1$ subgame faced by the present-biased naif in Section 3.

The static game described here can be interpreted as a two-player "volunteer's dilemma." In particular, action a — the analog to exerting effort at $t = 1$ in our dynamic game — represents a choice to volunteer to perform a costly task that generates a public good (and only requires one volunteer to do so). A common illustration of the volunteer's dilemma is a situation with multiple witnesses to a crime-in-progress. If at least one witness volunteers to alert the police, the crime will be thwarted, but if no one volunteers, the criminal will succeed to the detriment of the public interest.

In this story, the 'action selection' payoff matrix would represent a scenario in which player 1 discovers a higher-than-anticipated payoff from not volunteering in the event that player 2 volunteers. As it arises here, it may not be obvious how to interpret 'payoff naivete,' but one possible interpretation relates to biases in the maintenance of a positive self-image. To illustrate, suppose players derive utility from a positive self-image — i.e. a view of oneself as a 'socially responsible' person — that can be harmed by 'socially irresponsible' actions, which in this setting means choosing b (not volunteering). A 'naif' here initially believes there is an intrinsic cost from choosing b, but unexpectedly discovers a capacity to reconcile such socially irresponsible behavior with a positive self-image in the event that it does no harm (i.e., as long as their partner volunteers). Thus, "getting away" with not volunteering would not entail the intrinsic self-image cost, making the associated (net) payoff higher than anticipated: $\phi > \bar{\phi}$.[33] A revelation of this sort may be interpreted as a temporary lowering of one's personal ethical standards to permit a 'one-time' exemption, where naivete may allow such dynamically inconsistent perceptions to persist.

The interpretation of the 'naive volunteer's dilemma' outlined above can be understood in terms of established psychological concepts, some of which have appeared in economic theory. The idea that prosocial behavior is incentivized by a preference to maintain a positive self-image is at the heart of Brekke et al.'s (2003) "moral

[33] We presume that the self-image cost is implicitly reflected in the ex-ante payoff matrix, with the net utility associated with socially responsible behavior normalized to 0.

motivation" model, while the broader notion that a lower self-image carries a hedonic cost is inherent in some theories of overconfidence (Benabou and Tirole, 2002; Koszegi, 2006). In turn, applications of Liberman and Trope's (1998, 2003) temporal construal theory to ethical behavior demonstrate a human tendency to judge one's own future (hypothetical) transgressions in terms of rigid ethical principles, while situation-specific considerations can soften perceptions of present behavior (Eyal et al., 2008; Tenbrunsel et al., 2010).[34]

If we re-interpret the formal predictions of our (originally-conceived) dynamic collaboration model through the lens of this 'naive volunteer's dilemma,' it suggests that a positively-biased self-image, broadly speaking, can enhance collective efficiency. In a somewhat similar vein, Gervais and Goldstein (2007) show that positive self-perception biases, taken here to mean overestimating one's marginal return to effort, can mitigate free-riding and coordination failure in firms, while Benabou (2013) demonstrates that "wishful thinking" (i.e. ignoring bad news regarding a project's value) can also promote better outcomes and mitigate free-riding by boosting group morale. While these results present a similar flavor to our own, they rely on a positively-biased self-image to inflate the perceived payoff of socially responsible behavior. In contrast, our approach shows how a positively-biased self-image can, perhaps paradoxically, promote socially responsible behavior (leading to socially efficient outcomes) even when the bias increases the perceived payoff associated with socially irresponsible behavior.

7 Concluding Remarks

This paper presented a model of collaboration among present-biased players, each of whom may be fully-naive, fully-sophisticated, or a hybrid thereof, i.e. self-sophisticated regarding their own present bias yet other-naive regarding their partner's (or vice versa). Importantly, we allowed for variation in the "depths" and "directions" with which naive preconceptions (including higher-order beliefs) are reassessed when players confront their own present bias. In doing so, we avoided ad-hoc assumptions on

[34]Tenbrunsel et al. elaborate on this tendency, describing how transient influences can guide unethical behaviors that were overlooked in past predictions of current behavior and will also be overlooked in future recollections of current behavior, where selective memory, post-hoc justifications, and shifting ethical standards are subconsciously deployed to help sustain this so-called "ethical mirage." Rotela and Richeson (2013) document how selectively forgetting past wrongdoings can also serve to inflate one's perceptions of other in-group members, providing a potential basis for 'other-naivete' among players who share a common identity.

how naive beliefs are updated in a realm where Bayes' rule does not apply, while also identifying new behavioral elements that, according to our model, turn out to be crucial determinants of individual motivation and team efficiency.

To highlight one example, we showed that a team with two fully-naive players will be efficient, as long as the team is diverse with respect to the directions of players' reassessments. One noteworthy aspect of this prediction is that it shows how present-bias and naivete can enable efficient outcomes in a setting where a team with two time-consistent players (and also a team with two fully-sophisticated present-biased players) would be inefficient. Moreover, in achieving this efficient outcome, the 'extrospective' naif with outward reassessment is motivated to exert effort at the beginning of the project, while a time-consistent player would procrastinate with some positive probability. Since present-bias — particularly naive present bias — is known to make an agent prone to severe procrastination on individual projects, this heightened motivation to complete a task early in a team setting reveals a different 'side' of the oft-maligned naively present-biased agent.

Overall, we offer a distinct approach to diversity and team functionality that embraces the incentive issues (and related coordination issues) highlighted in the economics literature along with the notion that a key instrument to ensure efficiency within a team can be its composition. Our broad conclusion is that diversity among agents in their 'behavioral dispositions' can often be efficiency enhancing. With that said, our model also points to the potential value of moderation in this sense, as maximizing diversity on multiple dimensions may backfire — as we saw when pairing opposite types of partial sophisticates who also represent opposites in the depths with which they reassess their naive preconceptions.

A Appendix

A.1 Proof of Lemma 1

If a player believes their partner is unmotivated, the expected value of shirking is 0 and the expected value of effort is $-c + \beta(1 - c) > 0$. Therefore the player (who believes their partner is unmotivated) is motivated. If a player believes their partner is motivated, the expected value of shirking is $\beta(1 - c)$ and the expected value of effort is $-c + \beta(1 - c) < \beta(1 - c)$. Therefore the player (who believes their partner is motivated) is unmotivated. ■

A.2 Proof of Lemma 2

If the team completes more than three tasks, there is inefficient overprovision of effort. If the team completes less than three tasks, there is underprovision as the team fails to complete the project. Therefore, the team must complete three tasks in any first-best equilibrium. If the team completes zero tasks at $t = 1$ (i.e. both players shirk), this three-task outcome is impossible. If the team completes two tasks at $t = 1$ (i.e. both players exert effort), then both players mix at $t = 2$; therefore, in this case, the inefficient two-task and four-task outcomes, as well as the efficient three-task outcome all have a nonzero probability of occurring, so that the first-best is not assured. If the team completes one task at $t = 1$ (i.e. one player shirks and the other exerts effort), both players exert effort at $t = 2$, so that the three-task outcome is attained. It is assured that the team will complete one task at $t = 1$ if and only if one player is motivated and the other is unmotivated. Therefore, the team is efficient if and only if one player is motivated and the other player is unmotivated. ■

A.3 Proof of Corollary 1

Since behavior is uniquely determined by μ, if $\mu' = \mu$, the players described by these beliefs must be the same with respect to their status as motivated, unmotivated, or neither. Therefore, from Lemma 2, such a team must be inefficient. ■

A.4 Proof of Proposition 1

We first proceed by induction. In particular, we need to establish the following:

(i) Under zero reassessment, a naif is unmotivated

(ii) Take any $n = 0, 1, \ldots$. If a naif with $R_O = m$ and $R_I = n$ is unmotivated for any $m = 0, 1, \ldots, n$, then a naif with $R_O = n + 1$ and $R_I = m$ is motivated for any $m = 0, 1, \ldots, n$.

(iii) Take any $n = 1, \ldots$. If a naif with $R_O = n$ and $R_I = m$ is motivated for any $m = 0, 1, \ldots, n - 1$, then a naif with $R_O = m'$ and $R_I = n$ is unmotivated for any $m' = 0, 1, \ldots, n$.

Proof of (i). Under zero reassessment, a naif believes their partner is a time-consistent player who believes that his partner (i.e. the naif herself) is also time-consistent and that this is all mutual information. A hypothetical time-consistent player who believes he is paired with an identical player mixes, exerting effort with probability $1 - \frac{c}{1-c}$ (see Section 2.2). Therefore, the naif's perceived expected values of exerting effort and of shirking are $-c + \beta(1-c)$ and $\left(1 - \frac{c}{1-c}\right)\beta(1-c) = \beta(1-2c)$, respectively. Since $\beta(1-2c) > -c + \beta(1-c)$, the naif is unmotivated under zero reassessment.

Proof of (ii). For a naif with $R_O = n+1$ and $R_I = m$, we have that $\tilde{R}'_O = m$ and $\tilde{R}'_I = n$. Given $m \leq n$, this naif thus believes their partner is unmotivated, by assumption. From Lemma 1, the naif with $R_O = n+1$ and $R_I = m$ — and therefore with strictly outward reassessment — is motivated.

Proof of (iii). For a naif with $R_O = m$ and $R_I = n$, we have that $\tilde{R}'_O = n$ and $\tilde{R}'_I = m - 1$, provided $m \geq 1$. Given $m \leq n$, this naif thus believes their partner is motivated, by assumption. From Lemma 1, the naif with $R_O = m \geq 1$ and $R_I = n-1$ — and therefore with weakly inward reassessment — is unmotivated. Now if $m = 0$, the naif with $R_O = m$ and $R_I = n$ believes their partner is a time-consistent player where $\tilde{R}'_O = n$ and $\tilde{R}'_I = 0$ (although a hypothetical time-consistent player's beliefs would presumably not be based on reassessments triggered by unexpectedly discovering one's present bias, as they are for a naif, R_O and R_I can be defined from μ in the exact same way). Now we know that a naif with $R_O = n$ and $R_I = 0$ is motivated, by assumption. Therefore, if p is the probability that this naif with $R_O = n$ and $R_I = 0$ believes their partner will exert effort, $-c + \beta(1-c) > p\beta(1-c)$. Adding $(1-\beta)(1-c)$ to both sides, we get $1 - 2c > (p\beta + (1-\beta))(1-c) \geq p(1-c)$. Noting that, to the hypothetical time-consistent player with $R_O = n$ and $R_I = 0$, the perceived expected values of exerting effort and of shirking are $1 - 2c$ and $p(1-c)$, respectively, this type of (hypothetical) player is motivated. Thus, from Lemma 1, the naif who believes their partner is of this type is unmotivated.

Together, parts (i), (ii), and (iii) prove that a naif is motivated under strictly outward reassessment and unmotivated under weakly inward reassessment. Now a naif whose reassessment is neither strictly outward or weakly inward has unlimited reassessment, and under unlimited reassessment, a naif believes their partner is of the same type and thus mixes with probability $1 - \frac{c}{\beta(1-c)}$ (see Section 2.2). Therefore, by process of elimination, a strictly outward reassessment is not only sufficient, but also necessary for a naif to be motivated, and a weakly inward reassessment is likewise not only sufficient, but also necessary for a naif to be unmotivated. ■

A.5 Proof of Corollary 2

From Proposition 1, we know that a team consists of one motivated player and one unmotivated player if and only if one player's reassessment is weakly inward, the other player's reassessment is strictly outward, and both players' reassessments are limited. From Lemma 2, we know that being comprised of such players is necessary and sufficient condition for the team to be efficient. ■

A.6 Proof of Corollary 3

From Proposition 1, we know that a naif is motivated if and only if $R_O > R_I$. Therefore a naif must have nonzero reassessment with $R_O > 0$ to be motivated. Nonzero reassessment by at least one player is thus necessary for efficiency because, from Lemma 2, we know that a team must have one motivated player to be efficient. ■

A.7 Proof of Corollary 4

The assertion that a naif is neither motivated nor unmotivated under unlimited reassessment is a direct implication of Proposition 1 (the mixing behavior of such players is addressed in Proposition 1's proof). From Lemma 2, we know that any player on an efficient team must either be motivated or unmotivated. Therefore a team that includes a naif with unlimited reassessment must be inefficient. ■

A.8 Proof of Corollary 5

First note that Proposition 1 still applies for a naif with a time-consistent partner because a naif's status as motivated or unmotivated (or neither) does not depend on who their partner actually is. Next, since a time-consistent player's beliefs are correct, Lemma 1 implies that the time-consistent player will be motivated (unmotivated) if their fully-naive partner is in fact unmotivated (motivated). In turn, from Proposition 1 and Lemma 2, we get the desired result for the "mixed" team. As for the case of two time-consistent teammates, this team must be inefficient in light of Lemma 2 because these teammates must be the same with respect to their motivational statuses. ■

A.9 Proof of Proposition 2

Since a self-sophisticate does not reassess their naive outward beliefs, we know that $\mu_O(n) = 1$ for all $n = 1, 2, \ldots$. Therefore, for any μ_I, the self-sophisticate's beliefs are effectively weakly inward, and they are therefore unmotivated. ■

A.10 Proof of Proposition 3

We first consider the case of finite inward reassessment. Now the other-sophisticate's partner must be one of the following: a full-sophisticate, a self-sophisticate, an other-sophisticate, or a naif.

If the partner is self-sophisticated (i.e. a full-sophisticate or a self-sophisticate), we must have that $\mu_O(n) = \mu'_I(n) = \beta$ for all $n = 1, 2, \ldots$. Since the other-sophisticate's inward reassessment is finite, $\mu_I(n) = \beta$ if and only if $n < k$ for some finite $k \geq 0$ (and $\mu_I(n) = 1$ for all $n \geq k$). Therefore, the other-sophisticate's beliefs are effectively strictly outward, and thus the other-sophisticate is motivated if their partner is self-sophisticated.

If the other-sophisticate's partner is self-naive (i.e. an other-sophisticate or a naif), then $\mu_O(n) = \beta$ if and only if $n \leq R'_I + 1$. If the other-sophisticate's inward beliefs are weakly shallower than their partner's, i.e. $R_I \leq R'_I$, then $\mu_I(n) = \beta$ if and only if $n \leq R_I$ with $R_I < R'_I + 1$. Therefore, if $R_I \leq R'_I$, the other-sophisticate's beliefs are effectively strictly outward and thus the other-sophisticate is motivated. If $R_I > R'_I$, however, then $\mu_I(n) = \beta$ if and only if $n \leq R_I$ with $R_I \geq R'_I + 1$. Thus, if $R_I > R'_I$, the other-sophisticate's beliefs are effectively weakly inward and thus the other-sophisticate is unmotivated.

Next we consider infinite inward reassessment by the other-sophisticate, i.e. $\mu_I(n) = \beta$ for all $n = 1, 2, \ldots$. If $\mu_O \neq \mu_I$, then the other-sophisticate's beliefs must be effectively strictly inward, in which case they are unmotivated. If $\mu_O = \mu_I$, then $\mu = \tilde{\mu}'$, i.e. they believe their partner is of the same type. Thus, the other-sophisticate in this case is neither motivated nor unmotivated. ■

A.11 Proof of Proposition 4

For a full-sophisticate, $\tilde{\mu}' = \mu'$. Thus, since any player's motivation status is fully determined by their beliefs, the full-sophisticate believes their partner is motivated if and only if their partner truly is motivated (and similar if their partner is unmotivated). Therefore, from Lemma 1, we get our result. ■

A.12 Proof of Proposition 5

First note that, if one player is an other-sophisticate with infinite inward reassessment and the other player is a self-sophisticate, then we know from Proposition 3 that the former player is not motivated and from Proposition 2 that the latter player is also not motivated. Therefore, in light of Lemma 2, such a team is inefficient.

To show that alternate compositions are efficient, there are four potential cases to consider:

(a) Both players are other-sophisticates with $R_I \neq R'_I$. Since both players are other-sophisticates, we know that $\mu_O(n) = \beta$ if and only if $n \leq R'_I + 1$ and $\mu'_O(n) = \beta$ if and only if $n \leq R_I + 1$. Now take $R_I > R'_I$, without loss of generality. In this case, $\mu_I(n) = \beta$ if and only if $n \leq R_I$ with $R_I \geq R'_I + 1$. Therefore, the beliefs given by μ are effectively weakly inward, and thus this player is unmotivated. Meanwhile, $\mu'_I(n) = \beta$ if and only if $n \leq R'_I$ with $R'_I < R_I + 1$. Thus, the beliefs given by μ' are effectively strictly outward, and thus the partner is motivated. Consequently, from Lemma 2, this team is efficient.

(b) One player is an other-sophisticate and the other player is a full-sophisticate. Expressing our notation from the other-sophisticate's perspective (without loss of generality), first note that $\mu_O(n) = \mu'_I(n) = \beta$ for all $n = 1, 2, \ldots$. Next, since beliefs are not symmetric, we must have that $\mu_I \neq \mu'_I = (\beta, \beta, \beta, \ldots)$. Therefore, the other-sophisticate's inward reassessment is finite. Consequently, the outward-sophisticate is motivated, from Proposition 3. In turn, the full-sophisticate must be unmotivated, from Proposition 4. Finally, from Lemma 2, this team must be efficient.

(c) One player is a self-sophisticate and the other is a full-sophisticate. The self-sophisticate is unmotivated, from Proposition 2. In turn, the full-sophisticate must be motivated, from Proposition 4. Finally, from Lemma 2, this team must be efficient.

Note that, since self-sophisticates do not reassess their naive outward beliefs, a team with two self-sophisticates must have symmetric beliefs with $\mu_I(n) = \mu'_I(n) = \mu_O(n) = \mu'_O(n) = 1$ for all $n = 1, 2, \ldots$. Therefore the possibility of a team with self-sophisticates did not need to be considered in the above proof. ■

A.13 Proof of Proposition 6

If both players are full-sophisticates, it is common knowledge that both players are present-biased. Therefore, $\mu_I = \mu'_I = \mu_O = \mu'_O = (\beta, \beta, \beta, \ldots)$. From Corollary 1, such a team must be inefficient.

Now the full-sophisticate's partner is a different type of sophisticate. From Proposition 5, this team will be efficient.

Next, suppose the full-sophisticate's partner is a naif with limited reassessment. From Proposition 1, the naif is unmotivated if their reassessment is weakly inward, and motivated if their reassessment is strictly outward. Proposition 4, the full-sophisticate is then motivated for the former type of naif and unmotivated for the latter, which in turn implies that the team will be efficient in light of Lemma 2.

Lastly, suppose the full-sophisticate's partner is a naif with unlimited reassessment. Under unlimited reassessment, the naif's beliefs are given by $\mu'_I = \mu'_O = (\beta, \beta, \beta, \ldots)$. Thus, $\mu'_I(n) = \beta = \mu_O(n)$ for all $n = 1, 2, \ldots$, i.e. the naif is (effectively) self-sophisticated. Moreover, $\mu'_O(n) = \beta = \mu_I(n-1)$ for all $n = 1, 2, \ldots$, i.e. the naif is (effectively) other-sophisticated. Therefore, the naif is, in effect, a full-sophisticate under unlimited reassessment (when paired with an actual full-sophisticate) because all of their beliefs happen to be correct. Finally, since the naif is neither motivated nor unmotivated under unlimited reassessment (Corollary 4), the team will be inefficient (Lemma 2), as desired. ■

A.14 Proof of Proposition 7

From Proposition 2, the self-sophisticate is unmotivated. From Proposition 1, the naif is motivated if and only if their reassessment is strictly outward. Therefore, from Lemma 2, the team is efficient if and only if the naif's reassessment is strictly outward. ■

A.15 Proof of Proposition 8

First consider case (i). Since the naif's reassessment is weakly inward, the naif is unmotivated (Proposition 1). The other-sophisticate is motivated if and only if their inward reassessment is weakly shallower than that of the naif (Proposition 3, part (ii)), which in turn holds if and only if the team is efficient (Lemma 2).

Now consider case (ii). Since the naif's reassessment is strictly outward, the naif is motivated (Proposition 1). The other-sophisticate is unmotivated if and only if their inward reassessment is strictly deeper than that of the naif (Proposition 3, part (ii)), which in turn holds if and only if the team is efficient (Lemma 2). ■

A.16 Robustness to Generalized Reassessments

Here we consider relaxing our assumptions on how naive beliefs are reassessed in two ways. First, we no longer assume that reassessed beliefs must be β, and instead allow a reassessed belief to assume any value on $[\beta, 1)$. This may reflect some form of partial reassessment of a particular belief in that it may be reassessed downward from its prior belief (1), but not necessarily down to the level of the naif's recently-discovered present bias factor (β) which prompted the reassessment. Second, we relax our previous assumption that beliefs that are lower in order than the highest-order reassessments in its associated direction are also reassessed. That is, we now allow for the possibility that $\mu_z(k) = 1$ and $\mu_z(k') < 1$ with $k' > k$.

To establish the robustness of our results under these conditions, we first have to adapt our previous notation to the new setting: To start, we define, for each $z \in \{O, I\}$, a player's reassessment depths as follows:

$$R_z^* = \begin{cases} 0, & \text{if } \mu_z(n) = 1, \text{ for all } n = 1, 2, \ldots \\ \max\{n : \mu_z(n) < 1\}, & \text{otherwise.} \end{cases}$$

We presume that this generalization pertains to finite reassessments for which $R_z^* < \infty$, while maintaining our previous notion for cases of infinite reassessment along a dimension (i.e. $\mu_z = (\beta, \beta, \beta, \ldots)$) to avoid having to work out arbitrarily complicated yet uninstructive cases (such as $\mu_z(n) = 1$ if and only if n is prime).

Similar to our previous definitions, we let ${R_O^*}'$ and ${R_I^*}'$ denote the partner's reassessment depths and $\widetilde{R^*}'_O$ and $\widetilde{R^*}'_I$ denote the original player's beliefs regarding ${R_O^*}'$ and ${R_I^*}'$. Now define $f(x, y) \equiv \{(\widetilde{R^*}'_O, \widetilde{R^*}'_I) : R_O^* = x, R_I^* = y\}$. Thus, we can see that $f(R_O^*, R_I^*) = (\widetilde{R^*}'_O, \widetilde{R^*}'_I)$. Lastly, define $k^*(\mu) = \min\{k : f^{(k)}(R_O^*, R_I^*) = (0, 0)\}$. We already know that if $k^*(\mu) = 0$ — i.e. under zero reassessment so that $R_O^* = R_I^* = 0$ — then the player is unmotivated. Now suppose a player with beliefs μ^A is unmotivated (motivated) and $k^*(\mu^A) = n$. Then, if $k^*(\mu^B) = n + 1$, a player with beliefs μ^B must be motivated (unmotivated) because they believe their partner is unmotivated (motivated). Therefore, a player with beliefs μ must be unmotivated if $k^*(\mu)$ is even and motivated if $k^*(\mu)$ is odd. Now $k^*(\mu)$ is even if and only if $R_I^* \geq R_O^*$, and $k^*(\mu)$ is odd if and only if $R_O^* > R_I^*$. Hence, a naif who is weakly inward in the sense that $R_I^* \geq R_O^*$ remains unmotivated under our relaxed assumption on reassessments, while a naif who is strictly outward in the sense that $R_O^* > R_I^*$ remains motivated, which means that Proposition 1 is qualitatively robust to our generalizations. It is straightforward to confirm that the corresponding corollaries follow in turn.

References

[1] Akerlof, George, "Procrastination and Obedience," *American Economic Review: Papers and Proceedings*, 81 (1991), 1–19.

[2] Akin, Zafer, "Time Inconsistency and Learning In Bargaining Games," *International Journal of Game Theory*, 36 (2007), 275–299.

[3] Bagnoli, Mark and Barton Lipman, "Provision of Public Goods: Fully Implementing the Core through Private Contributions," *Review of Economic Studies*, 56 (1989), 583–601.

[4] Becker, Gary and Kevin Murphy, "The Division of Labor, Coordination Costs, and Knowledge," *Quarterly Journal of Economics*, 107 (1992), 1137–1160.

[5] Benabou, Roland, "Groupthink: Collective Delusions in Organizations and Markets," *Review of Economic Studies*, 80 (2013), 429–462.

[6] Benabou, Roland and Jean Tirole, "Self-Confidence and Personal Motivation," *Quarterly Journal of Economics*, 117 (2002), 871–915.

[7] Blume, Lawrence, Adam Brandenburger, and Eddie Dekel, "Lexicographic Probabilities and Equilibrium Refinements," *Econometrica*, 59 (1991), 81–98.

[8] Bonatti, Alessandro and Johannes Horner, "Collaborating,"*American Economic Review*, 101 (2011), 632-663.

[9] Brandts, Jordi, and David Cooper, "A Change Would Do You Good... An Experimental Study on How to Overcome Coordination Failure in Organizations," *American Economic Review*, 96 (2006), 669–693.

[10] Brekke, Kjell Arne, Snorre Kverndokk, and Karine Nyborg, "An Economic Model of Moral Motivation," *Journal of Public Economics*, 87 (2003), 1967–1983.

[11] Brocas, Isabelle and Juan Carrillo, "Rush and Procrastination Under Hyperbolic Discounting and Interdependent Activities," *Journal of Risk and Uncertainty*, 22 (2001), 141–164.

[12] Camerer, Colin, Teck-Hua Ho, and Juin-Kuan Chong, "A Cognitive Hierarchy Model of Games," *Quarterly Journal of Economics*, 119 (2004), 861–898.

[13] Costa-Gomes, Miguel and Vincent Crawford, "Cognition and Behavior in Two-Person Guessing Games: An Experimental Study," *American Economic Review*, 96 (2006), 1737–1768.

[14] Costa-Gomes, Miguel, Vincent Crawford, and Bruno Broseta, "Cognition and Behavior in Normal-Form Games: An Experimental Study," *Econometrica*, 69 (2001), 1193–1235.

[15] Crawford, Vincent, Miguel Costa-Gomes, and Nagore Iriberri, "Structural Models of Nonequilibrium Strategic Thinking: Theory, Evidence, and Applications," *Journal of Economic Literature*, 51 (2013), 5–62.

[16] Darley, John, and Bibb Latane, "Bystander Intervention in Emergencies: Diffusion of Responsibility.," *Journal of Personality and Social Psychology*, 8 (1968), 377–383.

[17] Diekmann, Andreas, "Volunteer's Dilemma," *Journal of Conflict Resolution*, 29 (1985), 605–610.

[18] Diekmann, Andreas, "Cooperation in an Asymmetric Volunteer's Dilemma Game Theory and Experimental Evidence," *International Journal of Game Theory*, 22 (1993), 75–85.

[19] Eyal, Tal, Nira Liberman, and Taacov Trope, "Judging Near and Distant Virtue and Vice," *Journal of Experimental Social Psychology*, 44 (2008), 1204–1209.

[20] Franzen, Axel, "The Volunteers Dilemma: Theoretical Models and Empirical Evidence," in Foddy, Smithson, Schneider, and Hogg, eds. *Resolving Social Dilemmas: Dynamics, Structural, and Intergroup Aspects*, Philadelphia, 135-148 (1999).

[21] Gervais, Simon and Itay Goldstein, "The Positive Effects of Biased Self-Perceptions in Firms," *Review of Finance*, 11 (2007), 453–496.

[22] Haan, Marco and Dominic Hauck, "Games with Possibly Naive Hyperbolic Discounters," working paper (2014).

[23] Holmstrom, Bengt, "Moral Hazard in Teams," *Bell Journal of Economics*, 13 (1982), 324–340.

[24] Karni, Edi and Marie-Louise Vierø, "'Reverse Bayesianism': A Choice-Based Theory of Growing Awareness," *American Economic Review*, 103 (2013), 2790–2810.

[25] Koszegi, Botond, "Ego Utility, Overconfidence, and Task Choice," *Journal of the European Economic Association*, 4 (2006), 673–707.

[26] Laibson, David, "Golden Eggs and Hyperbolic Discounting," *Quarterly Journal of Economics*, 112 (1997), 443–477.

[27] Liberman, Nira and Trope, Yaacov, "The Role of Feasibility and Desirability Considerations in Near and Distant Future Decisions: A Test of Temporal Construal Theory.," *Journal of Personality and Social Psychology*, 75 (1998), 5–18.

[28] Liberman, Nira and Trope, Yaacov, "Temporal Construal," *Psychological Review*, 110 (2003), 403–421.

[29] Mannix, Elizabeth and Margaret Neale, "What Differences Make a Difference? The Promise and Reality of Diverse Teams in Organizations," *Psychological Science in the Public Interest*, 6 (2005), 31–55.

[30] Myerson, Roger, "Multistage Games with Communication," *Econometrica*, 54 (1986), 323–358.

[31] Nagel, Rosemarie, "Unraveling in Guessing Games: An Experimental Study," *American Economic Review*, 85 (1995), 1313–1326.

[32] O'Donoghue, Ted and Matthew Rabin, "Doing It Now or Later," *American Economic Review*, 89 (1999), 103–124.

[33] O'Donoghue, Ted and Matthew Rabin, "Choice and Procrastination," *Quarterly Journal of Economics*, 116 (2001), 121–160.

[34] O'Donoghue, Ted and Matthew Rabin, "Procrastination on Long-Term Projects," *Journal of Economic Behavior and Organization*, 66 (2008), 161–175.

[35] Ortoleva, Pietro, "Modeling the Change of Paradigm: Non-Bayesian Reactions to Unexpected News," *American Economic Review*, 102 (2012), 2410–2436.

[36] Rotella, Katie and Jennifer Richeson, "Motivated to 'Forget' The Effects of In-Group Wrongdoing on Memory and Collective Guilt," *Social Psychological and Personality Science*, 4 (2013), 730–737.

[37] Sarafidis, Yianis, "Games With Time Inconsistent Players," working paper (2006).

[38] Stahl, Dale and Paul Wilson, "On Players' Models of Other Players: Theory and Experimental Evidence," *Games and Economic Behavior*, 10 (1995), 218–254.

[39] Tenbrunsel, Ann, Kristina Diekmann, Kimberly Wade-Benzoni, and Max Bazerman, "The Ethical Mirage: A Temporal Explanation as to Why We Are Not as Ethical as We Think We Are," *Research in Organizational Behavior*, 30 (2010), 153–173.

[40] Wittenbaum, Gwen, Sandra Vaughan, and Garold Strasser, "Coordination in Task-Performing Groups," *Theory and Research on Small Groups*, (2002), 177–204.