ECONtribute
MARKETS & PUBLIC POLICY

ECONtribute
Discussion Paper No. 213

Motivated Memory in Economics - a Review

Andrea Amelio Florian Zimmermann

December 2022 www.econtribute.de

UNIVERSITÄT BONN Universität zu Köln

Funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2126/1-390838866 is gratefully acknowledged.

Motivated Memory in Economics - a Review*

Andrea Amelio Florian Zimmermann

November 2022

Abstract

In this article, we review the economics literature on the motivated recall of information. Summarizing both theoretical and empirical work, we highlight key results this literature has produced. We also discuss methodological issues when studying motivated memory. We conclude by highlighting open questions and possibly exciting avenues for future research.

1 Introduction

The standard economics view on belief formation is that decision-makers strive to hold beliefs that are as accurate as possible given the decision environment. In contrast, the notion of motivated reasoning posits that decision-makers *want to* hold certain beliefs about themselves or the world (e.g., Kunda, 1990; Bénabou & Tirole, 2006b, 2011, 2016; Köszegi, 2006). Motivated reasoning can give rise to overconfidence (e.g., Malmendier & Tate, 2005, 2008) and partisan polarization (Kahan, 2013), and can facilitate moral transgression at the individual (e.g., Babcock et al., 2013; Konow, 2000; Dana et al., 2007; Haisley & Weber, 2010; Gneezy et al.; Di Tella et al., 2015) and collective (Bénabou,

*Review article prepared for special issue in *Games* on "Economis of Motivated Beliefs". For valuable comments and discussions we thank Katrin Goedker, Maximilian Mueller, Charlotte Saucet, Paul Smeets and Marie Claire Villeval. Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2126/1 – 390838866. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (ERC Starting Grant: 948424 — MEMEB).

2013) level.

A central question in the literature on motivated beliefs is how self-deception operates in the presence of feedback. To put it differently, how do people manage to preserve self-servingly biased beliefs in situations where unbiased information is available. Bénabou & Tirole (2006b) have termed the cognitive strategies that people employ to deceive themselves the "supply side" of motivated beliefs.[1] Perhaps the most obvious strategy to maintain a positive self-view is to avoid potentially unpleasant information altogether (e.g., Dana et al., 2007; Golman et al., 2017). When information cannot be avoided, individuals might engage in asymmetric information processing, placing more weight on positive than negative signals (e.g., Eil & Rao, 2011; Sharot, 2011; Möbius et al., forthcoming). Yet another supply side mechanism emphasizes the role of motivated memory (Bénabou & Tirole, 2002). Motivated memory can be defined as the selective retrieval of past experiences or information based on self-serving criteria.

In this article, we aim to review and assess the economics literature on motivated memory. We begin by outlining a general theoretical framework, based on Bénabou & Tirole's (2002) seminal work, to provide a formal illustration of the main intuitions in this literature. Relatedly, we review a set of theoretical works featuring a multi-selves agent with a beliefs-dependent utility who can (costly) manipulate her memory and does so strategically. We then focus on the empirical work on motivated memory in economics. This literature is partially aimed at testing theoretical predictions but also studies different shades of motivated memory mechanisms. We first consider lab evidence on motivated memory. This evidence ranges from prosocial decision-making to ego-relevant domains such as IQ to investment behavior. In a second step we then review field evidence on selective recall. Having summarized the existing empirical literature, we then take a step

[1] In turn, the "demand side" (Bénabou & Tirole, 2006b) of self-deception is concerned with *why* people might want to hold motivated beliefs. Several motives have been suggested. Köszegi (2006) and Brunnermeier & Parker (2005) formalize decision-makers who derive consumption utility from being optimistic about themselves and/or the future. Bénabou & Tirole (2002) emphasize the motivational value of optimistic beliefs. Schwardmann & van der Weele (2019) argued that people first need to deceive themselves in order to deceive others. More generally, the literature on social signaling emphasizes the role of (stated) optimism as a social signal (see Burks et al., Charness et al., 2018; Ewers & Zimmermann, 2015.)

back and (i) try to identify main takeaways from the literature and (ii) discuss methodological considerations related to the study of motivated recall. In the last section we conclude by highlighting some open research questions.

2 Theoretical Framework

A key challenge for the concept of motivated reasoning is how self-deception is even logically possible. After all, a decision-maker who tries to manipulate his views and beliefs needs to take action with the goal of self-deception. Hence, one might argue that self-deception is an impossibility, since the decision-maker must know why he or she took certain actions. Philosophers and psychologists have debated this question for decades. In this section, we present a theoretical framework based on Bénabou & Tirole's (2002) seminal paper, which is the first work in economics to present formally the concept of motivated memory. This framework formalizes how self-deception via selective recall can be effective, even if the decision-maker knows that he or she sometimes purposefully forgets things, providing a conceptual structure to approach and interpret motivated memory.

Consider a present-biased agent with imperfect memory, facing a decision problem such as picking a level of effort $e \in \mathbb{R}^+$ to exert in some task or project. Exerting effort is costly, with costs described by the cost function $c(e)$, and the agent's returns on her effort $V(e, \theta)$ are a function of her personal *skill* or *abilities* $\theta \in \Theta$. [2] The time horizon of the decision problem is split into three key moments, that is $t \in \{0, 1, 2\}$. At $t = 0$, the agent can observe a noisy signal $\sigma \in \Sigma$ about her skill level θ. The agent knows the signal generating process, that is she is aware of $p(\sigma \mid \theta)$ for each σ and θ. At $t = 1$, the agent decides how much effort to exert in the task, incurring in the related cost, and at $t = 2$ the utility from the exerted effort realizes.

The agent's memory is imperfect, in that at $t = 1$ she retrieves the signal $\hat{\sigma}$ from her memory: with probability λ, σ is forgotten, meaning no information is recalled, despite

[2] In Bénabou & Tirole's (2002) paper, the skill parameter θ is more specifically interpreted as the probability of success of the project. This is just a normalization of any other dense set $\Theta \subset \mathbb{R}$.

having observed some signal about θ in the previous period. Formally,

$$\hat{\sigma} = \begin{cases} \sigma, & \lambda \\ \emptyset, & 1-\lambda, \end{cases}$$

with $\emptyset$ representing the case of no signal being retrieved from memory. Additionally, it assumed that the agent at $t = 0$, or Self 0, may costly manipulate the probability of forgetting the signal λ. The incentive to manipulate memory stems from the fact that the agent's preferences are dynamically inconsistent. Formally, the agent's utility at $t = 0$ is

$$U_0(e) = c(e) + \mathbb{E}_{\sigma}[V_0(e, \theta)],$$

while her utility from $t = 1$ perspective is

$$U_1(e) = c(e) + \mathbb{E}_{\hat{\sigma}}[V_1(e, \theta)].$$

Hence, for any given expectation about θ, the optimal effort from Self 0's perspective is different than that from Self 1's perspective. Moreover, it is assumed that $V_0(\cdot)$ and $V_1(\cdot)$ are such that:

$$e_0^*(\theta) = \arg\max_{e \in \mathbf{R}^+} V_0(e, \theta) > \arg\max_{e \in \mathbf{R}^+} V_1(e, \theta) = e_1^*(\theta),$$

for any $\theta \in \Theta$. Hence, Self 0 will have an incentive to increase the recall probability of *positive* signals and decrease the one of *negative* signals. Self 0's strategy can then be described as a mapping assigning a recall probability λ to any possible signal realization σ, that is $\lambda : \Sigma \to [0, 1]$. When Self 1 recalls the signal, that is $\hat{\sigma} = \sigma$, her beliefs will just follow from combining her prior about θ and $p(\sigma \mid \theta)$ for each θ through Bayes' rule. On the other hand, when observing $\emptyset$, Self 1 may still update her beliefs. This is because Self 1 knows that she may have forgotten the signal and that $\sigma \neq \emptyset$. Moreover, she is aware of Self 0's incentives, and will not necessarily take $\hat{\sigma} = \emptyset$ at face value.

Given this framework, the papers we discuss in the next section study the existence and features of Bayes-Nash equilibria and how the set of existing equilibria changes with

problem parameters. Of particular interest are *corner* equilibria, namely *full-disclosure* and *babbling* or *self-deception* equilibria, in which, respectively, Self 0 always and never truthfully reports negative signals.

2.1 Strategic Memory Models

The economics literature on both motivated reasoning and motivated memory was pioneered by Bénabou & Tirole (2002) and Bénabou & Tirole (2004). The authors present a two-selves model with strategic memory management, like the one presented in a general fashion in the previous section. The agent faces an investment decision and is affected by present bias. Self 0 holds private information about her skill level, which increases the investment returns, and can send a coarse message to Self 1, who makes the investment decision. The signal space is binary, with a bad signal (B) and the absence thereof ($\emptyset$). The signal observed by Self 0 may be forgotten with some probability λ, which leads Self 1 to observe $\emptyset$, whatever the actual signal is. Self 0 may costly manipulate λ and has an incentive to do so when observing B, leading Self 1 to be overconfident in her skills, to contrast underinvestment caused by present bias. The authors use Perfect Bayesian Equilibrium as a solution concept and show a relationship between the existence of *self-deception equilibria* and the intensity of present bias and ego-utility. The equilibrium probability of misreporting a negative signal monotonically increases in present bias. Hence, a full-disclosure equilibrium exists for low levels of present bias.

Chew et al. (2020) extend Bénabou & Tirole's (2002) theoretical framework, enriching the signals space: Self 0 can observe and send also a good (G) signal about her skill level, in addition to a bad (B) one and the absence of signal ($\emptyset$). Hence, in addition to repressing negative memory (amnesia, using the authors' terminology), Self 0 can also send a good signal in absence of information (delusion). As opposed to Bénabou & Tirole (2002), in Chew et al. (2020) the probability of misreporting the signal is not strictly increasing in the level of present bias, as the presence of delusion decreases the value of amnesia, given that present bias is large enough to sustain a delusion equilibrium. Moreover, Chew et al. (2020) show that self-deception equilibria always exist if the ego-dependent component

of utility is large enough, implying that when beliefs have a strong impact on utility, self-deception is sustainable also for agents who do not suffer from present bias.

Hagenbach & Koessler (2022) also present a setting related to Bénabou & Tirole's (2002) framework, but focus on studying how different classes of *psychological utilities* affect the existence of a full-disclosure equilibrium. Moreover, the decision-maker depicted in Hagenbach & Koessler (2022) does not suffer from present bias, hence no dynamic inconsistency issues arise: the trade-off between perfect memory and forgetting arises because of belief-based utility, as in Chew et al. (2020). Finally, the authors extend their model by adding exogenous memory failures: with some probability, Self 1 will not be able to remember the state of the world, even though Self 0 decided to recall. It is shown how the absence of full-disclosure equilibrium is generally possible in this setting, and under which additional assumptions such equilibrium exists for the classes of utilities considered.

Fudenberg et al. (2022) theoretically study imperfect memory in a different framework from Bénabou & Tirole (2002). Fudenberg et al. (2022) illustrate a model featuring an agent who only recalls a subset of past experiences and, crucially, study the dynamics and long-run behavior of the agent's beliefs and actions. The recalled experiences depend on a collection of memory functions and on the currently observed information. The agent is naive about her imperfect memory, updating beliefs and selecting actions as if the recalled experiences are the only ones that occurred. Fudenberg et al. (2022) show how in their framework it is possible to generate, among other biases, motivated memory, and how, in this setting, the agent's long-run beliefs will be skewed in an ego-boosting direction and hence her actions will be also sub-optimal and biased that direction. Additionally, the authors note how an agent with a misspecified model is observationally equivalent to an agent with ego-boosting memory, but how their memory-based framework predicts a different impact of changes in information provision. For example, unlike the misspecification case, their framework predicts that providing jointly positive and negative information about different, unrelated, and ego-relevant dimensions, increases the likelihood that the agent retrieves that specific experience, inducing the long-term beliefs about the ego-decreasing dimension to be more precise.

3 Empirical Evidence

In the following, we review the empirical evidence on motivated memory. The bulk of existing evidence was obtained in controlled lab studies. We hence begin by reviewing those, separately for different decision contexts. Table 1 provides a summary of several methodological aspects of all the empirical papers reported in this section, both lab and field studies.

3.1 Motivated Memory and Prosocial Behavior

A first block of evidence in the literature explores the role of motivated memory in the domain of prosocial behavior. The underlying intuition is that most people want to think of themselves as following high moral standards. At the same time, people sometimes succumb to moral transgressions. This creates a tension between desired self-view and actual behavior, which can be resolved by the motivated recall of past prosocial behavior.

Li (2013): The first paper that investigates empirically the role of motivated memory in the economics literature is Li (2013). The article investigates experimentally how the recall accuracy of social interactions depends on a player's *own* behavior. Additionally, it investigates the dynamics of how actions by *others* are perceived over time.[3]

The experiment is divided into two phases: i) a trust game and ii) a recall phase. Participants in the first phase are not aware that they will take part in the second phase. The first mover in the game can choose to *not trust* (end the game) or to *trust* (allow the other player to play). Given that player 1 trusted her, player 2 can be *kind* (split the pay-off evenly) or *not kind* (keep all the pay-off). There were three main treatments, manipulating the delay between the first and the second phase: the recall phase could take place immediately after the first phase, after seven days, or after 43 days. In the recall phase, participants are asked to recall their and others' decisions in the game. For all but a control treatment, the recall is financially incentivized, granting participants a

[3]Recall of behavior of others might also be subject to motivated reasoning. People might want to believe that the world is just (Bénabou & Tirole, 2006b) and accordingly might have too rosy recollections of how they were treated in the past.

fixed amount of money for each correct answer. Moreover, participants rate the kindness of the recalled choices and are asked to guess the share of players playing each action. Finally, the two players played a dictator game.

The author finds that participants who received unkind acts are less likely to recall what the other player did and also to correctly recall their own actions. Participants who decided to be *not kind* in the first phase seem to change their perception of their action kindness level, as the delay increases. Relatedly, when there is no delay between the two phases of the experiment, participants' beliefs about the share of players picking each action are biased towards their experience (e.g. a player who received an unkind act tends to overestimate the share of players choosing to be *not kind*. However, this difference seems to decay over time below the significance threshold.

Carlson et al. (2020): Carlson et al. (2020) focus on motivated memory in dictator games and tackle the following questions. First, they investigate the direction and the extent of memory errors, using a continuous measure, the split of the pie in percentage points. Second, they study how the relationship between subjective fairness norms and actions impacts the extent and direction of memory errors. Finally, they study how active as opposed to forced choices differ in triggering motivated memory.

In total, the authors conducted five experiments, two in the lab and three online. In the first experiment, participants took five allocation decisions on how to split an amount of money between themselves and an anonymous partner. Afterward, they answered demographic questions and completed a numeracy test. Then, in an announced recall task, participants would be asked to report the average share of the pie assigned to the anonymous partner in the previous stage of the experiment. The other experiments are then implementing incremental changes to the first experiment. The second experiment is identical to the first, except that participants' fairness norms are elicited before the dictator game rounds. The third experiment features a different, more fine-grained, incentivization scheme for the recall phase: while in the first two experiments participants were rewarded if their recall was within 10% their actual average generosity, in the third

experiment the bonus was an increasing function of the accuracy, at the 1% level. The two final experiments are related and aimed at investigating the third question. In the fifth experiment, different from the previous ones, after the fairness norm elicitation, participants are informed that instead of deciding the splits themselves, the computer implements a randomly drawn choice from a participant of experiment number four.

The authors provide evidence that on average participants commit self-serving memory errors. Recalled generosity tends to be biased upwards. Additionally, the authors show that fairness norms matter in predicting motivated memory errors. Independently of the actual generosity level, violators (participants who were less generous than their reported norm) have positive memory errors on average, while upholders (participants who on average were more generous than the reported norm) do not. Finally, concerning the role of responsibility, participants' self-serving memory errors seem to fade out when choices were forced and hence real responsibility is absent.

Saucet and Villeval (2019): Saucet & Villeval (2019) focus on the role of responsibility for motivated memory distortions. They do so in the context of dictator games. Furthermore, they study how financial incentives and the level of prosociality[4] impact recall at the intensive margin: the object of the recall exercise is not binary, which allows measuring how these parameters modulate the magnitude and the direction of recall errors.

The experiment is split into 4 parts: i) 12 binary dictator games, ii) a filler task, iii) a recall task and iv) a memory assessment task. In the baseline condition, participants are assigned to the roles of dictator or receiver. In a control condition, the dictator's role is played by a computer. Each dictator game has a *selfish* and an *altruistic* option. Option A is fairer than option B if the receiver's pay-off for A is higher than for B, regardless of how the pie is split. During the recall task, which is not announced beforehand, participants are asked to recall the amount allocated to the receiver (or to themselves in

[4]The level of prosociality is captured in *absolute* terms. This differs from Carlson et al. (2020), who investigate the role of personal fairness norms. In contrast to Saucet & Villeval's (2019) results on prosociality, Carlson et al. (2020) find no effect of absolute fairness on recall.

one treatment). Unlike Li (2013), Saucet & Villeval (2019) do not vary the delay between the dictator games and the recall task. Instead, they make all participants undergo the same filler task for eight minutes.

Dictators are found to have a higher recall accuracy when they picked the fair option. Interestingly, this difference is not present when allocations are made by the computer. The effect also does not seem to be explained or moderated by participants' education level or performance in an additional, unrelated, memory task. This is an important step towards a causal interpretation where past behavior is indeed the driver of recall accuracy. We discuss this further at the end of this section. Moreover, the authors provide evidence that partially supports the hypothesis of incentives playing a role in determining recall accuracy: incentives significantly increase the share of correctly recalled amounts by dictators, but only for those who picked the fair option. Finally, the authors do not find evidence for biased recollections of memory, beyond the selective nature of recall.

Galeotti et al. (2020): Galeotti et al. (2020) focus on a different aspect, namely mechanisms behind incorrect memories following antisocial behavior. The authors argue that unethical amnesia, defined as imprecise memory of morally dubious past actions, is due to instrumental reasons rather than self-image concerns. In other words, they hypothesize that faulty memory, in their setting, is caused by the need to justify future immoral decisions as opposed on top of the preservation of self-image.

This is shown through an experiment divided into two stages. First, participants undergo twenty sessions of a *wheel game*. Afterwards they take part in a recall stage. The time elapsed between the first and the second stage is three weeks. The wheel game is structured as follows. Participants are shown a wheel divided into six sections. Their task is to pick in their mind one of those sections. After they picked, each section is associated with a number from 1 to 6. Participants have to report the number associated with the section they picked, and pay-off increases in the reported number. Hence, potentially, participants could always state to have picked the section of the wheel with the largest number and receive the largest possible pay-off. In the recall stage, participants have to

recall how many times they reported each outcome in the twenty wheel game sessions. To disentangle hedonic and instrumental memory errors, the authors present two treatments. In the *Instrumental* treatment, right before the memory stage starts, participants are told that at the end of the session they may give up part of their pay-off and are encouraged to do so if they cheated in the first stage. In the *Hedonic* treatment, instead, participants are not given this possibility. Hence, memory errors in the *Instrumental* treatment may also depend on the desire to perpetuate unethical behavior, while in the *Hedonic* treatment, any memory error should not derive from strategic considerations, as participants are not given the chance to change their pay-off. Additionally, for both *Hedonic* and *Instrumental* treatments, there is a corresponding control condition, which is identical to the respective treatment, except for the fact that participants can not cheat in the wheel game, as they are required to click on the screen, as opposed to picking the section of the wheel in their mind.

To illustrate the results, the authors provide a type classification of participants: participants in treatment conditions are classified as dishonest if the distribution of reported outcomes in the first part is right-skewed and significantly different from a uniform with a 3.5 mean. First, the authors report no significant difference in memory errors comparing honest participants from the treatment conditions and participants from control conditions. Comparing memory errors from the pooled control conditions with those from dishonest participants in *Intrumental* and in *Hedonic*, memory errors in *Intrumental* are significantly higher. No such effect is found for *Hedonic*. This provides direct evidence for the hypothesis that the main drivers of memory errors, in this framework, are instrumental motives.

Taken together, the bulk of existing evidence in the context of prosocial behavior suggests that people self-servingly recall prosocial acts better than antisocial ones. A challenge in this literature, however, is to obtain causal evidence. The key reason this concern arises is that actual moral behavior is of course endogenous and hence might relate to factors that correlate with recall accuracy. For instance, it is conceivable that more prosocial

people, in general, have a better memory[5]. Saucet & Villeval (2019) actually measure recall performance in an unrelated task and show that their results remain unaffected when controlling for recall performance, strengthening a causal interpretation of their results. As we will see in the next sections, a key focus in the domains of investment behavior and IQ is to establish causality in the relationship between feedback and recall accuracy.

3.2 Motivated Memory and Financial Decisions

A second decision context where motivated memory has been studied is financial investments. Intuitively, investors want to think of themselves as being good at what they do. At the same time, some investments fail, creating a motivation to selectively recall past investment outcomes.

Gödker et al. (2022): In a series of controlled lab experiments, Gödker et al. (2022) provide experimental evidence of systematic memory bias in the domain of financial decisions. In this context, the authors aim to relate memory, beliefs, and reinvestment choices.

The authors introduce the following paradigm. The experiment is split into four stages: i) first investment and observation of returns ii) memory elicitation iii) belief elicitation and iv) second investment choice.[6] In the first stage, participants allocate an initial endowment between a risky asset and a safe bond. The stock can be good or bad, with the good stock being more likely to generate positive returns. Afterward, participants observe a set of noisy signals about the stock quality and, with different degrees of delay, are asked to recall the returns realizations and to provide their assessment of the probability of the stock being good or bad. The delay between the first and the second stage of the experiment can be either zero or one week. This allows the authors to isolate

[5]There is mixed evidence on the relationship between prosocial behavior and IQ. For example, Han et al. (2012) report evidence in favor of this relation only for tasks with higher complexity. Guo et al. (2019), instead, report a significant relationship between IQ and self-report measures of pro-social behavior.

[6]Points (ii) to (iv) were elicited in random order.

the effect of memory from the role of any other factor that may impact the recollection through encoding (e.g. attention, salience). The authors argue that any difference in recall performance between the conditions with different delays is to be attributed to memory, as any other factor which may influence information acquisition or processing is held constant.

Participants who decide to invest in the stock are found to disproportionately recall positive returns and, as a consequence, hold more optimistic beliefs about the stock being good. Moreover, the level of memory bias exhibited by participants predicts their probability of making an incorrect re-investment decision in the third phase of the experiment, showing how memory bias is connected not only to beliefs but also to investment choices. The authors also study how stake size has an impact on the size of the memory distortion and show that memory bias disappears on average when the incentives for accurate recall are substantially increased.

3.3 Motivated Memory and IQ

Another context where motivated recall has been studied is IQ test performance. Most people like to think of themselves as being smart, in particular relative to others. Hence, the presence of unbiased feedback about IQ test performance might create a motive to remember good news and misremember bad ones.

Chew et al. (2020): In a first attempt to empirically study memory about IQ-related feedback, Chew et al. (2020) propose and implement an experiment that closely follows the theoretical framework they present in the paper (see section 2.1 of this review). The goal of the experiment is to document three memory error types i) positive amnesia (not remembering a bad signal) ii) delusion (remembering a signal that was never observed) and iii) positive confabulation (remembering a good signal after observing a bad one) and relate these errors to the degree of present bias exhibited by participants.

The main task in the experiment is Raven's IQ test, with an additional recall phase. The delay between the two stages of the experiments is stated to be "several months",

which suggests that different participants experienced different delays in this order of magnitude. In the recall phase, participants are shown six questions and the related correct answers and may choose one of the following: a) *My response was correct* b) *My response was incorrect* c) *I didn't see this question* d) *I don't remember*. Four out of the six questions were drawn from the actual test, while two were not present in it.

The results show that participants tend to exhibit positive rather than negative memory errors. In other words, positive amnesia, delusion, and confabulation are relatively more frequent than the negative corresponding phenomena and are relevant in terms of incidence, ranging from 25% to 50%. Additionally, it is shown that the likelihood of exhibiting positive false memories increases in present bias, while this relationship does not hold for positive amnesia.

Zimmermann (2020): Zimmermann (2020) takes a different approach and studies the dynamics of updating from feedback about IQ test performance. He tries to map out how immediate updating differs from updating in the long run. In a second step, he then asks how selective recall can explain the dynamics of motivated beliefs, and what the limits of motivated memory are.

In the experiment, participants take an IQ-related test. Participants then state prior beliefs about their relative test performance. They then receive noisy feedback on relative test performance. The noise component in feedback is crucial, as it creates exogenous variation in whether subjects obtained good or bad news, hence allowing for the identification of a causal effect of feedback on belief dynamics and recall accuracy. In a between-subject design, participants then (unannounced) state posterior beliefs, either immediately after obtaining the feedback or 1 month later. In a separate treatment, instead of stating posterior beliefs 1 month after the feedback, subjects are asked to recall the feedback. In additional treatment variations (i) the belief elicitation 1 month later is announced (ii) incentives for recall accuracy are substantially increased.

Zimmermann (2020) finds that while in the short-run beliefs respond to both positive and negative feedback, over time beliefs after negative feedback move back to prior levels.

Positive feedback instead has a more long-lasting effect. This dynamic pattern seems well-explained by selective recall. After 1 month, participants more accurately recall positive feedback compared to negative feedback. Importantly, these effects are substantially mitigated when (i) future belief elicitation is announced and (ii) incentives for recall are high.

3.4 Field Evidence on Motivated Memory

While the majority of existing studies on selective recall are lab-based, in the following we review the existing work that studies motivated memory in the field.

Huffman et al. (2022): Huffman et al. (2022) are the first to study the role of motivated memory in a field setting. The context of the study is managerial overconfidence and how it responds to feedback. More specifically, the authors study persistent overconfidence of in-store managers and whether such overconfidence can be explained through motivated memory.

They found a very interesting setting where managers have access to a ranking concerning their performances and are asked to: i) state a prediction of the quintile of their ranking for the following quarter and ii) recall their past ranking quintile.

Managers are shown to be overall overconfident. This overconfidence is persistent despite the existence of feedback. Managers also exhibit upward memory bias. While managers' memory is precise when recalling positive past performances, memory errors arise and increase as the rank decreases. This provides a possible explanation for the persistence of memory bias.

Müller (2022): Müller (2022) provides field evidence on motivated memory in the domain of fertility desires, combining a large Kenyan panel covering more than ten years with experimental methods.

He shows how participants' recollection of their past fertility desires is a function of their actual fertility rate. In particular, participants tend to overestimate their past

desires, as on average they had more children than stated ten years before. Additionally, financial incentives do not seem to improve memory performances for fertility rates, while they seem to have a significant impact on ego-neutral recall questions, pointing towards mechanisms of selective forgetting.

Müller (2022) also reports evidence concerning information avoidance in this domain. Survey participants are randomly assigned to different conditions in which they were to learn their past fertility desires at the end of the survey, with or without financial incentives to do so. The results show that financial incentives impact the take-up rate only for participants who did not exceed their past fertility desires: for participants with excess fertility rates financial incentives seem to have no significant effect.

Roy-Chowdhury (2022): Roy-Chowdhury (2022) provides additional field evidence on motivated memory in the domain of school performances and relates it to prior beliefs and preferences. The analysis is based on longitudinal data, collected biannually, of teenagers' memories of their mathematics grades.

Similarly to Huffman et al. (2022) and Müller (2022), the author shows how recall of past performances is imperfect, with memory errors being skewed towards ego-enhancing recalls: on average, students recall better grades than the correct ones, and students with worse grades exhibit on average more memory errors. Roy-Chowdhury (2022) additionally exploits the different lags between the observation of the actual grade by the children and the moment in which the survey was conducted each semester, that is the recall date. The Fall wave of the survey asked about grades that the children observed approximately five months before, compared to the three weeks of the Spring wave. This way, the author can perform a within-subject analysis, showing how the evidence on motivated memory is robust to within-individual estimation.

The paper also reports evidence of memory distortion impacting the dynamics of self-confidence levels. Self-confidence shows high degrees of autocorrelation, but also depends greatly on recalled grades, and not on actual grades, suggesting that memory distortion plays a role in feasibly keeping high self-confidence levels. The author interprets these dy-

namics in the light of self-persuasion motives: students distort their memory of their past grades to impact their self-confidence level, which in turn may impact future academic performance.

Finally, Roy-Chowdhury (2022) estimates a structural model which decomposes the probability of recalling a specific grade, given the actual grade, into two components, along the lines of Bénabou & Tirole (2002): the ego benefit of recalling a high grade and a cost associated to distorting memory. The parameters associated with these factors are significant, supporting the idea of strategic memory manipulation, with students with a stated preference in math, or higher self-confidence, exhibiting higher gains from memory distortions.

4 Discussion - Taking Stock

The literature on motivated memory is still in its infancy. Nonetheless, in the following we will highlight empirical patterns that have emerged and discuss both conceptual and methodological aspects related to the study of motivated memory:

1. Motivated memory has by now been found in a rather broad set of contexts, ranging from investment behavior to prosocial decision-making. What distinguishes these contexts is the object of and the reason for self-deception. This highlights that studies of motivated memory always jointly test the existence of motivated reasoning and selective recall as a supply-side mechanism. This provides an empirical challenge since an implicit assumption in empirical studies of selective recall is that subjects do have a motive for self-deception to begin with. The same challenge is of course also present in the study of other supply-side mechanisms. In fact, part of the reason why the literature on asymmetric updating has produced somewhat mixed results might be the joint testing of motivated reasoning plus asymmetry in updating.[7]

[7] Eil & Rao (2011) test how individuals acquire and process information in ego-relevant contexts, intelligence and beauty, showing evidence for asymmetric updating. Möbius et al. (forthcoming) report similar evidence of asymmetric updating and conservatism in the domain of beliefs concerning relative

2. Despite the pervasive evidence for motivated memory in different contexts, the literature has also identified important limits to selective recall. In particular, two patterns emerge: (i) substantial increases in stake size as in Gödker et al. (2022) and Zimmermann (2020) mitigate motivated memory and instead induce people to accurately remember both good and bad news. (ii) the announcement of a future recall task diminishes motivated memory, see Coffman et al. (2021) and Zimmermann (2020). In other words, when future recall tasks are anticipated, asymmetries in recall appear to vanish. In sum, the prevalence of motivated memory seems to depend heavily on features of the decision environment.

3. A methodological pattern that emerges from existing work on motivated memory is that the role of selective recall is identified via delay between action taken or feedback provided and subsequent memory elicitation. Two comments seem in order: (i) so far, no consensus has emerged on the delay that is necessary for selective recall to emerge. In most existing studies, delay ranges between one and several weeks (ii) implementing memory constraints via delay is different from the typical approach in studies of non-motivated recall, where memory constraints are made binding via interference (similar memory entries compete and interfere with each other, making retrieval of a specific memory cue less likely).[8]

4. A further methodological pattern seems to be that in most studies, recall comes as a surprise. This again is in contrast with studies of non-motivated memory, where typically the entire experiment, including future recall elicitation is explained in advance. Importantly, as discussed, motivated memory appears to be mitigated when future recall tasks are announced.

performance in an IQ test. Ertaç (2011) also reports evidence of asymmetric updating, but in the opposite direction: participants react more to bad news about their performance, compared to good news. Sharot (2011) show how people update their beliefs asymmetrically, reacting more to positive than negative news, relating this asymmetry to previous expectations and to variations in neural coding of positive and negative news. Schwardmann & van der Weele (2019) reproduce results concerning asymmetric updating in the domain of intelligence tests, showing that such asymmetry towards positive self-image is fostered by motives to persuade others. Coutts (2019) reports evidence of asymmetry and conservatism in belief updating also in financially relevant and not value-relevant domains, arguing that asymmetric updating in ego-relevant domains could also be not due to motivated reasoning. Barron (2021) investigates updating in a financial domain, reporting no significant evidence of asymmetric updating.

[8]See, for example, Enke et al. (2022) and Bordalo et al. (forthcoming).

5. One empirical challenge is to establish causality in selective recall. Studies that have attempted to identify causality predominantly exploited noisy feedback provision (Gödker et al. (2022) and Zimmermann (2020)), where participants obtain noisy feedback about their own test performance on investment success, and the noise component allows for a causal identification of selective recall. In action-based paradigms such as prosocial behavior, causality is more difficult to establish, but causal interpretations have been bolstered by adding controls for memory performance (Saucet & Villeval (2019)).

6. In terms of the measurement of memory, the literature has mostly used incentivized recall paradigms, where participants receive a reward if recall accuracy is high. In terms of question format, different approaches have been followed (similar to some of the leading approaches in memory research, see Kahana's (2012) book): (i) in recognition paradigms, respondents are asked to recall past behavior or feedback received and are provided with several possible answers. The task then is to pick the correct one; (ii) free recall paradigms are similar in that subjects are also asked to recall past behavior or feedback received. However, they are not provided with a list of possible answers, but need to recall freely; (iii) in open texts paradigms, instead of being asked to recall specific events (e.g. past action or feedback), subjects are asked to describe certain aspects of an experiment or a decision environment. All these formats can be incentivized.[9] Interestingly, as we discuss in more detail below, stake size has been found to matter in at least some studies.

5 Concluding Remarks

In this review, we summarize the still small but growing economics literature on motivated memory. While clear limits to selective recall appear to exist, the evidence suggests the prevalence of selective recall in a broad set of domains. In other words, motivated memory seems to be an important input into the development and maintenance of motivated belief

[9]Open text messages can be more difficult to incentivize. In Zimmermann (2020), subjects receive a reward if a group of RAs judges the text as an accurate description of the object of interest.

patterns.

In the following, we will try to summarize a few holes in the literature that might point towards promising directions for future research. First, until now, the literature on motivated and non-motivated essentially moved in parallel with essentially no overlap, both conceptually and in terms of methodology. Bringing these strands of literature together might be fruitful in many different ways. For instance, the literature on non-motivated recall has identified a series of key features of how memory operates, such as associativeness and interference (Bordalo et al. (2020); Enke et al. (2022)), and studied implications, for instance for the reaction to news or stereotypical thinking. It would be interesting if motivated agents were able to exploit these key patterns in memory to self-servingly bias what they recall. Second, as this review highlights, there is until now a clear dominance of lab-based work on selective recall. While the lab has many advantages for the study of memory, more field work will be needed to assess the importance and limits of motivated memory for economic applications. Third, the demand side of motivated memory remains partially unexplored. In this direction, we see two promising avenues. The first is to better understand what are the determinants of the demand side value for motivated memory, that is which factors impact the demand for motivated memory and its values. Carlson et al. (2020), Saucet & Villeval (2019) and Gödker et al. (2022) provide evidence on how responsibility or active choices may impact the values of the demand side, but there may be more relevant factors at play. The second is to study individual heterogeneity in demand value and how this impacts heterogeneity in the ability to repress unpleasant information.[10] Fourth and finally, until now, the bulk of empirical evidence is reduced form. In light of the existing theoretical work on motivated memory, it might be fruitful to follow more structural approaches that allow for more direct tests of model predictions and for the identification of model parameters.

[10]With the current paradigms, it would be difficult to distinguish the demand value of motivated memory and the ability to forget information. How to disentangle these two elements is an open empirical challenge.

	Delay	Announced Recall	Continuous Recall Object	Incentivized Recall	Causal Identification
Li (2013)	*Treatment dependent: 0, 7 and 43 days*	✗	✗	✓	✗
Carlson et al. (2018)	*Participant dependent: ∼5 minutes*	✗	✓	✓	✗
Chew et al. (2018)	*Months (not further specified)*	*Not specified*	✗	✓	✗
Saucet and Villleval (2019)	*8 minutes*	✗	✓	*Treatment dependent*: ✓/✗	✗
Galeotti et al. (2020)	*3 weeks*	✗	✓	✓	✗
Zimmermann (2020)	*Treatment dependent: No Delay or 1 month*	Treatment Dependent: ✓/✗	✓	✓	✓
Gödker et al. (2021)	*Treatment dependent: 0 and 1 week*	✗	✓	✓	✓
Huffman et al. (2021)	*3 months*	✗	✗	✓	✗
Müller (2021)	*10 years*	✗	✓	✓	✗
Roy-Chowdhury (2022)	*Treatment dependent: 5 months and 3 weeks*	✗	✗	✗	✗

Table 1: *Empirical papers summary table. The Delay column reports the delay between the elicitation of the object to be recalled and the recall task. Announced Recall states whether the recall task was announced to participants beforehand. Continuous Recall Object concerns the type of information that participants are asked to retrieve from memory: a binary information (e.g. the amount was fair/unfair) or a continuous one (e.g. a specific amount, a percentage). The Incentivized Recall column provides information about whether the memory task was financially incentivized. Finally, the Causal Identification column concerns the presence, or absence, of an exogenous manipulation of motivated behavior which allows for establishing causality by design.*

References

Babcock, L., Loewenstein, G., Issacharoff, S. & Camerer, C. (2013), 'Biased judgments of fairness in bargaining', American Economic Review **85**(5), 1337–43.

Bordalo, P., Conlon, J. J., Gennaioli, N., Kwon, S. Y. & Shleifer, A. (forthcoming), 'Memory and probability', The Quarterly Journal of Economics .

Bordalo, P., Gennaioli, N. & Shleifer, A. (2020), 'Memory, attention, and choice', The Quarterly Journal of Economics **135**(3), 1399–1442.

Brunnermeier, M. K. & Parker, J. A. (2005), 'Optimal expectations', American Economic Review **95**(4), 1092–1118.

Burks, S. V., Carpenter, J. P., Goette, L. & Rustichini, A. (2013), 'Overconfidence and social signalling', The Review of Economic Studies **80**(3), 949–983.

Bénabou, R. (2013), 'Groupthink: Collective delusions in organizations and markets', Review of Economic Studies **80**(2), 429–462.

Bénabou, R. & Tirole, J. (2002), 'Self-confidence and personal motivation', The Quarterly Journal of Economics **117**(3), 871–915.

Bénabou, R. & Tirole, J. (2004), 'Willpower and personal rules', Journal of Political Economy **112**(4), 848–886.

Bénabou, R. & Tirole, J. (2006a), 'Belief in a just world and redistributive politics', The Quarterly Journal of Economics **121**(2), 699–746.

Bénabou, R. & Tirole, J. (2006b), 'Incentives and prosocial behavior', American Economic Review **96**(5), 1652–1678.

Bénabou, R. & Tirole, J. (2011), 'Identity, morals, and taboos: Beliefs as assets', The Quarterly Journal of Economics **126**(2), 805–855.

Bénabou, R. & Tirole, J. (2016), 'Mindful economics: The production, consumption, and value of beliefs', Journal of Economic Perspectives **30**(3), 141–64.

Carlson, R. W., Maréchal, M. A., Oud, B., Fehr, E. & Crockett, M. J. (2020), 'Motivated misremembering of selfish decisions', Nature Communications **11**(1), 2100.

Charness, G., Rustichini, A. & van de Ven, J. (2018), 'Self-confidence and strategic behavior', Experimental Economics **21**, 72–98.

Chew, S. H., Huang, W. & Zhao, X. (2020), 'Motivated false memory', Journal of Political Economy **128**(10), 3913–3939.

Coffman, K. B., Ugalde Araya, P. & Zafar, B. (2021), A (dynamic) investigation of stereotypes, belief-updating, and behavior, Working Paper 29382, National Bureau of Economic Research.

Coutts, A. (2019), 'Good news and bad news are still news: experimental evidence on belief updating', Experimental Economics **22**(2), 369–395.

Dana, J., Kuang, J. X. & Weber, R. A. (2007), 'Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness', Economic Theory **33**, 67–80.

Di Tella, R., Perez-Truglia, R., Babino, A. & Sigman, M. (2015), 'Conveniently upset: Avoiding altruism by distorting beliefs about others' altruism', American Economic Review **105**(11), 3416–42.

Eil, D. & Rao, J. M. (2011), 'The good news-bad news effect: Asymmetric processing of objective information about yourself', American Economic Journal: Microeconomics **3**(2), 114–38.

Enke, B., Schwerter, F. & Zimmermann, F. (2022), Associative memory and belief formation, Working Paper 26664, National Bureau of Economic Research.

Ertaç, S. (2011), 'Does self-relevance affect information processing? experimental evidence on the response to performance and non-performance feedback', Journal of Economic Behavior Organization **80**(3), 532–545.

Ewers, M. & Zimmermann, F. (2015), 'Image and misreporting', Journal of the European Economic Association **13**(2), 363–380.

Fudenberg, D., Lanzani, G. & Strack, P. (2022), Selective memory equilibrium, Working paper, SSRN.

Galeotti, F., Saucet, C. & Villeval, M. C. (2020), 'Unethical amnesia responds more to instrumental than to hedonic motives', Proceedings of the National Academy of Sciences **117**(41), 25423–25428.

Gneezy, U., Saccardo, S., Serra-Garcia, M. & van Veldhuizen, R. (2020), 'Bribing the self', Games and Economic Behavior **120**, 311–324.

Golman, R., Hagmann, D. & Loewenstein, G. (2017), 'Information avoidance', Journal of Economic Literature **55**(1), 96–135.

Guo, Q., Sun, P., Cai, M., Zhang, X. & Song, K. (2019), 'Why are smarter individuals more prosocial? a study on the mediating roles of empathy and moral identity', Intelligence **75**, 1–8.

Gödker, K., Jiao, P. & Smeets, P. (2022), Investor memory, Working papers, SSRN.

Hagenbach, J. & Koessler, F. (2022), 'Selective memory of a psychological agent', European Economic Review **142**(2), 104012.

Haisley, E. C. & Weber, R. (2010), 'Self-serving interpretations of ambiguity in other-regarding behavior', Games and Economic Behavior **68**(2), 614–625.

Han, R., Shi, J., Yong, W. & Wang, W. (2012), 'Intelligence and prosocial behavior: Do smart children really act nice?', Current Psychology **31**.

Huffman, D., Raymond, C. & Shvets, J. (2022), 'Persistent overconfidence and biased memory: Evidence from managers', American Economic Review **112**(10), 3141–75.

Kahan, D. M. (2013), 'Ideology, motivated reasoning, and cognitive reflection', Judgment and Decision Making **8**(4), 407–424.

Kahana, M. J. (2012), Foundations of Human Memory, Oxford University Press.

Konow, J. (2000), 'Fair shares: Accountability and cognitive dissonance in allocation decisions', American Economic Review **90**(4), 1072–1091.

Kunda, Z. (1990), 'The case for motivated reasoning', Psychological Bulletin **108**(3), 480–498.

Köszegi, B. (2006), 'Ego utility, overconfidence, and task choice', Journal of the European Economic Association **4**(4), 673–707.

Li, K. K. (2013), 'Asymmetric memory recall of positive and negative events in social interactions', Experimental Economics **16**(3), 248–262.

Malmendier, U. & Tate, G. (2005), 'Ceo overconfidence and corporate investment', The Journal of Finance **60**(6), 2661–2700.

Malmendier, U. & Tate, G. (2008), 'Who makes acquisitions? ceo overconfidence and the market's reaction', Journal of Financial Economics **89**(1), 20–43.

Möbius, M., Niederle, M., Niehaus, P. & Rosenblat, T. (forthcoming), 'Managing self-confidence: Theory and experimental evidence', Management Science .

Müller, M. W. (2022), Selective memory around big life decisions, Working papers.

Roy-Chowdhury, V. (2022), Self-confidence and motivated memory loss: Evidence from schools, Cambridge Working Papers in Economics 2213, Faculty of Economics, University of Cambridge.

Saucet, C. & Villeval, M. C. (2019), 'Motivated memory in dictator games', Games and Economic Behavior **117**(C), 250–275.

Schwardmann, P. & van der Weele, J. (2019), 'Deception and self-deception', Nature Human Behavior **3**(10), 1055–1061.

Sharot, T. (2011), 'The optimism bias', Current Biology **21**(23), 941–945.

Zimmermann, F. (2020), 'The dynamics of motivated beliefs', American Economic Review **110**(2), 337–61.