(Dis)honesty in individual and collaborative settings

A behavioral ethics approach

Leib, M.

Publication date
2021

Document Version
Other version

License
Other

Link to publication

Citation for published version (APA):
Leib, M. (2021). *(Dis)honesty in individual and collaborative settings: A behavioral ethics approach*.

Chapter 2

Loss framing increases self-serving mistakes (but does not alter attention)

This chapter was published as Leib, M., Pittarello, A., Gordon-Hecker, T., Shalvi, S., & Roskes, M. (2019). Loss framing increases self-serving mistakes (but does not alter attention). *Journal of Experimental Social Psychology, 85*, 103880. https://doi.org/10.1016/j.jesp.2019.103880

Mistakes can be beneficial or harmful to those making them. For instance, when claiming costs associated with a business trip, one may accidentally submit a hotel receipt twice or completely forget to submit it. When motivated to be honest, people should be as likely to make a self-serving mistake and submit a receipt twice, as they are to make a self-hurting mistake and forget to submit a receipt altogether. When motivated by self-interest, however, self-serving mistakes are more likely to occur than self-hurting mistakes. Furthermore, in some organizations, employees pay from their own pockets and get reimbursed for the costs they have incurred after the trip. In other organizations, employees receive an allowance before the trip, and pay back money they have not used after the trip. Because receiving an allowance before the trip makes people feel that the money they do not claim is a loss (Thaler, 1980), especially in such settings people may be likely to make self-serving mistakes and submit their hotel receipt twice.

Recent work revealed that people violate rules and lie more to avoid or minimize their losses (Folmer & De Cremer, 2012; Grolleau, Kocher, & Sutan, 2016; Kern & Chugh, 2009; Schindler & Pfattheicher, 2017) compared to secure or maximize their gains. It is unclear, however, whether making self-serving (vs. hurting) mistakes will lead to a similar pattern of results. We define self-serving mistakes as judgments or behaviors that are misguided or wrong and benefit the self. Such mistakes are likely to occur in ambiguous settings, where the rules or the information presented is unclear. As such, those mistakes may be considered as a milder type of unethical behavior, compared with more blatant types like outright lying and cheating. Focusing on this understudied type of unethical behavior, the first question we tackle here is: Do people make more self-serving mistakes when doing so minimizes losses compared with maximizes gains?

Assessing the process underlying self-serving mistakes revealed that, in ambiguous settings, people allocate more attention to tempting (vs. non-tempting) information, which in turn shapes their self-serving mistakes (Pittarello, Leib, Gordon-Hecker, & Shalvi, 2015; and relatedly Bazerman & Sezer, 2016; Bazerman & Tenbrunsel, 2011; Chugh, Bazerman, & Banaji, 2005; Sezer, Gino, & Bazerman, 2015). The theoretical explanation to this finding is that people want to think of themselves as honest individuals while simultaneously benefit from ethical rule violations (Mazar, Amir, & Ariely, 2008). When they can justify their rule violations – for example, by claiming to not notice which piece of information was the one they should have used – they violate rules more than when such justifications are not available (Bassarak, Leib, Mischkowski, Strang, Glöckner, & Shalvi, 2017; Shalvi, Dana, Handgraaf, &

De Dreu, 2011; Shalvi, Gino, Barkan, & Ayal, 2015). The second question we tackle is therefore: Is the increase in self-serving mistakes when people attempt to minimize losses vs. maximize gains driven by an increased attention to tempting information?

Loss aversion and self-serving mistakes: The role of attention

Abundant evidence shows that people are loss averse. The pain from losing exceeds people's pleasure from gaining equal goods, and thus they are more inclined to take actions that prevent (or reduce) losses than actions that secure (or increase) gains (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001; Cameron & Miller, 2009; Kahneman, Knetsch, & Thaler, 1990; Kahneman & Tversky, 1979; Ritov, Baron, & Hershey, 1993; Schurr & Ritov, 2013; Tversky & Kahneman, 1981). As such, loss aversion drives people to engage in more goal-oriented behavior when they are facing a loss, compared to an equal sized gain (Rothman, Bartels, Wlaschin, & Salovey, 2006; Sherman, Mann, & Updegraff, 2006).

People violate ethical rules more to prevent or minimize losses than to secure or maximize gains (Folmer & De Cremer, 2012; Grolleau et al, 2016; Schindler & Pfattheicher, 2017; see related Van Yperen, Hamstra, & van der Klauw, 2011). For instance, Kern and Chugh (2009) had participants take the role of an entrepreneur interested in acquiring a business owned by a competitor. Participants indicated whether they would hire a consultant holding private inside information about the competitor's company that would ultimately help to complete the acquisition. Half the participants learned they would have a 25% chance of gaining the acquisition; the other half learned they would have a 75% chance of losing the acquisition. Although the two scenarios were identical in terms of the expected outcome, participants were more likely to hire the consultant when their chance of acquiring the company was framed as a potential loss compared to a gain. Clearly, a loss prospect affected people's hypothetical willingness to violate the rules. However, the effect of framing on a financially incentivized, more subtle type of unethical behavior remains an open question. Does seeking to minimize a loss, compared to maximize a gain, lead people to make more self-serving mistakes? And if so, are these self-serving mistakes driven by people's increased attention to tempting information in a loss than in a gain framing?

To assess attention, we focus on a well-established proxy, namely *visual* attention, in particular gaze dwell time – the time people spend looking at a particular piece of information (Glöckner, Fiedler, Hochman, Ayal, & Hilbig, 2012; Krajbich & Rangel, 2011; Mogg, Bradley, Field, & De Houwer, 2003; Papies, Stroebe, & Aarts, 2008; Raab & Johnson, 2007, Townshend & Duka, 2001). Prior work found that people pay more attention and look longer at tempting, goal relevant information (Mogg et al., 2003; Townshend & Duka, 2001; Vogt, De Houwer, Moors, Van Damme, & Crombez, 2010). Further, the longer the relative time people spent looking at a piece of information, the more weight they gave to that piece of information when making decisions (Fiedler, Glöckner, Nicklisch, & Dickert, 2013; Glöckner et al., 2012; Halevy & Chou, 2014; Hochman, Glöckner, Fiedler, & Ayal, 2016; Krajbich & Rangel, 2011).

Pittarello and colleagues (2015) revealed that longer time spent looking at tempting (vs. not tempting) information drives people's self-serving mistakes. Employing the ambiguous dice paradigm, over multiple trials participants saw a fixation cross on a computer screen followed by six die-roll outcomes. They were asked to report the value of the die-roll that appeared closest to the fixation cross and were paid based on their reports (higher reports corresponded to higher payoffs). Participants made more self-serving than self-hurting mistakes. Specifically, they were more likely to report the value second closest to the fixation cross when it was higher than the one closest to the fixation cross (i.e., a tempting alternative) compared to when it was lower than the one closest to the fixation cross (i.e., a non-tempting alternative). These mistakes were driven by participants' increased attention (i.e., longer gaze dwell) to the tempting alternative.

But how does a loss, compared to a gain framing, affect people's attention to tempting information? We consider three possibilities. The first possibility is that framing affects the attention people allocate to tempting information. Specifically, people look longer at tempting information in a loss, than in a gain framing. In turn, the longer time spend looking at tempting information leads to more self-serving mistakes in a loss (vs. a gain) framing. Supporting this possibility, research shows that framing affects people's psychological arousal and attention patterns. Facing a loss, compared to a gain, leads to increased heart rate and pupil dilation (Hochman & Yechiam, 2011, Löw, Lang, Smith, & Bradley, 2008; Satterthwaite et al., 2007). Further, losses increase the overall attention people dedicate to a task, decrease their random mistakes, and facilitate behaviors that are consistent with the task's (payoff) structure (Yechiam & Hochman, 2013a, 2013b). Since people look longer at goal relevant and tempting information (Vogt et al., 2010), and loss framing increases people's goal-oriented behavior (Rothman et al., 2006), it is plausible that people will look longer at tempting information in loss, compared to gain framing. If this is indeed the case, we should find an interaction between temptation (yes vs. no) and framing (gain vs. loss), predicting gaze dwell time. Gaze dwell time, in turn, should predict the likelihood of making self-serving mistakes. We label this the *loss increases attention to tempting information* hypothesis (H1a, see Figure 2.1).

The second possibility is that framing does not alter people's attention to tempting information, but rather people's likelihood to self-servingly use the tempting information they looked at. Supporting this possibility Müller, Rothermund, and Wentura (2016) asked people to look at stimuli on a computer screen. On each trial, a color indicated whether they could gain (or, lose) 20 points if they provided (or, failed to provide) a correct answer. When people observed colors associated with losses, they exhibited attentional bias (in terms of response time) similar to when they observed colors associated with gains (see also Wentura, Müller, & Rothermund, 2014). The authors conclude that "attention is biased towards both positive gain cues and negative loss cues" (p. 761), suggesting that a stimulus' valence (positive or negative) does not influence attention. Brosch, Sander, Pourtois, and Scherer (2008) reached similar conclusions when using brain activity as a proxy for attention. Taken together, these lines of work suggest framing should not alter the attention people pay to tempting information. If people make more self-serving mistakes in a loss compared to a gain framing, but framing does not increase the attention to tempting information, we should not find an interaction between temptation (yes vs. no) and framing (gain vs. loss) predicting gaze dwell time. Rather, we should find an interaction between gaze dwell time and framing predicting self-serving mistakes. We label this the *loss increases the use of tempting information* hypothesis (H1b, see Figure 2.1).

Finally, a third possibility is that framing affects both the attention people allocate to tempting information, and their use of this information. If so, we should find an interaction between temptation (yes vs. not) and framing (gain vs. loss) predicting gaze dwell time, as well as an interaction between gaze dwell time and framing predicting self-serving mistakes. We label this the *loss increases attention to and use of tempting information* hypothesis (H1c, see Figure 2.1).

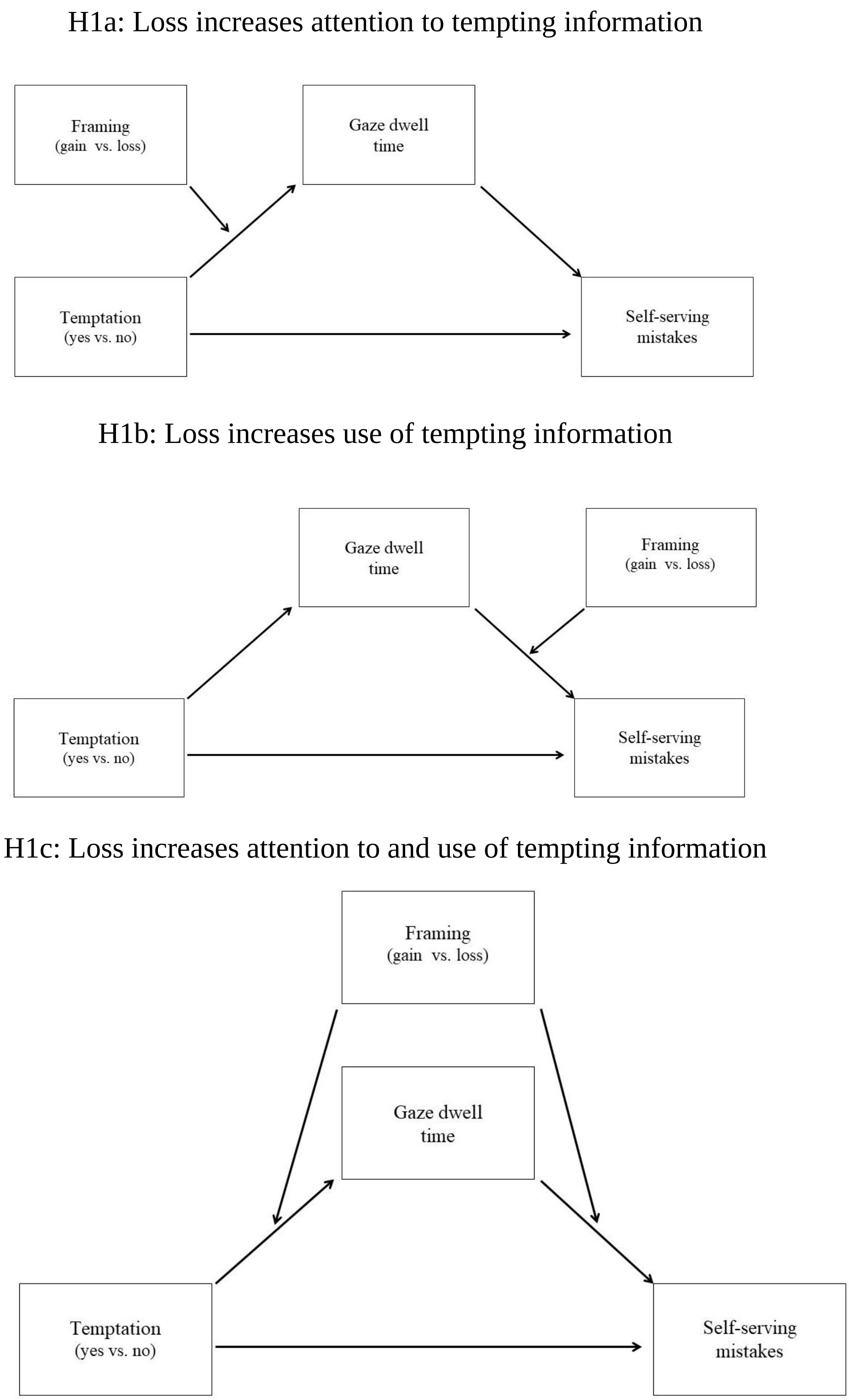

*Figure 2.1***.** Theoretical models: loss increases attention to tempting information (H1a), loss increases use of tempting information (H1b), and loss increase attention to, and use of tempting information (H1c) hypotheses.

Overview of experiments

We conducted two experiments employing the ambiguous dice paradigm (Pittarello et al., 2015). In both experiments, participants observed a fixation cross on a computer screen followed by six dice and were asked to report the value appearing closest to the fixation cross (with higher reports corresponding to higher payoffs). Across multiple trials, the value second closest to the fixation cross was either higher or lower than the one closest to the fixation cross. In the gain framing condition, participants started the task without a monetary endowment and earned money based on their reports. In the loss framing condition, participants started with an endowment and lost money based on their reports.

The goal of Experiment 2.1 was to test the effect of loss vs. gain framing on people's likelihood to make self-serving mistakes. We further added a control condition in which participants were paid based on their accuracy, rather than their reports. The control condition allowed us to examine whether people, for whatever reason, report higher values than the ones they are supposed to, even when there is no monetary incentive to do so. The goal of Experiment 2.2 was to test how attention is associated with self-serving mistakes in both loss and gain framings. To do so, we focused on the conditions in which participants were paid based on their reports and traced their eye movements during the task. We then fitted the three theoretical models (H1a, H1b, and H1c) to the obtained eye tracking data to determine which model has the best fit to the data. The task, data, and code for all analyses are available on Open Science Framework[1]. In the main text, we report how we determined our sample size, all data exclusions, all manipulations, and all measures.

Experiment 2.1

Method

One hundred and twenty university students (78.33% females, $M_{age} = 23.06$, $SD_{age} = 1.27$) engaged in the ambiguous dice paradigm in exchange for course credit. Upon arriving to the lab, participants learned they could earn an additional pay based on their performance. Thus, the incentive to sign up and participate in the experiment was fixed and unrelated to the condition participants were assigned to. The sample size was determined based on previous work. Since the ambiguous dice paradigm contains multiple trials, the data were analyzed

[1] This paper has Open Materials and Open Data badges for transparent practices. The materials are available at https://osf.io/473gq/. The data and code are available at https://osf.io/uk5c7/.

using a generalized linear mixed model that clusters trials at the individual level, treating each participant's reports as interdependent (Baayen, Davidson, & Bates, 2008). A priori power analysis for a generalized linear mixed model is not yet available on the G*power calculator. Thus, we followed Pittarello et al. (2015) who also used a between-subjects design using the same task and collected 30 participants per between-subjects cell. At the end of the results section of Experiment 2.2 we report a sensitivity analysis for the sample sizes used in Experiments 2.1 and 2.2 showing that our samples were sufficient to detect a medium to large effect size.

In the task, participants saw a fixation cross on the computer screen (1000 ms), followed by six die roll outcomes (2000 ms). After the die disappeared, participants were asked to report the outcome that appeared closest to the fixation cross (the target). Participants engaged in multiple trials, presented in a random order, and learned that at the end of the experiment, one trial would be randomly selected for payment. Between participants we manipulated the payoff structure and the framing of the task.

Participants in the pay-for-report conditions were incentivized according to the value they reported, with higher values leading to higher payoffs. In the pay-for-report, gain-framing condition (n = 30), participants started the task with 0 Israeli Shekels (ILS; 1 ILS = ~€0.25) and earned money according to the value they reported on a randomly selected trial, with reporting "1"= 5 ILS, "2"= 10 ILS, "3" = 15 ILS, "4" = 20 ILS, "5" = 25 ILS, and "6"= 30 ILS. In the pay-for-report, loss-framing condition (n = 30), participants started the task with 35 ILS and lost money based on their reports on one randomly selected trial, with reporting "1"= losing 30 ILS, "2"= losing 25 ILS, "3" = losing 20 ILS, "4" = losing 15 ILS, "5" = losing 10 ILS, and "6"= losing 5 ILS. In both pay-for-report conditions, the payoffs associated with reporting a given value was identical but framed as either a gain or a loss. For example, reporting "6" yielded 30 ILS both in the gain (0 ILS + 30 ILS = 30 ILS) and in the loss framing conditions (35 ILS - 5 ILS = 30 ILS). Further, in both pay-for-report conditions, on a given trial, participants could earn between 5 and 30 ILS. Whereas the minimum payment of 5 ILS was achieved by reporting 1 in both conditions, it was framed as a gain of 5 ILS in the gain condition (in which participants had an initial endowment of 0 ILS) and a loss of 30 ILS in the loss framing (in which participants had an initial endowment of 35 ILS). Therefore, by making self-serving mistakes, participants in the gain condition could increase their gains whereas participants in the loss condition could decrease their losses compared with this minimal payment.

Participants in the pay-for-accuracy conditions were incentivized to be accurate in the task. In the pay-for-accuracy, gain-framing condition (*n* = 30), participants started the task with 0 ILS and could earn 10 ILS if they reported the target correctly on one randomly selected trial. Participants in the pay-for-accuracy, loss-framing condition (*n* = 30) started the task with 10 ILS and could lose that money if they reported the target incorrectly on one randomly selected trial.

After reading the instructions, participants engaged in three practice trials. After each practice trial, they received feedback indicating how much money they would have earned had that trial been selected for pay. Next, to make sure participants understood the payoff rule, they answered two questions: (1) “Imagine that the value closest to the fixation cross is ‘two’, and you reported the number ‘two’. How many Shekels would you earn (lose)?” (2) “Imagine that the value closest to the fixation cross is ‘five’, and you reported the number ‘two’, how many Shekels would you earn (lose)?” After each answer, participants received feedback (“you were correct” or “you were incorrect”) and saw the correct answer to the question on the computer screen.

Next, participants engaged in 196 trials of the task (96 experimental and 100 filler). On each trial, a black fixation cross was displayed on the screen, followed by six dice. Only one die was objectively closest to the fixation cross (i.e., the target). In the experimental trials, the target was always 3, whereas in the filler trials to diversify the numbers appearing on the screen, the target was 1, 2, 4, 5, or 6. Across trials we varied the value second closest to the fixation cross (henceforth, value next to the target) to be either higher than the target (4 or 5, i.e., tempting in the pay-for-report condition) or lower than the target (1 or 2, i.e., not tempting in the pay-for-report condition). As in Pittarello et al. (2015), we varied the locations of the target and the fixation cross. The location of the target was the second, third, fourth, or fifth die from the left. The location of the fixation cross was 20, 40, or 60 pixels away from the center of the target. In all three locations, the fixation cross was always closer to the target than to the value next to the target. Finally, as in Pittarello et al., each combination (of value next to the target, location of the fixation cross, and location of target) was repeated twice; see Figure 2.2.

The complete experimental design included the within-subjects factors of temptation (value next to the target: higher vs. lower than the target) × target location (second vs. third vs. fourth vs. fifth location from the left) × location of the fixation cross (20 vs. 40 vs. 60 pixels away from the center of the target die) × combination (first vs. second repetition). Payment (pay-for-accuracy vs. pay-for-report) and framing (gain vs. loss) were manipulated between subjects. In the main text we focus on the effects of temptation, payment, and framing to predict participants' reports. Additional exploratory analyses assessing the effects of the location of the fixation cross and the location of the target are reported in the supplementary online material (SOM)[2].

After completing the task, participants answered the following two control questions: (1) "How much money did you start the task with?" and (2) "What was the criterion according to which you received your payment?" (participants had to choose between "according to whether I was correct about the value that was closest to the fixation cross" and "according to the value that I reported as the closest to the fixation cross, regardless of whether I was correct or not"). Here, we report analyses including all participants. Excluding those who did not answer the control questions correctly did not change the results, see SOM. Lastly, a few weeks after the experiment, participants were recruited by email to complete an online version of the approach and avoidance temperament questionnaire (Elliot & Thrash, 2010), see SOM for results.

In the pay-for-report conditions high die-roll values are tempting. Reporting a higher number secures a higher gain or a smaller loss than reporting honestly. We thus expect participants to be more likely to "mistakenly" report the value next to the target (instead of the target) when doing so is tempting (a self-serving mistake) than when it is not tempting (a self-hurting mistake). Further, we expect this pattern to be amplified in the loss, compared to the gain framing condition.

[2]SOM appear in https://www.sciencedirect.com/science/article/pii/S0022103119300563#s0110 as "Appendix A. Supplementary data".

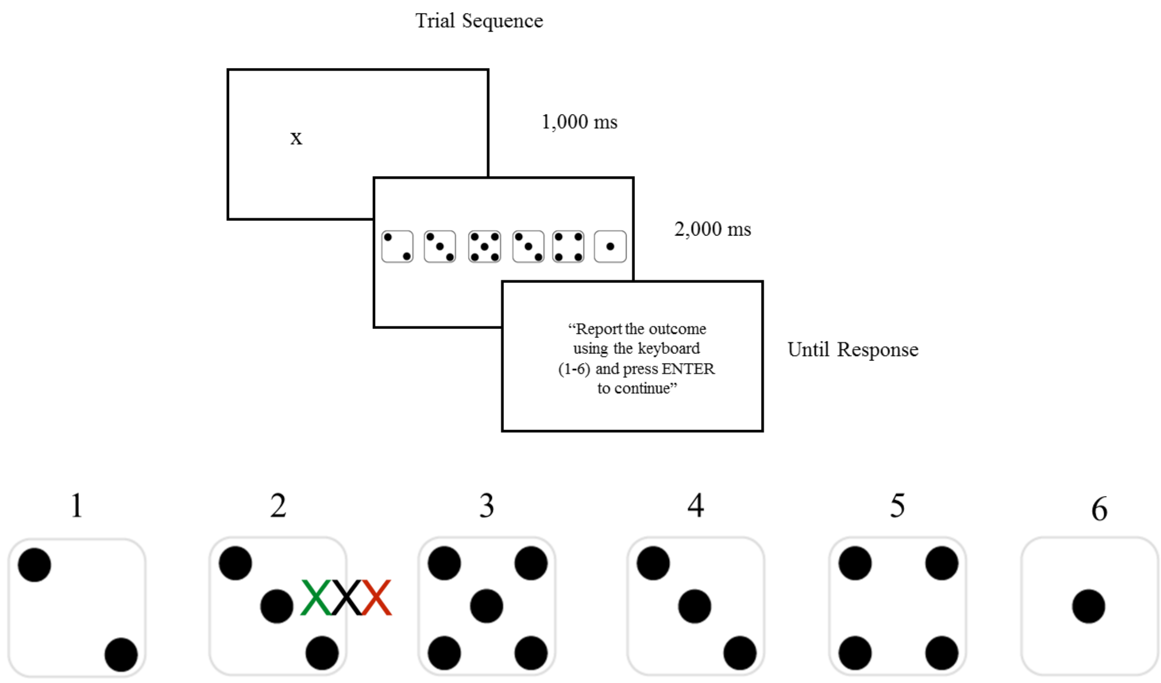

Figure 2.2. The procedure of the task and an example trial. In the example trial, the target = 3; the value next to the target = 5. In each trial, the location of the fixation cross was 20 (green), 40 (black), or 60 (red) pixels away from the center of the target.

Results

A generalized linear mixed model with Payment (pay-for-accuracy vs. pay-for-report), Framing (gain vs. loss), and Temptation (value next to the target: higher vs. lower than the target) predicting the Likelihood to report the value next to the target revealed a three-way Payment × Framing × Temptation interaction, $F(1, 11512) = 4.41$, $p = .036$, $b = -0.53$, 95% CI [-1.026, -.036]. In the pay-for-report conditions, the Framing × Temptation interaction was significant, $F(1, 5756) = 26.81$, $p < .001$, $b = 0.95$, 95% CI [.591, 1.310]. Specifically, in the gain-framing condition, participants reported the value next to the target in 19.02% of the cases when it was tempting (i.e., higher than the target), and in 8.05% of the cases when it was not tempting (i.e., lower than the target), $b = .08$, $t(5756) = 4.52$, $p < .001$, 95% CI [.050, .127]. The pattern was amplified in the loss-framing condition, in which the gap was twice as large with participants reporting the value next to the target in 27.29% of the cases when it was tempting, and only in 5.55% of the cases when it was not tempting, $b = .18$, $t(5756) = 5.33$, $p < .001$, 95% CI [.115, .248], see Figure 2.3. In the pay-for-accuracy condition the Framing × Temptation interaction attenuated, see SOM and Figure 2.3.

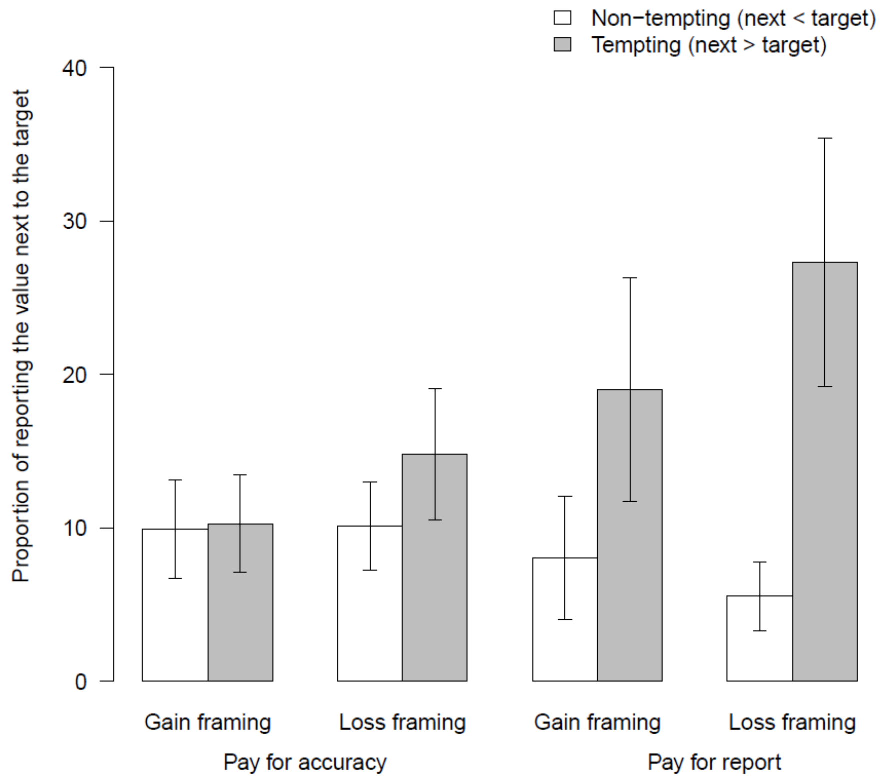

Figure 2.3. Behavioral results of Experiment 2.1. Proportion (and 95% confidence intervals) of the trials in which participants reported the value next to the target as a function of payment (pay-for-accuracy vs. pay-for-report), framing (gain vs. loss), and temptation of the value next to the target (non-tempting: next < target vs. tempting: next > target). The CIs are calculated for each participant across all trials, and then averaged across participants.

Discussion and Experiment 2.2

Experiment 2.1 revealed that, when paid according to the value they report, participants were more likely to make self-serving than self-hurting mistakes. Further, participants made more self-serving mistakes when facing a potential loss, compared to an equally sized gain. Experiment 2.2 aimed at replicating the framing effect obtained in Experiment 2.1 and investigating how attentional processes may shape it. Specifically, we tested which of three theoretical models gains empirical support: loss framing increases attention allocated to tempting information (H1a); the use of tempting information (H1b); or both the attention allocated to, and use of tempting information (H1c). We employed the same procedure as in Experiment 2.1, focusing on the pay-for-report conditions, and traced participants' eye movements, measuring gaze dwell time – the time people spend looking at a piece of information.

Method

A total of 89 university students (55.05% females, M_{age} = 24.20, SD_{age} = 2.50) participated in the experiment in exchange of a show-up fee of 20 ILS. As in Experiment 2.1, upon arriving to the lab participants learned they could earn additional pay based on their performance. Thus, the incentive to sign up and participate in the experiment was fixed and unrelated to the condition participants were assigned to. Following Pittarello et al. (2015), we estimated a data loss of ~20% of the sample for the analysis of participants' eye movements, due to low eye tracking accuracy (below 70%). We therefore aimed to increase our sample and collect data from 50 participants per between-subjects cell in order to take into consideration this potential data loss. In the time allocated to running the study, we were able to collect data from 89 participants: 49 in the gain framing and 40 in the loss framing condition. The unequal cell size was due to a programming error.

Upon arriving to the lab, participants were seated in a private cubicle, 60 cm from a 24-in. computer monitor (maximum resolution =1280 × 1024 pixels). Participants engaged in the ambiguous dice paradigm, while a Tobii T120 eye tracker (Tobii Technology, Danderyd, Sweden; sampling rate = 120 Hz; accuracy = 0.45°; standard nine-point eye tracking calibration) recorded their eye movements. Participants were randomly assigned to a gain or a loss framing condition, within the pay-for-report payoff scheme. All measures, instructions and questions were identical to those in Experiment 2.1. As in Experiment 2.1, here we report analyses of the full sample. Excluding participants who did not answer the control question correctly did not change the obtained results, see SOM.

Results

A generalized linear mixed model with Framing (gain vs. loss) and Temptation (value next to the target: higher vs. lower than the target) predicting the Likelihood to report the value next to the target revealed a Framing × Temptation interaction, $F(1, 8540) = 111.50$, $p < .001$, $b = 1.73$, 95% CI [1.409, 2.052]. Replicating the results of Experiment 2.1, in the gain-framing condition, participants reported the value next to the target more often when it was tempting (22.02%) than when it was not (7.86%), $b = .12$, $t(8540) = 5.48$ $p < .001$, 95% CI [.076, .161]. This pattern was amplified in the loss-framing condition in which participants reported the value next to the target in 37.81% of the trials when it was tempting, and only in 4.58% of the trials when it was not, $b = .27$, $t(8540) = 5.85$, $p < .001$, 95% CI [.183, .366], see Figure 2.4.

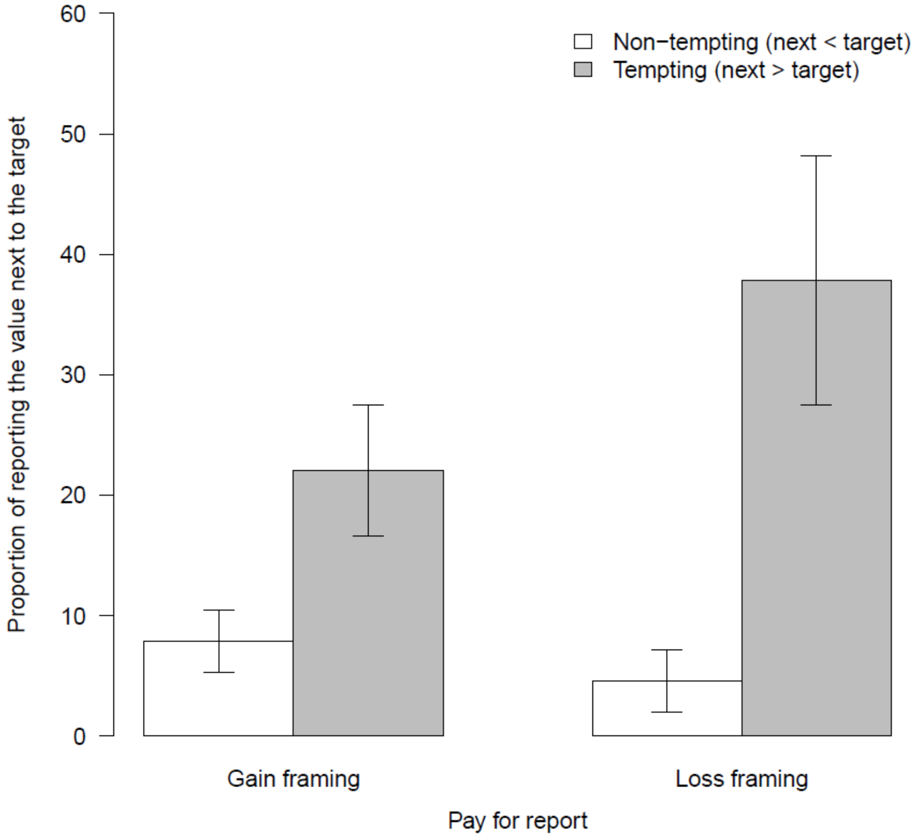

Figure 2.4. Behavioral results of Experiment 2.2. Proportion (and 95% confidence intervals) of the trials in which participants reported the value next to the target as a function framing (gain vs. loss) and temptation of the value next to the target (non-tempting: next < target vs. tempting: next > target). The CIs are calculated for each participant across all trials, and then averaged across participants.

Eye tracking. As in Pittarello et al., (2015) we excluded participants with low eye tracking accuracy (below 70%) from the analyses, which left 38 participants in the gain-framing and 26 participants in the loss-framing conditions. Including all participants did not change the obtained results, see SOM.

Figure 2.5 presents the amount of time participants spent looking at the target and the value next to the target, as a function of framing, and whether the value next to the target was tempting (higher than the target) or not (lower than the target). As can be seen in the figure, overall, participants spent more time looking at the target (M = 689.23ms, SD = 560.51) than at the value next to the target (M = 189.83ms, SD = 301.23), $F(1, 12286) = 4{,}085.93$, $p < .001$, b = -.499, 95% CI [-0.515, -0.484]. In the remaining analyses we focus on participants' attention to the value next to the target and test between models H1a, H1b, and H1c.

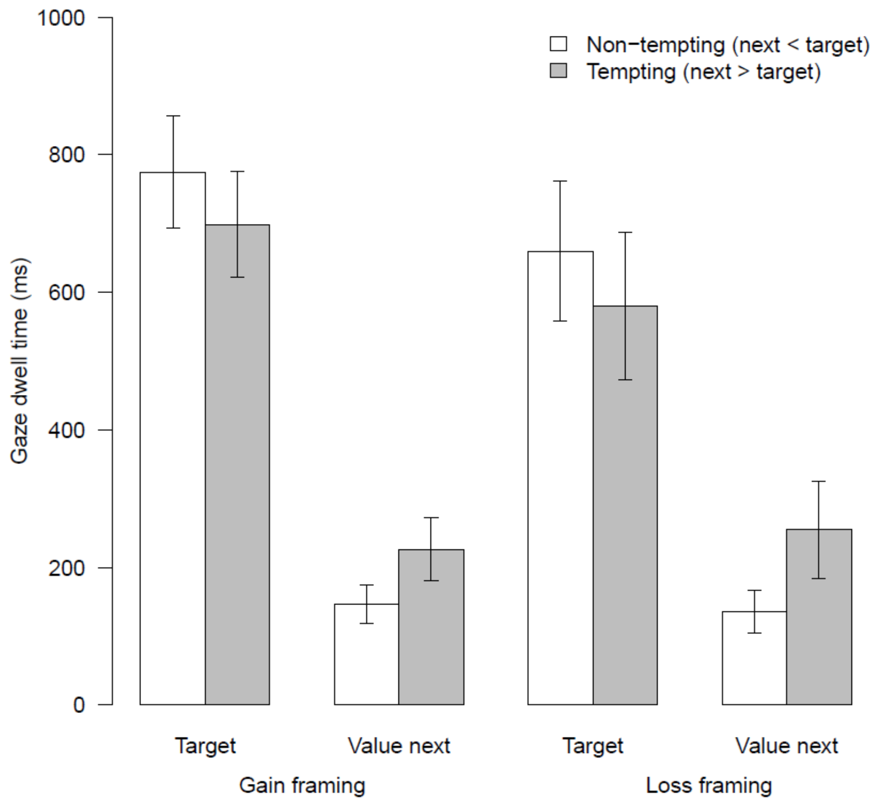

Figure 2.5. Eye tracking results of Experiment 2.2. The mean time (and 95% confidence intervals) participants spent looking (in milliseconds) at the target (correct value), and the value near the target, as a function of framing (gain vs. loss) and temptation of the value next to the target (non-tempting: next < target vs. tempting: next > target). The CIs are calculated for each participant across all trials, and then averaged across participants. Out of the total 2000 ms, participants looked at the target and the value next to the target for ~ 900 ms.

The eye tracking data allows testing how people divide their attention between the target value (i.e., the information they are instructed to report) and the value appearing second closest to the target in loss and gain framing conditions. That is, whether people spend more time looking at tempting, yet incorrect, values in the loss compared to gain framing condition. To do so, we calculated for each participant (1) the average time spent looking at the value next to the target when it was tempting (higher than the target) and not tempting (lower than the target), and (2) the proportion of self-serving and self-hurting mistakes, out of the potential mistakes participants could make (48 for each type of mistakes). We then ran three moderated mediation analyses using PROCESS macro (Hayes, 2013; Preacher, Rucker, & Hayes, 2007) and assessed how framing affects the attentional process underlying self-serving mistakes.

Does loss increase attention to tempting information (H1a)? A moderated mediation analysis using a bootstrapping procedure with 5000 iterations (Hayes, 2013; model 7) revealed that Temptation predicts the Likelihood to report the value next to the target, b = .206, $t(126) = 5.57$, $p < .001$, 95% CI = [.133, .279]. Participants reported the value next to the target more often when it was tempting (26.46%) than when it was not (5.89%). Temptation further predicted the Gaze dwell on the value next to the target, b = .079, $t(124) = 2.69$, p = .008, 95% CI = [.021, .137]. Participants looked longer at the value next to the target when it was tempting (M = 237.58ms, SD = 160.37ms) than when it was not (M = 142.08ms, SD = 83.38ms). The Temptation × Framing interaction, however, did not predict the Gaze dwell on the value next to the target, b = .039, p = .394. That is, the loss (versus gain) framing did not alter the time participants looked at the tempting information. Further, Gaze dwell on the value next to the target predicted the Likelihood to report it, b = 1.07, $t(124) = 9.86$, $p < .001$, 95% CI = [.863, 1.296], and Temptation predicted the Likelihood to report the value next to the target as well, b = .102, $t(124) = 3.45$, $p < .001$, 95% CI = [.043, .161]. The indirect effect of Temptation on the Likelihood to report the value next to the target via Gaze dwell was positive and significant both in the gain, 95% CI = [.029, .146], and the loss-framing conditions, 95% CI = [.041, .239]. The difference between the two conditional indirect effects was not significant, 95% CI = [-.049, .159], providing no support for the moderated mediation model suggested by H1a, see Figure 2.6.

Does loss increase the use of tempting information (H1b)? A moderated mediation analysis using a bootstrapping procedure with 5000 iterations (Hayes, 2013; model 14) revealed that Temptation predicts the Likelihood to report the value next to the target, b = .206, $t(126) = 5.57$, $p < .001$, 95% CI = [.133, .279], as well as the Gaze dwell on the value next to the target, b = .095, $t(126) = 4.22$, $p < .001$, 95% CI = [.050, .140]. The Framing × Gaze dwell on the value next to the target interaction predicted the Likelihood to report the value next to the target, b = .780, $t(123) = 4.05$, $p < .001$, 95% CI = [.399, 1.161]. Specifically, the effect of Gaze dwell on the value next to the target on the Likelihood to report it was positive and significant in the gain-framing condition, b = .683, $t(123) = 4.84$, $p < .001$, 95% CI = [.404, .962], and was amplified in the loss-framing condition, b = 1.463, $t(123) = 10.40$, $p < .001$, 95% CI = [1.184, 1.741]. That is, equal time spent looking at tempting information translated to more self-serving mistakes in the loss than in the gain-framing condition. Temptation further predicted the Likelihood to report the value next to the target, b = .102, $t(123) = 3.68$, $p < .001$, 95% CI = [.047, .158]. The indirect effect of Temptation on the Likelihood to report the value next to the target via Gaze dwell was positive and significant in the gain-framing, 95% CI = [.023, .115], and the loss-framing condition, 95% CI = [.069, .217]. The difference between the two conditional indirect effects was significant 95% CI = [.023, .143], supporting the moderated mediation H1b hypothesis, see Figure 2.6.

Does loss increase attention to and use of tempting information (H1c)? A moderated mediation analysis using a bootstrapping procedure with 5000 iterations (Hayes, 2013; model 58) revealed that Temptation predicts the Likelihood to report the value next to the target, $b = .206$, $t(126) = 5.57$, $p < .001$, 95% CI = [.133, .279]. Further, Temptation predicted Gaze dwell on the value next to the target, $b = .079$, $t(124) = 2.69$, $p = .008$, 95% CI = [.021, .137]. The Temptation × Framing interaction predicting Gaze dwell on the value next to the target was not significant, $b = .039$, $p = .394$. Further, Gaze dwell on the value next to the target predicted the Likelihood to report it, $b = .683$, $t(123) = 4.84$, $p < .001$, 95% CI = [.404, .962], and the Framing × Gaze dwell interaction also predicted the Likelihood to report the value next to the target, $b = .780$, $t(123) = 4.05$, $p < .001$, 95% CI = [.399, 1.161]. With Framing, Temptation, and Gaze dwell in the model, Temptation predicted the Likelihood to report the value next to the target, $b = .102$, $t(123) = 3.68$, $p < .001$, 95% CI = [.047, .158]. The indirect effect of Temptation on the Likelihood to report the value next to the target via Gaze dwell was positive and significant both in the gain, 95% CI = [.013, .110], and the loss-framing condition, 95% CI = [.056, .298]. The difference between the two conditional indirect effects was not significant, 95% CI = [-.009, .254], providing no support for the moderated mediation model suggested by H1c, see Figure 2.6.

Comparing the models. To compare models H1a, H1b, and H1c we calculated the Akaike Information Criterion (AIC) for each model. The AIC is used to compare the relative fit of different models to the data, with smaller scores indicating a better fit of the model (Akaike, 1974). Results revealed an AIC of -265.56 for model H1a, an AIC of -282.85 for model H1b, and an AIC of -279.75 for model H1c, suggesting that model H1b fits the data better than models H1a and H1c.

Based on Burnham and Anderson (2004), the larger the gap between the model's AICs the more support there is to the model with the lower AIC over the model with a higher AIC. A gap larger than 10 essentially indicates no support for the model with the higher AIC. Further, it is possible to assess the relative probability that the model with the lowest AIC will fit the data better (i.e., minimize information loss) than models with higher AICs by calculating the exponential of $((AIC_{\text{model with the smallest AIC}} - AIC_{\text{model I}})/2)$. Such value can range between 100% (equal support for both models) and 0% (no support for the model with the higher AIC to be a better fit for the data compared to the model with the lower AIC).

The gap between the AICs of model H1b and H1a is 17.29, indicating that there is essentially no support for model H1a over H1b. Specifically, there is only 0.01 probability [exp((-282.85+265.56)/2) = 0.0001] that model H1a fits the data as good as model H1b. The gap between the AICs of model H1b and H1c is 3.1, indicating more support for model H1b than H1c. Specifically, there is only 0.21 probability [exp((-282.85+279.75)/2) = 0.2122] that model H1c fits the data as good as model H1b. That is, model H1b is ~5 times more likely than model H1a. Taken together, our results suggest that model H1b fits the eye tracking data the best, and that loss (versus gain) framing does not increase people's attention to tempting information, but rather their use of the tempting information, resulting in self-serving mistakes.

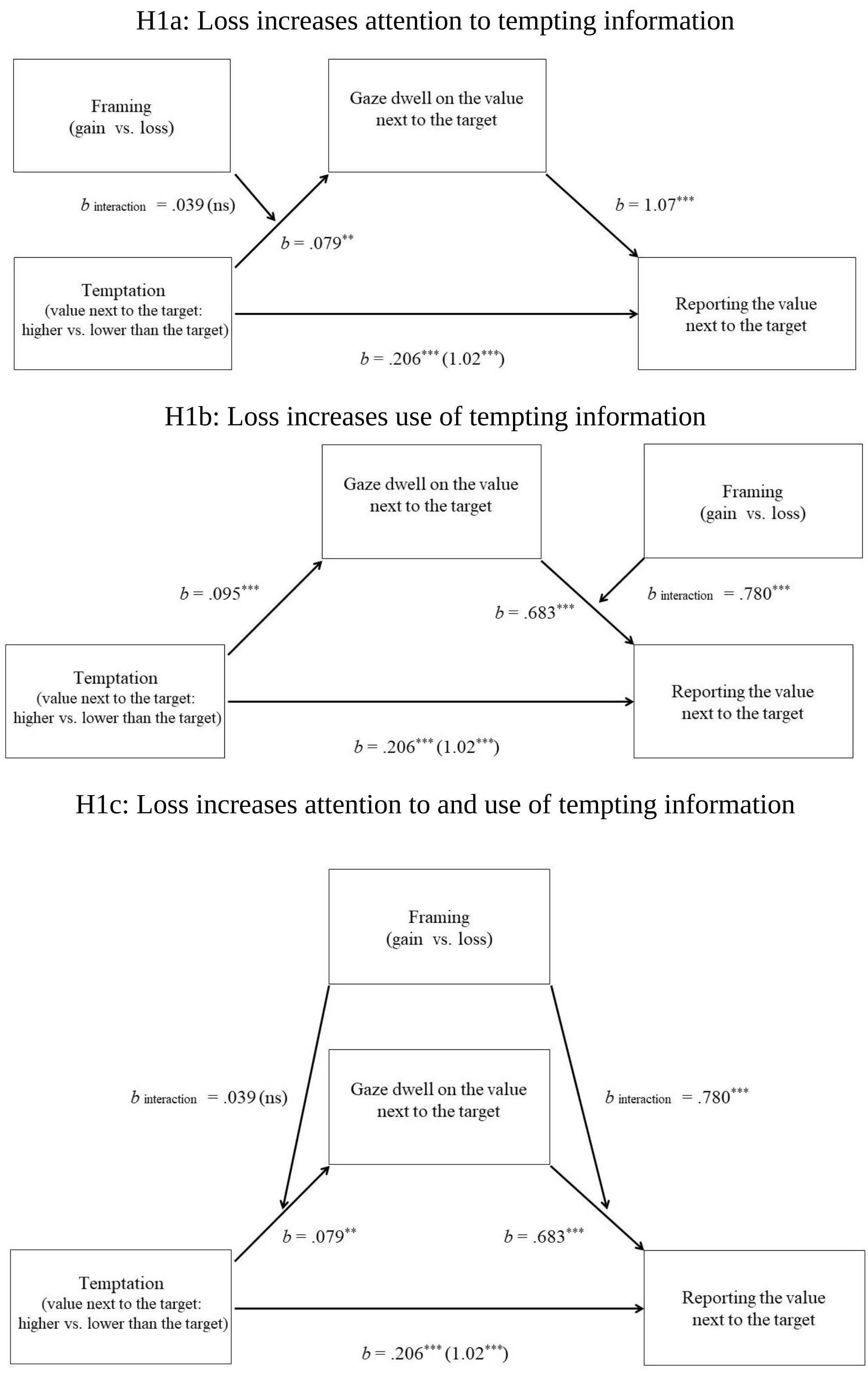

Figure 2.6. Results for the (H1a) loss increases attention to tempting information, (H1b) loss increases use of tempting information, and (H1c) loss increase attention to, and use of tempting information hypotheses. * $p < .05$, $^{**}p < .01$, $^{***}p < .001$.

Sensitivity analyses. We conducted additional sensitivity and Bayesian analysis, as well as an equivalence test in order to make sure (1) we had sufficient power to detect the effect of Framing on Self-serving mistakes, and (2) that we can meaningfully interpreted the lack of interaction between Framing and Temptation, predicting Gaze dwell.

The effect of framing on self-serving mistakes. We ran a sensitivity analysis to determine the effect size that our sample sizes in Experiments 2.1 and 2.2 allowed detecting. The test does not exist in G*power for a generalized linear mixed model, therefore we ran a sensitivity test for an ANOVA with Framing (gain vs. loss) predicting Self-Serving (vs. Self-Hurting) mistakes. We then compared the effect size that we could detect, to the overall effect size we found.

A sensitivity analysis for an ANOVA with 80% power to detect an effect and significance level of 0.05 revealed that the sample of Experiment 2.1 (n = 60 in the pay-for-report conditions) was sufficient to detect a medium-large effect size of f = 0.36, and that the sample size of Experiment 2.2 (n = 89) was sufficient to detect a medium-large effect size of f = 0.30. We then compared those calculated effect sizes to the observed effect size of Framing on Self-Serving mistakes. To directly compare the effect sizes, we used an ANOVA to analyze the data from Experiments 2.1 and 2.2. Specifically, focusing on the pay-for-report conditions, we computed for each participant the proportion of self-serving and self-hurting mistakes (out of the maximum number of potential mistakes participants could make, which equaled 48). We then computed for each participant the gap between the proportion of self-serving and self-hurting mistakes. The gap between self-serving and self-hurting mistakes could range between -100% (indicating that a participant made the maximum possible number of self-hurting mistakes and no self-serving mistakes) and 100% (indicating that a participant made the maximum possible number of self-serving mistakes and no self-hurting mistakes). A gap of 0% indicates that a participant made the same number of self-serving and self-hurting mistakes.

We ran a 2 (Framing: gain vs. loss) × 2 (Experiment: 2.1 vs. 2.2) ANOVA predicting the Gap between self-serving and self-hurting mistakes. Results revealed a main effect for Framing, $F(1, 205) = 16.67$, $p < .001$, $\eta^2 = .075$. The Gap between self-serving and self-hurting mistakes was larger in the loss framing (M = 21.21%, SD = 28.83) than in the gain framing (M = 9.48%, SD = 19.20). The Framing × Experiment interaction was not significant, $F(1, 205) = 3.13$, $p = .078$. Thus, the overall effect size of framing on self-serving (vs. hurting) mistakes observed in our data was of a medium-large size ($\eta^2 = .075$, equivalent to f

= 0.28). Taken together, the data we have collected allowed detecting a medium-large effect, and analyzing our data using ANOVA, we find a medium-large effect of framing on self-serving mistakes.

Bayesian analyses. To further increase our ability to interpret the non-significant Framing × Temptation interaction on Gaze dwell time we conducted Bayesian analyses comparing a model where only Temptation is the predictor for Gaze dwell with a model that includes Temptation, Framing, and Temptation × Framing interaction as predictors for Gaze dwell. Results revealed a Bayes factor of $BF_{10} = .084$, suggesting strong evidence in favor of a model where only Temptation is the predictor. Specifically, the data was 11.90 times more likely to occur when Temptation is the only predictor for Gaze dwell time than when Temptation, Framing, and a Temptation × Framing interaction predict Gaze dwell.

Equivalence test. Lastly, we conduct an equivalence test (Lakens, Scheel, & Isager, 2018; Lakens, McLatchie, Isager, Scheel, & Dienes, 2018) to assess whether we can reject a meaningful effect of framing on attention to tempting information. Drawing on the vast literature on loss aversion, we use the “loss aversion ratio” reported by Kahneman (2011), which ranges between 1.5 and 2.5. To be on the conservative side, we use the lower end of the range (1.5) as a meaningful effect of interest.

In the gain framing condition, on average, participants spent 225.97ms (SD = 183.30ms) looking at the value next to the target when it was tempting. Thus, for the effect to be of a meaningful size, participants needed to look at the value next to the target when it is tempting in the loss framing condition 1.5 times longer. A meaningful gap of interest is accordingly 112.98ms. We thus set the boundaries of interest as -112.98ms and 112.98ms and test whether the gap between the time participants looked at tempting information in the loss and gain framing conditions was both significantly larger than the low boundary, and significantly lower than the high boundary. The equivalence test was significant, showing that the actual gap obtained (28.58ms) was significantly higher than the lower bound, $t(62) = 2.05$, $p = .021$, and significantly lower than the higher bound, $t(62) = -3.45$, $p < .001$. Thus, the effect of framing on attention to tempting information was not of a size of interest.

Discussion

Experiment 2.2 replicated the results of Experiment 2.1 revealing that participants made more self-serving mistakes in a loss, compared to gain framing. Tracing participants' eye movements allowed us to test the relationship between framing and attention in predicting self-serving mistakes. Results supported the *loss increases use of tempting information* (H1b) hypothesis. Specifically, participants looked a similar amount of time at tempting information under loss and gain framing. The similar time spent looking at tempting information, however, translated to more self-serving mistakes in the loss, compared to the gain framing condition. Thus, the loss (versus gain) framing did not alter the way people divide their attention between correct and incorrect (yet tempting) information. Rather, framing altered people's likelihood to use the tempting information in a self-serving way.

General discussion

In ambiguous situations, temptation leads people to make self-serving mistakes. In two experiments, participants could make self-serving and self-hurting mistakes and were motivated to either minimize their losses or maximize their gains. Results revealed that overall, people made more self-serving than self-hurting mistakes. Importantly, this pattern was amplified when self-serving mistakes reduced the amount of money people could lose, compared to increase the amount of money they could gain. In particular, self-serving mistakes occurred twice as often when participants' payoff was framed as a loss, compared to a gain.

Tracing participants eye movements allowed to gain insight into the process underlying self-serving mistakes in loss and gain framing. Consistent with work on motivated attention (Fiedler et al., 2013; Glöckner et al., 2012; Mogg et al., 2003; Townshend & Duka, 2001), and replicating Pittarello et al., (2015), we found that people look longer at tempting, compared to non-tempting information, which in turn shapes their self-serving mistakes. Assessing attention allocation patterns allowed us to test three competing theoretical models that are derived from prior research. Results showed that the model in which framing moderates the relationship between the time people spend looking at tempting information and the likelihood to report it was the most supported by the data. The model in which framing moderated the relationship between tempting information and time spend looking at it, and the model in which framing moderated both relationships were not supported. As such, employing eye tracking was a useful tool for theory testing and improved our understanding of the process underlying self-serving mistakes. In particular, loss framing seems to increase the likelihood of using tempting information to boost payoff, but not the attention allocated to tempting information.

The results are in line with prior work showing that the valance of information, whether it is associated with losses or gains, does not affect attention processes (Brosch et al., 2008; Müller et al., 2016; Wentura et al., 2014). Whereas prior work found that losses increase *overall* attention dedicated to a task (Yechiam & Hochman, 2013a, 2013b), it seems to not affect the way people *divide* their attention between correct information and tempting, yet incorrect information. One feature of the Ambiguous dice paradigm is that participants had a limited time to look at the die roll outcomes (2,000ms), after which they had to report their response. As such, participants could only look at any given piece of information for a maximum of 2,000ms. An interesting avenue for future research would be to allow people an unlimited time before making a decision. Doing so should allow people to deliberate about their choice for as long as they feel needed. It will further allow assessing whether, in line with prior work (Yechiam & Hochman, 2013a, 2013b) loss framing leads to *overall* more deliberation and attention allocated to making choices that carry ethical consequences. Another avenue to explore would be to use more complex stimuli (i.e., that require more attention to digest) to further test the role of attention in shaping self-serving mistakes.

Here we focused on the *visual* attention process by measuring gaze dwell – a well-established proxy for attention (e.g., Glöckner et al., 2012; Papies et al., 2008; Raab & Johnson, 2007). We found no difference in the visual attention people allocate to tempting information in the loss and gain framing conditions. It is unclear, however, whether other phases of information processing are similar between loss and gain framing. For instance, it might be that people cognitively process tempting information differently under loss and gain framing. People might look at the tempting outcome to the same extent in loss and gain framing, but perhaps perceive the tempting information as closer to the fixation cross in the loss than gain framing. Prior work shows that people's motivation can influence their cognitive processing and perception of reality (Balcetis & Dunning, 2006; 2007; 2009). As such, testing whether people's perception differs when they aim to maximize their gains, compared with minimize their losses is an interesting avenue for future research.

Unlike outright lies that are by definition conscious, whether self-serving mistakes are conscious or not remains an open question. Hochman and colleagues (2016) find support for both conscious and unconscious elements when people make self-serving mistakes. First, they find increased physiological arousal when people make self-serving mistakes, indicating that, at least to an extent, people are aware of the mistake they are making. Second, they find that people pay less attention to self-hurting information, indicating a rather unconscious process. Although not a focal interest of the current work, understanding whether people make self-serving mistakes intentionally, is an interesting avenue for future work. Asking participants to complete the ambiguous dice paradigm and then financially incentive them to accurately estimate whether they think they made more, less, or a similar amount of self-serving and self-hurting mistakes can be a promising first step in addressing this question.

Lastly, uncovering the attentional process underlying self-serving mistakes—a mild form of unethical behavior—can help creating interventions to reduce them. Prior work found that manipulating people's attention affects their ethical judgment and decision making (e.g., Pärnamets et al., 2015). Thus, one might focus on interventions that highlight information that promotes ethical behavior, and make information that promotes unethical behavior less noticeable (e.g., by changing its' perceptual features; see Pittarello, Frătescu, & Mathôt, 2019). Other research, however, finds that externally manipulating attention may have a rather small effect on judgment and decision making, if at all. For instance, exogenously manipulating participants' visual attention pattern resulted in only 1.19% change in their decisions in moral dilemmas (Ghaffari, & Fiedler, 2018). Similarly, complex moral judgment tasks were not affected by manipulated attention (Newell & Le Pelley, 2018). Together, those results and ours suggest that manipulating attention might not be a promising intervention when aiming to reduce unethical behavior. The attentional process underlying self-serving mistakes appear to be similar in different framing settings. This suggests that addressing the motivational factors, such as the motivation to secure gains vs. prevent losses or the motivation to avoid making self-serving mistakes, is a more promising intervention than diverting people's attention from tempting information.

Conclusion

In ambiguous settings that allow for self-serving interpretations, people make more self-serving mistakes to minimize a loss than to maximize a gain. Rather than causing tempting information to capture *attention*, loss framing increases people's propensity to *use* the tempting information they have observed to boost their profits. Framing goals and incentive schemes as potential gains can help to decrease the occurrence of self-serving mistakes, and in turn foster environments with more accuracy and less motivated mistakes.