IZA Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 15292

Do Losses Trigger Deliberative Reasoning?

Jeffrey Carpenter
David Munro

MAY 2022

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 15292

Do Losses Trigger Deliberative Reasoning?

Jeffrey Carpenter
Middlebury College and IZA

David Munro
Middlebury College

MAY 2022

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Do Losses Trigger Deliberative Reasoning?*

There is a large literature evaluating the dual process model of cognition, including the biases and heuristic it implies. To advance this literature, we focus on what triggers decision makers to switch from the intuitive process (aka System 1) to the more deliberative process (aka System 2). Based on previous studies indicating that potential losses increase cognitive effort, we posit that losses may also differentially trigger System 2 reasoning. To evaluate this hypothesis, we design an experiment based on a task that has been developed to distinguish between System 1 and System 2 thinking – the cognitive reflection task. Replicating previous research, we find that losses elicit more effort (measured by the time spent on the task and the incidence of correct answers). However, we also find that losses differentially reduce the incidence of intuitive answers, consistent with triggering System 2. To complement these results, we provide tests of the robustness of our results using aggregated data, subgroup analysis and the imposition of a cognitive load to hinder the activation of System 2.

JEL Classification: C9, D9
Keywords: dual process theory, cognitive effort, loss, experiment

Corresponding author:
David Munro
Department of Economics
Middlebury College
14 Old Chapel Rd.
Warner Hall 505
Middlebury, VT 05753
USA
E-mail: dmunro@middlebury.edu

* We thank seminar participants at the Liberal Arts Experimental Conference. Our study was registered at the AEA RCT registry (ID:AEARCTR-0007839).

1 Introduction

It is well documented that individuals make systematic errors in various important decision contexts. These biases lead to sub-optimal outcomes and thus harm decision makers. A prominent conceptualization of biased decision making is the "dual process" or "dual system" framework first mentioned by James (1890) and popularized more recently in Kahneman (2011). In this framework, System 1 controls our "gut" responses and acts quickly, requiring limited cognitive effort. For less obvious or intuitive choices, relying too heavily on System 1 can lead to biased decisions. By contrast, System 2 represents more effortful, contemplative cognition, and consumes more time, energy and attention but tends to drive us towards better choices. While this framework now dominates behavioral research in economics and the other social sciences, our understanding of what causes effortful thinking is rudimentary (e.g., Westbrook and Braver, 2015). A particularly relevant question relating to the dual system framework involves understanding what causes individuals to switch from System 1 to System 2, since this distinction is thought to be at the core of many flawed choices.

In this study we explore how the context of incentives, gains or loses, marshal cognitive resources, influence the transition between System 1 and 2 and, ultimately, impact decision making. In an online experiment with approximately 1200 participants and three incentive treatments we test performance on seven Cognitive Reflection Test (CRT) questions, designed to differentiate between System 1 and 2 thinking by having an intuitive, but incorrect, answer. To formulate hypotheses about what economically-relevant factors might trigger deliberative System 2 reasoning, we rely on a literature in economics, cognitive psychology and neuroscience showing that financial rewards and the framing of rewards can affect the amount of cognitive effort people devote to decisions – the basic idea being that if financial incentives can slow decision makers down and elicit more effort from them, they might also activate System 2. For example, Smith and Walker (1993) showed that across 31 studies, increased rewards can shift choices toward the predictions of rational models, an outcome consistent with overcoming the opportunity cost of providing more cognitive effort. More recently, Clay et al. (2017) focused on how decision makers demonstrate increased vigilance and attention when losses are possible, Lejarraga et al. (2019) showed that losses prompt more attention and cognitive effort than gains, and Chen et al. (2020) extended these results showing that experimental participants expend disproportionately more effort to avoid losses even when the task is mostly physical.[1]

[1] Considering other factors that might activate System 2, Alter et al. (2007), for instance examined whether the ease with which information comes to mind when facing a problem (i.e., the fluency of information) and the perceived difficulty of the judgement regulate when System 2 will switch on. Similarly, Bourgeois-Gironde and Van Der Henst (2009) explored whether giving the decision maker hints in the framing of the problem

There is also a large literature examining the neural processing of gains and losses. For example, Gehring and Willoughby (2002) find that the medial frontal cortex shows greater activation to losses, relative to gains, within 265 milliseconds of the reward stimuli. Sallet et al. (2007) and Foti et al. (2015) find evidence of distinct facets of reward processing related to gains and losses in the anterior cingulate cortex (ACC) and Fujiwara et al. (2009) find evidence of separate coding of monetary rewards and punishment in distinct subregions in the cingulate cortex. Further, it is also thought that the ACC moderates the voluntary selection of behavior and encodes signals related to decision making (e.g. Amiez et al. (2005) and Wallis and Kennerley (2011)). These studies are relevant to our research as they highlight the differential neural responses to losses in regions of the brain which also influence decision making and attention allocation.

To examine the efficacy (and symmetry) of financial rewards in the provision of cognitive effort and the activation of System 2, we designed three experimental treatments. In the no reward or "None" treatment, subjects receive no additional compensation for correct CRT answers, in the "Gain" treatment subjects are awarded an additional sum for each correct response, and in the "Loss" treatment subjects who begin with an endowment are financially penalized for each incorrect response. In addition, we conducted an auxiliary experiment to test the extent to which any effects of our treatments are on the extensive margin of triggering System 2 and not just on the intensive margin of eliciting more cognitive effort. This second experiment was identical to the one just described except that it imposed a 20 second time constraint per question. This time constraint introduces a "cognitive load" that engages the working memory of our participants (Gillard et al., 2009; Kalyuga, 2011) and limits individuals' ability to access System 2 (e.g., Arsalidou et al., 2013; Bago and De Neys, 2017; Deck et al., 2021). We hypothesized that compromising System 2's ability to respond in this second experiment would result in our incentive treatments performing equally well (or poorly) if they differ primarily in their ability to trigger System 2.

Replicating the previous literature, we first find that financial incentives lead our participants to exert more cognitive effort. They spend more time working on individual problems and are more likely to work to the correct answer when they are incentivized financially. We also find that the effects of gains and losses on cognitive effort are asymmetric – the treatment effect of losses is twice as large as the treatment effect of gains. This too is in line with the previous literature which suggested that participants pay particular attention to potential losses. Our biggest contribution, however, is that we find that participant behavior is consistent with our hypothesis that losses stimulate individuals to switch to System 2 thinking. Subjects in the Loss treatment give fewer intuitive (and incorrect) responses,

cues System 2.

suggesting that losses induce a switch between System 1 and 2.

To test the robustness of our main results, we aggregate the data to the level of the individual participant and show that the effects of losses, in particular, remain large and highly significant. We then explore heterogeneous treatment effects among a subpopulation of participants commonly associated with higher impulsivity – young males – and find that they are indeed more likely to give intuitive answers (i.e., think fast) but their behavior is particularly sensitive to losses. For young men, potential losses dramatically reduces intuitive responses. One interpretation of these results is that more impulsive types are predisposed to System 1 decision making (possibly because they face higher costs to switch) and losses, which tend to loom larger for most individuals, are necessary for them to cross the threshold into System 2, whereas less impulsive types tend to occupy System 2 more readily, making losses less necessary and impactful. As a final robustness check on the hypothesis that losses cause individuals to switch systems, we examine the results of our second experiment with time constraints. Here we find strong evidence that our manipulation imposes a significant cognitive load on participants and that the load leads to more intuitive responses. We then show that once we overload System 2, the effect of losses disappears. This result gives us more confidence that the main effect of losses is to trigger deliberative, System 2 thinking.

2 The Experiment

Approximately 1200 subjects were recruited via Amazon's MTurk platform and the experiment was conducted on Qualtrics.[2] To begin, subjects gave consent (the protocol was approved by the Middlebury College IRB), responded to basic demographic questions and were randomized into the different treatment cells. The treatment cells involved a different incentive prompt and a simple mathematical question to ensure subjects comprehended the incentives. Subjects who failed this comprehension question were removed from the experiment. The three incentive prompts are reproduced in the Appendix. In all incentive treatments subjects were awarded \$1.00 for completing the experiment. In the None treatment subjects received no additional compensation for correct CRT answers. In the Gain treatment, subjects were awarded an additional \$0.25 for each correct CRT response. And finally, in the Loss treatment subjects were endowed with \$2.75 and were penalized \$0.25 for each incorrect CRT response. To avoid any possible concerns related to wealth effects, subjects were only informed about their performance after the completion of the experiment.

After finishing the 7 CRT questions, subjects responded to the NASA TLX survey, which

[2]Subjects were based in the United States and were required to have prior experience in at least 100 HITs and have a HIT approval rating of 95%. An initial captcha question was utilized to eliminate bots.

was intended to measure work load. The survey instrument, developed by NASA's Human Performance Group, asked participants to subjectively rank the effort they put into a task (on a 0-10 scale) across six demand dimensions (mental, physical, temporal, performance, effort and frustration level). Because our task was entirely mental, we pruned the physical demand question. On average, subjects spent approximately 5 minutes on the experiment and earned almost $2, which translates to an hourly wage of approximately $24/hr.

In the second experiment, we implemented a cognitive load on our participants by restricting the time allocated to each question to 20 seconds. The choice of 20 seconds was a tradeoff between giving subjects enough time to read and comprehend the questions, but also making them feel rushed to answer. The average reading speed of an adult is around 250 words per minute, which would result in a reading time of 13.2 seconds for our longest question. The median question response time in our unconstrained experiments was 16 seconds. So the 20 second limit seemed to be a reasonable choice to allow question comprehension, but also cause subjects to feel rushed. In addition to the time constraint, the experiment included a clock counting down the time remaining to answer a question. This served to distract attention, further loading System 2. Manipulating response time is a standard method for imposing a cognitive load. In the context of answering questions like the CRT, Gillard et al. (2009) showed that reducing the time spent on similar questions to the median time spent by unconstrained participants significantly reduced correct answers (and increased intuitive responses). More broadly, Kalyuga (2011) review the literature on working memory and explain how time constraints are an essential part of any cognitive load.

As noted, the experiment involved seven CRT questions. Three were the original CRT questions formulated by Frederick (2005).[3] Because these CRT questions are so ubiquitous researchers have become concerned about repeated exposure of subjects to these questions. Toplak et al. (2014) propose four new CRT questions, which we also include in our set.[4] Given the online implementation of the experiment an additional concern is the ease at which subjects could search for answers to these questions online, given the ubiquity of these questions. To diminish this concern to some degree, we slightly modified the numerical values and changed the language in the questions. See Table I for the seven CRT questions and our slightly modified versions.

[3] These original questions are the first three in Table I.

[4] We find some evidence that subjects perform better, on average, on the original CRT questions relative to the four new questions. Across all incentive treatments the probability of a correct response on the original questions is 64% versus 47% on the new questions. This could reflect some prior exposure to the original CRT questions, however, even with the original questions more than a third of subjects still respond incorrectly.

	CRT Questions	Modified CRT Questions
1	A bat and a ball cost $1.10 in total. The bat costs a dollar more than the ball. How much does the ball cost? __ cents	A rod and a reel cost $3.50 in total. The rod costs three dollars more than the reel. How much does the reel cost? __ dollars
	CA: 5 cents, IA: 10 cents	CA: 0.25 dollars, IA: 0.5 dollars
2	If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? __ minutes	If it takes 7 squirrels 7 minutes to collect 7 nuts, how long would it take 50 squirrels to collect 50 nuts? __ minutes
	CA: 5 minutes, IA: 100 minutes	CA: 7 minutes, IA: 50 minutes
3	In a lake, there is a patch of lily pads. Everyday, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? __ days	There is a virus in the city you live. Every day, the number of people infected doubles in size. If it takes 64 days for the whole city to be infected how long would it take for half the city to be infected? __ days
	CA: 47 days, IA: 24 days	CA: 63 days, IA: 32 days
4	If John can drink one barrel of water in 6 days, and Mary can drink one barrel of water in 12 days, how long would it take them to drink one barrel of water together? __ days	If Jose can drink one bottle of Coke in 4 hours, and Beth can drink one bottle of Coke in 12 hours, how long would it take them to drink one bottle of Coke together? __ hours
	CA: 4 days, IA: 9 days	CA: 3 hours, IA: 8 hours
5	Jerry received both the 15th highest and the 15th lowest mark in the class. How many students are in the class? __ students	Michael finished both the 10th highest and the 10th lowest in the tournament. How many people are in the tournament? __ people
	CA: 29 students, IA: 30 students	CA: 19 people, IA: 20 people
6	A man buys a pig for $60, sells it for $70, buys it back for $80, and sells it finally for $90. How much has he made? __ dollars	A woman buys a goat for $30, sells it for $40, buys it back for $50, and sells it finally for $60. How much has she made? __ dollars
	CA: $20, IA: $10	CA: $20, IA: $10
7	Simon decided to invest $8,000 in the stock market one day early in 2008. Six months after he invested, on July 17, the stocks he had purchased were down 50%. Fortunately for Simon, from July 17 to October 17, the stocks he had purchased went up 75%. At this point, Simon has: a. broken even in the stock market, b. is ahead of where he began, c. has lost money	Deborah decided to invest $5,000 in Bitcoin one day early in 2016. Five months after she invested, on August 27th, the value of her Bitcoin investment was down 50%. After this, from August 27th to November 27th, the Bitcoin she had purchased went up 75%. At this point, compared to her starting position, Deborah has: a. broken even in her Bitcoin investment, b. is ahead of where he began, c. has lost money
	CA: c, IA: b	CA: c, IA: b

Table I. *CRT questions. CA and IA denote "Correct Answer" and "Intuitive Answer," respectively.*

3 Results

3.1 Data overview

	Total	Gain (1)	No Reward (2)	Loss (3)	Diff. (1)-(2)	Diff. (1)-(3)	Diff. (2)-(3)
Age	39.053	39.332	40.194	37.626	-.862	1.706	2.569
Female	0.442	0.434	0.468	0.424	-.0348	.0100	.0448
Black	0.084	0.087	0.075	0.091	.0115	-.0044	-.0159
Asian	0.073	0.066	0.058	0.096	.0081	-.0297*	-.0378**
Hispanic	0.080	0.102	0.075	0.064	.0268	.0380	.0112
Advanced degree	0.128	0.133	0.138	0.113	-.0057	.0194	.0250
High income	0.548	0.548	0.553	0.542	-.0049	.0066	.0115
F-test for joint significance of characteristics						F-stat= 1.55	
F-test for joint significance of characteristics						p-value= 0.147	

Note: means or frequencies; $n = 1210$; * p<0.10, ** p<0.05, *** p<0.01.

Table II. Participant characteristics.

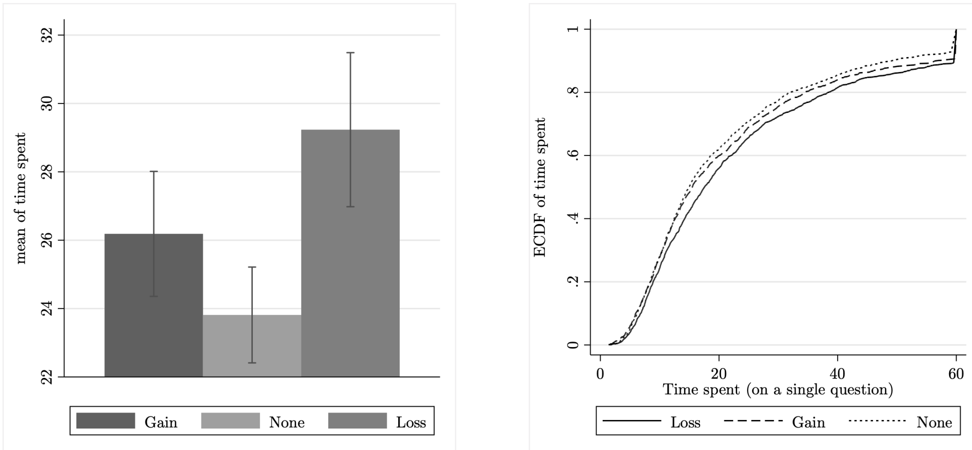
A total of 1210 people participated in the experiment, more or less evenly split between the two cognitive load treatments (609 with time constraints and 601 without) and the three incentive treatments (392 received gains, 406 received losses and 412 were not incentivized). In Table II, we see that our average participant was slightly more than 39 years old, 44% were female, 8% were Black, 7% were Asian, 13% identified as Hispanic, 13% held advanced degrees (masters or above), and 55% reported having an income above \$50k per year. Considering treatment balance on these observables, only 2 of the 21 between-column treatment comparisons were significant at the 10% level (two-sided t-tests). Further, an F-test for the joint significance of characteristics in predicting assignment to treatment is insignificant, which suggests we achieved randomization to treatment (regardless, we will present regression results with and without these controls).

Considering the choices that our participants made, the average participant spent 251 seconds completing the experiment and the treatments differed in this average completion time as expected. The participants in the no time constraints treatment spent an average of 311 seconds in the experiment while those participants in the time constraints treatment spend only 191 seconds, on average. On the CRT task, the overall average number of correct responses was 3.31 out of 7 and participants performed better without time constraints (3.81 correct answers versus 2.82 with time constraints).[5] Finally, people earned an average

[5]This rate of correct response of 47% (3.31/7) is somewhat larger than that reported in Toplak et al. (2014). They report a probability of correct response of 17 and 24% for the original 3 CRT questions and

of \$1.57 for their participation.

3.2 Incentives and cognition



(a) Mean Time Spent

(b) Empirical CDF of Time Spent

Figure 1. *Time spent answering single CRT questions for the three incentive treatments.*

In this section, we present our main results assessing whether losses trigger deliberative thinking. We begin our analysis by examining the extent to which our incentive treatments affected the amount of effort provided by our participants when there were no time constraints, measuring effort initially by the time spent on an individual CRT question and then by whether the individual correctly answered the question. On the left side of Figure 1, we report the differences in the mean time spent on each CRT question when there was not time constraint. Overall, we see that incentives matter – participants spent more time when they could earn gains or take losses. Combining treatments, the average time spent increases by almost 4 seconds or 16% when an incentive is provided ($t = 3.40$, $p < 0.01$), indicating participants are thinking longer about the questions when there are financial incentives. Splitting the incentive treatments, we see that losses differentially affected how long a participant thought about the questions. While the Gain treatment increased the average time spent by 2.37 seconds compared to the control ($t = 2.03$, $p = 0.04$), the Loss treatment increased the average by 5.42 seconds ($t = 4.00$, $p < 0.01$).

On the right of Figure 1, we examine the empirical cumulative distribution functions (ECDF) for the three incentive treatments (top coding at the 95 percentile or 60 seconds) and find that, because the ECDFs don't cross, the distributions have similar shapes (i.e., standard deviations) but different means. In fact, it is clear that the Loss treatment ECDF

the 4 new CRT questions they propose, respectively.

stochastically dominates both that of the None treatment and the Gain treatment. These conclusions are supported by Kolmogorov-Smirnov tests ($p < 0.01$ for both).

To more conservatively estimate the treatment effects of our incentive treatments, we regress the time spent on a question on incentive treatment indicators in Table III (having no reward is the omitted category). In the first column, we replicate what Figure 1 and the associated t-tests indicated – implementing gains increased the average time spent on a question by 2.37 seconds or 10% and implementing losses increased this time by 5.42 seconds or 23%. Here we also see that the difference in these point estimates is significant at the 5% level. In the second column, we add controls for race, income and education and find only small differences in our estimates as anticipated because our survey software correctly randomized participants to treatment. Finally, in the third column of Figure III, we account for the fact that each participant answered 7 of these questions by clustering our standard errors.[6] Accounting for the panel nature of our data does increase the standard errors of our estimates somewhat, the difference between Gain and Loss becomes somewhat weaker (p-value=0.136) but our main finding, that losses increase the amount of time spent answering questions (i.e., effort), survives. In sum, like the previous literature, losses induce decision-makers to devote more effort to the task.

	(1)	(2)	(3)
Gain	2.372**	2.271*	2.271
	(1.175)	(1.193)	(1.661)
Loss	5.420***	5.308***	5.308***
	(1.354)	(1.364)	(1.772)
Constant	23.813***	23.018***	23.018***
	(0.715)	(0.938)	(1.281)
Gain - Loss	-3.047**	-3.037**	-3.037
	(1.480)	(1.488)	(2.034)
Controls	No	Yes	Yes
Clustered Errors	No	No	Yes
Observations	4207	4207	4207

Dependent variable is time spent answering (in seconds).
Controls include race, income and education.
OLS with robust standard errors reported.
* p<0.10, ** p<0.05, *** p<0.01.

Table III. The effect of gains and losses on time spent answering.

We find that losses also elicit more effort in terms of participants being more likely

[6] An alternative to clustering is to collapse data to the subject level, which we investigate below.

to answer the CRT questions correctly. In Figure 2, we see that our participants were slightly better than equally likely to answer the average CRT question correctly. In fact, the likelihood of getting a question correct in the None treatment is 0.498 and adding any incentive increases this by 7 percentage points, on average ($t = 4.32$, $p < 0.01$). As with the time spent answering a question, when we separate the Gain and Loss treatments, we find that the participants who could suffer losses perform better. Here the effect of the Gain treatment is to increase the chance of answering correctly by 4.5pp ($t = 2.40$, $p = 0.02$) and the effect of the Loss treatment is to increase this chance by 9.3pp ($t = 5.02$, $p < 0.01$).

In Table IV, we estimate the treatment effects more carefully by regressing a correct answer indicator on the incentive conditions. Column (1) reproduces the results from Figure 2 and provides us an estimate for the treatment difference. Here we see that the Loss treatment effect is almost exactly twice as large as the Gain treatment effect ($p < 0.01$). In the second column, we confirm that adding demographic controls changes these estimates only slightly, as expected, and in the third column we find that clustering the standard errors on the participant level, results in just the main effect of the Loss treatment persisting, as was the case with the time spent answering a question. Based on this measure of cognitive effort we also find that losses cause participants to work harder.

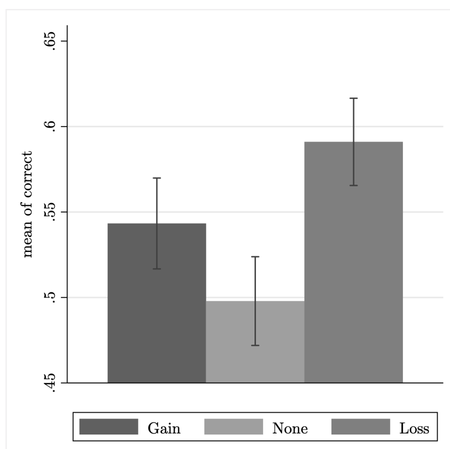


Figure 2. *The likelihood of answering CRT questions correctly for the three incentive treatments.*

At this point, we have two pieces of evidence suggesting that losses differentially elicit more effort from decision-makers: participants who face losses think longer and harder about the CRT problems. While these results are consistent with the previous literature discussed above on losses and decision effort, we are mostly interested in whether losses can flip the decision-maker from System 1 to System 2. Anticipating the difficulty of separating explanations based on cognitive effort from those based on cognitive process, we chose the CRT

	(1)	(2)	(3)
Gain	0.045**	0.049***	0.049
	(0.019)	(0.019)	(0.031)
Loss	0.093***	0.088***	0.088***
	(0.019)	(0.019)	(0.030)
Constant	0.498***	0.495***	0.495***
	(0.013)	(0.016)	(0.026)
Gain - Loss	-0.048***	-0.039**	-0.039
	(0.019)	(0.019)	(0.030)
Controls	No	Yes	Yes
Clustered Errors	No	No	Yes
Observations	4207	4207	4207

Dependent variable is a correct response indicator.
Controls include race, income and education.
OLS with robust standard errors reported.
* p<0.10, ** p<0.05, *** p<0.01.

Table IV. The effect of gains and losses on thinking correctly.

task so that we can also examine the frequency with which the incentive treatments elicit intuitive answers. Giving intuitive answer indicates being "stuck" in System 1.

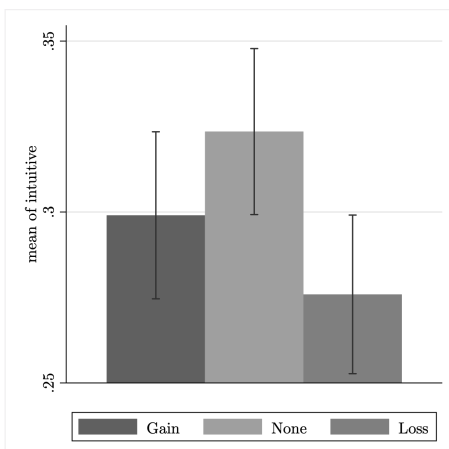


Figure 3. *The likelihood of answering CRT questions intuitively for the three incentive treatments.*

In Figure 3, we illustrate how the incentive treatments affect the likelihood of responding to a CRT question with the intuitive, but incorrect, answer. In this case, adding a monetary incentive reduces the probability of responding intuitively by 3.6pp, overall ($t = 2.44$, $p = 0.02$). As in both of the previous cases, the Loss treatment is more effective at reducing intuitive responses – it more effectively triggers System 2. The difference between the None and Gain treatments is just 2.5pp ($t = 1.39$, ns) but the difference between the None and Loss treatments is twice this magnitude, 4.8pp ($t = 2.78$, $p < 0.01$).

	(1)	(2)	(3)
Gain	-0.024	-0.026	-0.026
	(0.018)	(0.018)	(0.026)
Loss	-0.048***	-0.045***	-0.045*
	(0.017)	(0.017)	(0.025)
Constant	0.324***	0.319***	0.319***
	(0.012)	(0.015)	(0.021)
Gain - Loss	0.023	0.019	0.019
	(0.017)	(0.017)	(0.026)
Controls	No	Yes	Yes
Clustered Errors	No	No	Yes
Observations	4207	4207	4207

Dependent variable is an intuitive response indicator.
Controls include race, income and education.
OLS with robust standard errors reported.
* p<0.10, ** p<0.05, *** p<0.01.

Table V. The effect of gains and losses on thinking intuitively.

In Table V, we first confirm the results of our summary tests and see that the differential efficacy of losses to trigger System 2 thinking is robust to the addition of controls in column (2) and clustering the standard errors in column (3). In sum, our main results indicate that losses not only elicit more cognitive effort from decision makers, they also cause them to flip from fast and intuitive System 1 thinking to slower, more deliberative, System 2 thinking.

3.3 Robustness

To explore the robustness of our main results, we first aggregate the data for each individual and examine the total time spent solving CRT questions, the total number of questions our participants answer correctly and the total number of intuitive responses our participants give. We then look for heterogeneous treatment effects of our incentives among a subgroup of our participants known for relying a little too much on System 1 reasoning – young men. Lastly, we examine what happens to our treatment effects when we imposed a cognitive load by limiting the time participants are given to respond to each CRT question. In principle, if losses differentially trigger System 2 thinking and we impose a load on the working memories of our participants, making it much harder to switch to System 2, the treatment effects should vanish.

	(1)	(2)	(3)	(4)	(5)	(6)
Gain	16.607	15.898	0.087	0.096	-0.079	-0.078
	(11.523)	(11.694)	(0.061)	(0.060)	(0.084)	(0.084)
Loss	37.938***	37.159***	0.171***	0.163***	-0.159*	-0.151*
	(12.283)	(12.478)	(0.057)	(0.057)	(0.086)	(0.085)
Constant	166.691***	161.124***	1.249***	1.243***	0.817***	0.803***
	(6.390)	(9.021)	(0.045)	(0.051)	(0.056)	(0.069)
Gain - Loss	-21.331	-21.261	-0.084	-0.067	0.080	0.073
	(14.213)	(14.323)	(0.054)	(0.054)	(0.090)	(0.090)
Dep. Var.	Time	Time	Correct	Correct	Intuitive	Intuitive
Controls	No	Yes	No	Yes	No	Yes
Observations	601	601	601	601	601	601

Dependent variable varies between total time spent, number of correct responses and number of intuitive responses. Controls include race, income and education. OLS or negative binomial with robust standard errors reported. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table VI. Treatment effects using aggregated data.

An analysis of the aggregated data can be found in Table VI. In the first two columns we examine the total time spent answering CRT questions, first without controls and then with them. In both cases, we see that aggregating the data has two effects. First, the point estimate on the Gain treatment – spending 16.61 seconds more than in the None treatment – is no longer significant. Second the point estimate on the Loss treatment – an additional 37.94 seconds above what is spent in the control – remains highly significant, even when the covariates are added. Columns (3) and (4) report the results of analyzing the number of correct responses to the CRT questions. Again, we find in these two columns that regardless of whether controls are added to the regression or not, the effect of the Gain treatment is no longer significant while the sizable Loss treatment effect remains highly significant. In the final two columns of Table VI, we see the same pattern as in the previous columns. When considering the number of intuitive responses that participants give, The Gain treatment effect loses significance in the aggregated data but the Loss effect remains significant at least at the 10% level. The punchline of this analysis is that only the Loss treatment effects are robust to aggregating the data across individual participants. Further, we continue to find evidence that losses differentially trigger System 2 thinking.

Mounting evidence suggests that brains develop slowly and differentially by sex (Cowell et al., 2007). It is now thought that young men, in particular, have trouble with planning and

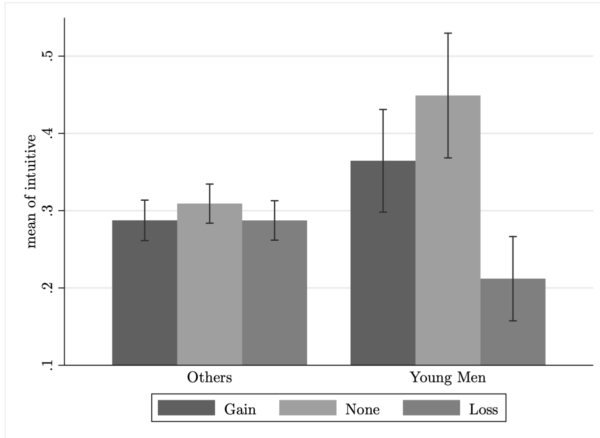


Figure 4. *The likelihood of answering CRT questions intuitively separately for young men.*

other executive functions associated with working memory and the prefrontal cortex (Paus, 2005; Kaller et al., 2012). Based on these findings, we hypothesized that if losses truly help decision-makers switch into System 2, their effects should be more pronounced among people who might find this transition difficult. That is, if older people, for example, are predisposed to a higher likelihood of System 2 thinking, or who switch to System 2 more readily, they may not need the extra stimulus provided by losses (required by younger people), to access System 2.

To assess whether the effects of losses on switching to System 2 are more pronounced among young men, we split the sample into two groups, men under the age of 30 (16% of the sample) and the rest. In Figure 4, we see that the treatment effects we found in the previous section persist among the "others," but become more symmetrical. Among the other participants on the left of the figure, gains reduce the number of intuitive responses by 2.2pp and losses reduce them by 2.2pp as well. In this case, however, these more modest differences are not significant. The results differ for young men. Here, on the right of Figure 4, we see that young men respond more intuitively that the others, overall. Considering the treatment differences for young men, we find that gains reduce the frequency of intuitive responses by 8.4pp ($t = 1.59$, $p = 0.11$) and losses reduce intuitive responses by 24pp ($t = 4.95$, $p < 0.01$). Put differently, the treatment effect of losses is 12 times greater among young men than it is in the rest of the population, exactly what we hypothesized.

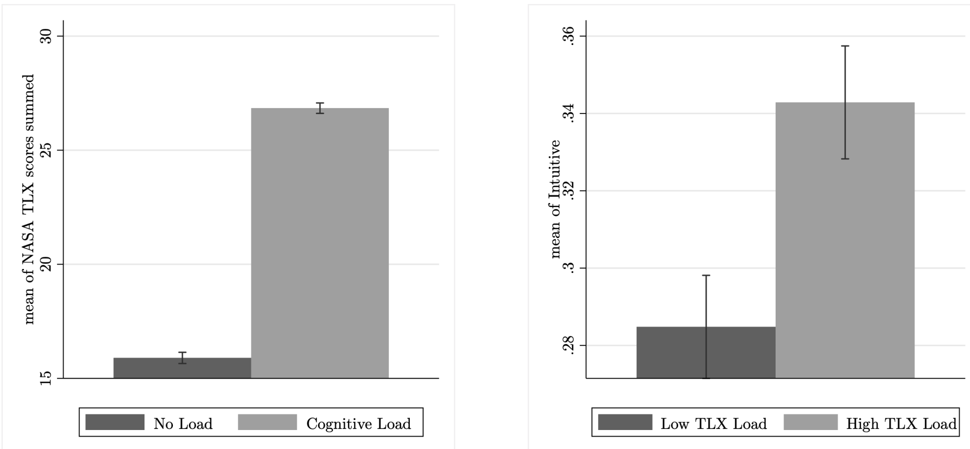
Combining all the data in a regression setting with interactions, if Table VII, we find that among the other participants, the Gain treatment effect of reducing intuitive responses by 2.4pp is significant at the 10% level but the effect of the Loss treatment is a fairly precisely zero. Further when we include our controls for race, income and education, neither treatment effect is significant for the other participants. The interaction of being young, male and being

	(1)	(2)
Gain	-0.024*	-0.025
	(0.013)	(0.019)
Loss	0.004	-0.018
	(0.013)	(0.018)
Young Man	0.135***	0.137***
	(0.032)	(0.043)
Gain × Young Man	-0.026	-0.052
	(0.041)	(0.057)
Loss × Young Man	-0.207***	-0.219***
	(0.040)	(0.053)
Constant	0.312***	0.305***
	(0.009)	(0.015)
Gain + (Gain × Young Man)	-0.049	-0.076
	(0.039)	(0.053)
Loss + (Loss × Young Man)	-0.202***	-0.238***
	(0.009)	(0.049)
Controls	No	Yes
Observations	4207	4207

Dependent variable is an intuitive response indicator.
Controls include race, income and education.
OLS with robust standard errors reported.
* $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table VII. The differential effect of losses on thinking intuitively among young men.

in the Loss treatment is highly significant however, regardless of what controls are added. Near the bottom of Table VII, we report the estimated treatment effects of the young men. Though sizable (between 4.9pp and 7.6pp), the estimates of the Gain treatment effect are not significant for young men but the estimated effect of the Loss treatment is highly significant for the young men and varies from 20.2pp to 23.8pp.



(a) Load effect on experienced effort (b) Effort effect on intuitive answers

Figure 5. *Manipulation check: does the load impose cognitive costs and affect System 2 usage?*

As a final robustness test, we ran the experiment a second time and added a 20-second time constraint to each CRT question as a cognitive load. Prior literature has highlighted that time constraints are an essential part of any cognitive load (e.g. Kalyuga (2011)). Imposing a cognitive load on System 2 should prevent subjects from being able to access it regardless of the incentive treatment. As a result, we expect the Loss treatment to lose its differential ability to trigger System 2 in this second experiment when it is, more or less, unavailable. We begin our analysis of the data from this second experiment by testing whether the time constraint manipulation had the intended effect: to load up working memory and make the task seem more effortful. On the left of Figure 5, we see that imposing the time constraint cognitive load increased the NASA TLX work load scores reported by our participants dramatically. Specifically, we sum the subjective responses (reported on a 0-10 scale) that participants report to the questions rating mental, temporal, performance and effort demand and compare these sums between participants that did and did not experience time constraints (using factor analysis to aggregate the individual questions works just as well and revealed that most of the value from the scale came from just these four items). The average load experienced by participants in the no time constraints experiment is relatively mild – 15.90 out of a possible 40 – but the load reported by the participants that answered the

CRT questions with time constraints is significantly more severe – 26.84 out of 40 ($t = 24.15$, $p < 0.01$). This difference suggests that the time constraints did impose a cognitive load on participants. To corroborate this inference, in the right panel of Figure 5, we see that participants who report higher than median NASA TLX scores were 5.8pp more likely to respond to CRT questions with the intuitive response ($t = 5.77$, $p < 0.01$).

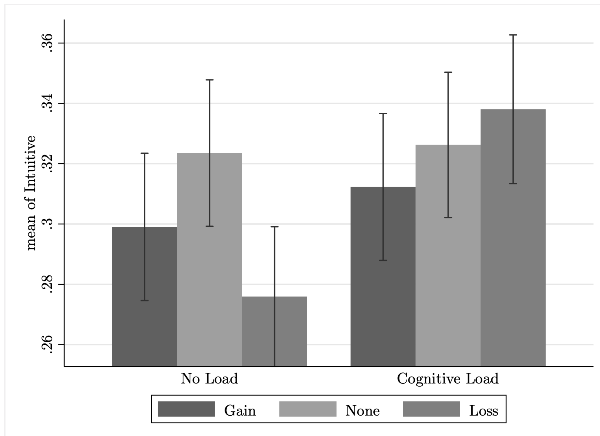


Figure 6. *A cognitive load attenuates the treatment effects.*

Comparing between cognitive load treatments, we see the anticipated results in Figure 6. On the left, we reproduce our main results from the initial experiment, without any cognitive load (i.e., Figure 3), in which both gains and losses affect the frequency of intuitive CRT responses but losses reduce the frequency twice as much. On the right of Figure 6, we illustrate the results of the second experiment with a cognitive load. Here we find that the Gain treatment reduces intuitive responses by just 1.4pp ($t = 0.80$, ns) and the Loss treatment actually increases intuitive responses slightly by 1.2pp ($t = 0.67$, ns). However, none of the these treatment differences are significant.

In Table VIII, we combine the data and estimate treatment effects for both experiments. As in Table V, we see that the Loss treatment significantly reduces the chance that a participant will respond intuitively (4.8pp or 4.6pp depending on whether controls are added or not) but the Gain treatment does not affect intuitive responses significantly. Adding the point estimate on the correct interaction term at the bottom of Table VIII, reveals that the loss treatment no longer triggers System 2 when it is unavailable because of a cognitive load. In other words, our results indicate that the prospect of losses does not just elicit more effort from decision makers, it helps them transition to deliberative System 2 thinking.

	(1)	(2)
Gain	-0.024	-0.025
	(0.018)	(0.018)
Loss	-0.048***	-0.046***
	(0.017)	(0.017)
Time constraint	0.003	0.004
	(0.017)	(0.017)
Gain × Time	0.011	0.015
	(0.025)	(0.025)
Loss × Time	0.059**	0.054**
	(0.025)	(0.025)
Constant	0.324***	0.328***
	(0.012)	(0.014)
Gain + (Gain × Time)	-0.014	-0.011
	(0.017)	(0.017)
Loss + (Loss × Time)	0.012	0.009
	(0.017)	(0.017)
Controls	No	Yes
Observations	8470	8470

Dependent variable is an intuitive response indicator.
Controls include race, income and education.
OLS with robust standard errors reported.
* $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table VIII. The differential effect of a cognitive load on thinking intuitively.

4 Discussion

The dual system approach has been an extremely influential conceptual framework in understanding decision making. However, little is known about what factors cause individuals to shift between Systems 1 and 2. Our experiment, designed to specifically examine System 1 and 2 thinking, highlights that the framing of rewards is an important factor in determining whether subjects provide intuitive responses (System 1) or engage in more effortful, contemplative thinking (System 2).

Since dual system thinking has been shown to be correlated with biased decision making, our results may be helpful in highlighting reward mechanisms which nudge individuals to towards System 2 thinking and, thus, better outcomes. Our results also align nicely with the neuro-imaging literatures which highlight the differential neural processing of gains and losses in regions of the brain which also moderate decision making and attention. Interesting follow-up work would be to examine the neural activity related to System 1 and System 2 thinking, and our results highlight an experimental mechanism to differentially elicit these types of decision making.

References

Alter, A. L., D. M. Oppenheimer, N. Epley, and R. N. Eyre (2007): "Overcoming intuition: metacognitive difficulty activates analytic reasoning." *Journal of experimental psychology: General*, 136, 569.

Amiez, C., J.-P. Joseph, and E. Procyk (2005): "Anterior cingulate error-related activity is modulated by predicted reward," *European Journal of Neuroscience*, 21, 3447–3452.

Arsalidou, M., J. Pascual-Leone, J. Johnson, D. Morris, and M. J. Taylor (2013): "A balancing act of the brain: Activations and deactivations driven by cognitive load," *Brain and behavior*, 3, 273–285.

Bago, B. and W. De Neys (2017): "Fast logic?: Examining the time course assumption of dual process theory," *Cognition*, 158, 90–109.

Bourgeois-Gironde, S. and J.-B. Van Der Henst (2009): "How to open the door to System 2: Debiasing the bat-and-ball problem," in *Rational animals, irrational humans*, ed. by S. Watanabe, A. P. Blaisdell, L. Huber, and A. Young, Keio University, chap. 14.

Chen, X., S. Voets, N. Jenkinson, and J. M. Galea (2020): "Dopamine-dependent loss aversion during effort-based decision-making," *Journal of Neuroscience*, 40, 661–670.

Clay, S. N., J. A. Clithero, A. M. Harris, and C. L. Reed (2017): "Loss aversion reflects information accumulation, not bias: A drift-diffusion model study," *Frontiers in psychology*, 8, 1708.

Cowell, P. E., V. A. Sluming, I. D. Wilkinson, E. Cezayirli, C. A. Romanowski, J. A. Webb, S. S. Keller, A. Mayes, and N. Roberts (2007): "Effects of sex and age on regional prefrontal brain volume in two human cohorts," *European Journal of Neuroscience*, 25, 307–318.

Deck, C., S. Jahedi, and R. Sheremeta (2021): "On the consistency of cognitive load," *European Economic Review*, 134, 103695.

Foti, D., A. Weinberg, E. M. Bernat, and G. H. Proudfit (2015): "Anterior cingulate activity to monetary loss and basal ganglia activity to monetary gain uniquely contribute to the feedback negativity," *Clinical Neurophysiology*, 126, 1338–1347.

Frederick, S. (2005): "Cognitive reflection and decision making," *Journal of Economic perspectives*, 19, 25–42.

Fujiwara, J., P. N. Tobler, M. Taira, T. Iijima, and K.-I. Tsutsui (2009): "Segregated and integrated coding of reward and punishment in the cingulate cortex," *Journal of Neurophysiology*, 101, 3284–3293.

Gehring, W. J. and A. R. Willoughby (2002): "The medial frontal cortex and the rapid processing of monetary gains and losses," *Science*, 295, 2279–2282.

Gillard, E., W. Van Dooren, W. Schaeken, and L. Verschaffel (2009): "Proportional reasoning as a heuristic-based process: time constraint and dual task considerations," *Experimental Psychology*, 56, 92–99.

James, W. (1890): *The principles of psychology*, Henry Holt and Company.

Kahneman, D. (2011): *Thinking, fast and slow*, Macmillan.

Kaller, C. P., K. Heinze, I. Mader, J. M. Unterrainer, B. Rahm, C. Weiller, and L. Köstering (2012): "Linking planning performance and gray matter density in mid-dorsolateral prefrontal cortex: moderating effects of age and sex," *Neuroimage*, 63, 1454–1463.

Kalyuga, S. (2011): "Cognitive load theory: How many types of load does it really need?" *Educational Psychology Review*, 23, 1–19.

Lejarraga, T., M. Schulte-Mecklenbeck, T. Pachur, and R. Hertwig (2019): "The attention–aversion gap: how allocation of attention relates to loss aversion," *Evolution and Human Behavior*, 40, 457–469.

Paus, T. (2005): "Mapping brain maturation and cognitive development during adolescence," *Trends in cognitive sciences*, 9, 60–68.

Sallet, J., R. Quilodran, M. Rothé, J. Vezoli, J.-P. Joseph, and E. Procyk (2007): "Expectations, gains, and losses in the anterior cingulate cortex," *Cognitive, Affective, & Behavioral Neuroscience*, 7, 327–336.

Smith, V. L. and J. M. Walker (1993): "Monetary rewards and decision cost in experimental economics," *Economic Inquiry*, 31, 245–261.

Toplak, M. E., R. F. West, and K. E. Stanovich (2014): "Assessing miserly information processing: An expansion of the Cognitive Reflection Test," *Thinking & Reasoning*, 20, 147–168.

Wallis, J. D. and S. W. Kennerley (2011): "Contrasting reward signals in the orbitofrontal cortex and anterior cingulate cortex," *Annals of the New York Academy of Sciences*, 1239, 33–42.

Westbrook, A. and T. S. Braver (2015): "Cognitive effort: A neuroeconomic approach," *Cognitive, Affective, & Behavioral Neuroscience*, 15, 395–415.

Appendix A Experiment Instructions

After the subjects completed an experiment consent form and filled out demographic information, they were prompted with the following incentive information regarding payments for correct answers.

You will now be awarded a bonus to answer seven questions. The amount of bonus money you will earn in this HIT depends on your performance. To start, you have been awarded $1.00. However, for **each** correct answer you will **gain** an additional $0.25.

As an attention check, if you get one question correct, how much will you be awarded? (If you answer incorrectly the survey will end.)

- ○ $0.50
- ○ $1.25
- ○ $5.00

Figure A.1. *Incentive prompt for Gain treatment.*

You will now be awarded a bonus to answer seven questions. The amount of bonus money you will earn in this HIT depends on your performance. To start, you have been awarded $2.75. However, for **each** incorrect answer you will **lose** $0.25.

As an attention check, if you get **one** question incorrect, how much will you be awarded? (If you answer incorrectly the survey will end.)

- ○ $1.00
- ○ $2.50
- ○ $8.00

Figure A.2. *Incentive prompt for Loss treatment.*

You will now be awarded a bonus to answer seven questions. The amount of bonus money you will earn in this HIT does not depend on your performance. You will be awarded $1.00 for completing the questions.

As an attention check, how much will you be awarded for completing the questions? (If you answer incorrectly the survey will end.)

- ○ $0.50
- ○ $1.00
- ○ $2.00

Figure A.3. *Incentive prompt for No Reward treatment.*

After subjects received on of the prompts above, they were sequentially presented with the CRT questions in Table I and the order of questions was the same for each subject.

After completing the CRT questions, the subjects were prompted with the following NASA TLX questions:

Figure A.4. *NASA TLX Questions*