Supplemental Materials

1. APPENDIX A: STUDY STIMULI
2. APPENDIX B: PRE-TEST RESULTS
3. APPENDIX C: DATA CLEANING PROTOCOL
4. APPENDIX D: FULL DETAILS OF HIERARCHICAL REGRESSIONS
5. APPENDIX E: STUDIES USED IN META-ANALYSIS
6. APPENDIX F: ROBUSTNESS CHECKS USING META-ANALYSIS DATA
7. APPENDIX G: NON-PARAMETRIC MODEL FOR ESTIMATIMNG MOMENTARY POST-REWARD ENGAGEMENT REDUCTION
8. APPENDIX H: EFFECT OF TEMPORARY INCENTIVES ON ACCURACY
9. AND NET OUTCOME
10. APPENDIX I: MODERATION BY INITIAL MOTIVATION
11. APPENDIX J: HETEROGENEITY IN POST REWARD BEHAVIOR
12. APPENDIX K: ADDITIONAL CHARTS FOR ALL STUDIES
13. APPENDIX L: STUDY WITH PAYING FOR A LEISURE TASK
14. APPENDIX M: STUDY WITH FRAMING BOTH CHOICE OPTIONS AS IMPORTANT
15. APPENDIX N: ALTERNATIVE PARAMETERIZATION TO ESTIMATE INITIAL REDUCTION IN ENGAGEMENT IN THE POST-REWARD PERIOD
16. APPENDIX O: RESULTS OF FIELD STUDIES THAT HAVE MEASURED BEHAVIOR AFTER CONTINGENT INCENTIVES ENDED

APPENDIX A: STUDY STIMULI

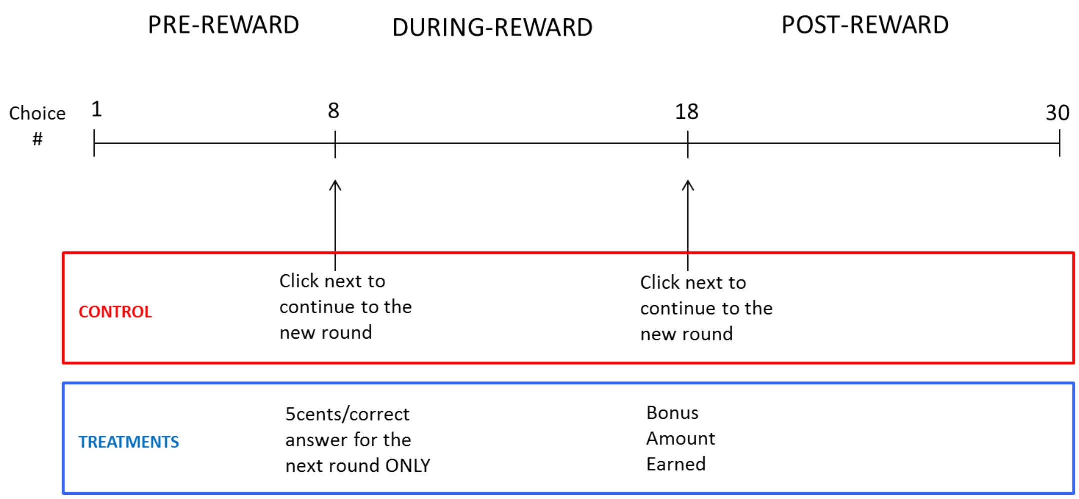


Exhibit A.1: The exhibit shows how the series of 30 repeated choices (or trials) were organized in all the studies. The trials were divided into three rounds for all the experimental groups, with 8, 10, and 12 trials in the pre-, during-, and post-reward rounds respectively. The reward group(s) learnt about the total rewards earned before the start of the third round. However, all payments were made at the very end of the study, and *not* after the immediate end of the second round.

PLEASE READ THE INSTRUCTIONS CAREFULLY.

In this survey you will be a given a series of **choices between doing cognitive tasks and watching videos of interesting television advertisements** collected from across the world.

The cognitive task will train your mental reasoning skills, and we will use your results to calibrate and standardize a training test. **You can do as many of them as you want**, or can just enjoy the videos.

A typical cognitive task in this survey looks like this.

The task will require **searching and selecting two numbers in a grid such that they add up to 10**. You can select a number by clicking on the box containing the number. An example is shown below.

8.63	4.38	2.68
5.72	1.67	7.38
7.32	3.62	1.29
7.02	5.17	1.62

Note:

1. **<u>All</u>** cognitive tasks in this study have an unique pair of answer.

2. You should **select ONLY two numbers in a grid**, and no more, as shown above.

You will get **30 seconds to complete each such tasks** after which the survey will automatically advance to the next screen. If you are done with the task before the time limit, you can click NEXT to proceed.

Alternatively, you can choose to watch a video clip for the same duration, and click NEXT when you are done viewing.

Click NEXT to see a typical video clip we have in this survey. This will also help test if the video loads properly in your browser.

NOTE: These clips are from different years and various countries, and therefore some of them might not be of very high video quality. But you should be able to view them without any issues.

Exhibit A.2: Four-stage instructions provided to all participants before the start of all the studies. Participants were informed that they could do as many of the target tasks (i.e., math tasks) as they wanted, or enjoy the videos during the whole study. They were also informed that all the cognitive tasks had one solution. Finally, they were informed that both tasks were of the same total duration. All participants were shown a sample video and asked if they were able to view and hear it properly - the experiment was aborted for those who reported having a problem.

Task Type = Writing

Please indicate your choice of a topic below.

- ○ Your favorite genre of Music
- ○ Your idea of favorite vacation

Please indicate your choice of a topic below.

- ○ Why are social media sites so important in today's popular culture?
- ○ Why are tattoos so popular in today's society?

Please indicate your choice below about which you would like to give your opinion.

- ○ Should kids be given smartphones?
- ○ Should the minimum age for teenagers to get a driver's license be increased?

Task Type = Brand-name Matching

Please choose the product category for which you would like to do the matching between logos and brand names.

- ○ Automobile
- ○ Sports Good

Please choose the product category for which you would like to do the matching between logos and brand names.

- ○ Oil and Gas Companies
- ○ Luxury and Accessories

Please choose the product category for which you would like to do the matching between logos and brand names.

- ○ Technology Companies
- ○ Banks and Financial Institutions

Exhibit A.3: The exhibit shows the types of tasks used for unrelated activities break in Study 2. Participants did all the three tasks of a particular type (Writing or Brand-name Matching), and each one of these tasks had a time limit of 30 seconds. The exhibit shows the choice condition. In the no-choice condition, one of the options was randomly pre-selected for the participant.

BONUS INFORMATION: PLEASE READ CAREFULLY.

In a previous version of this survey were able to pay as much as 5 cents for every correct answer, but in this version **we are unable to pay more**.

You will get 1 cent for every cognitive task that you answer correctly.

BONUS INFORMATION: PLEASE READ CAREFULLY.

In a previous version of this survey were able to pay only 5 cents for every correct answer, but in this version **we are able to pay a LOT more**.

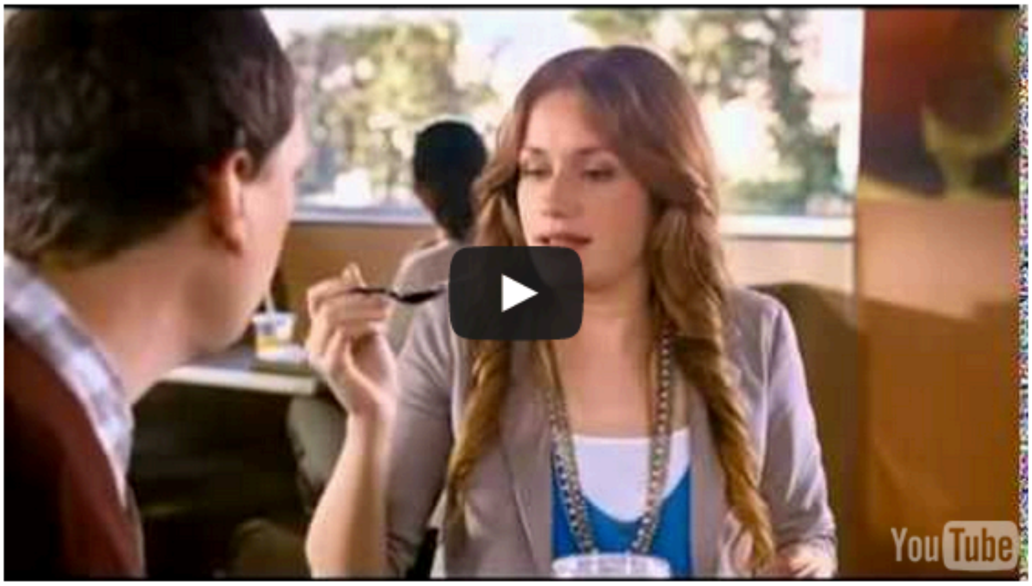
You will get 50 cents for every cognitive task that you answer correctly.

Exhibit A.4: The exhibit shows the manipulations used to make the perceived reward magnitude salient in Study 3.

Target Task = Math
PLEASE READ THE INSTRUCTIONS CAREFULLY. In this survey you will be asked to do a task. The task is to **solving cognitive math tasks**. WE WILL USE YOUR RESPONSES TO DESIGN EXPERIMENTAL STIMULI FOR A SPATIAL REASONING STUDY. Since doing the task can be tiring, you will also have an option of a different task, **evaluating videos of television advertisements,** so that you can take a break. It is completely up to you to choose which task you want to do in each round.
Target Task = Video
PLEASE READ THE INSTRUCTIONS CAREFULLY. In this survey you will be asked to do a task. The task is to **evaluate videos of television advertisements**. WE WILL USE YOUR RESPONSES TO DESIGN EXPERIMENTAL STIMULI FOR AN ATTENTION AND PERCEPTION STUDY. Since doing the task can be tiring, you will also have an option of a different task, **solving cognitive math problems,** so that you can take a break. It is completely up to you to choose which task you want to do in each round.

Exhibit A.5: The exhibit shows the two types of framing used for the math (work) and the video (leisure) task in Study 4.

Please watch the video and rate it to indicate how much you liked it.



Your rating ☆☆☆☆☆

Exhibit A.6: The exhibit shows a typical video *task* used in Study 4. Unlike other studies (Study 1- 3), in this study the participants in the target-task = video condition were asked to rate the video in order to get their rewards.

PLEASE READ CAREFULLY.

In the next round instead of making repeated choices between doing a cognitive task and watching a video, **you will need to choose now what you want to do next during the ENTIRE round.**

If you choose Cognitive Tasks, you will be presented with ONLY Cognitive Tasks in the next round.

If you choose Videos, you will be presented with ONLY Videos in the next round.

Please indicate what would you like to do during the ENTIRE NEXT ROUND.

○ Do only Cognitive Tasks

○ Watch only Videos

Exhibit A.7: The exhibit shows the instruction given to participants at the end of Round2 in Study 5. Participants were asked to choose which task they would like to do during the entire ensuing period of Round 3.

APPENDIX B: PRE-TEST RESULTS

Task Pretest (Math task, Video task)

A pretest (N=47) was done to examine how people felt about the math task (i.e., the target task) and the video task (i.e., the alternative task). A random sample of participants was chosen from the same population and they judged the two tasks on several attributes. Participants judged the math task as relatively more work-like compared to the video ($M_{math} = 6.87, SD = 2.19$ $vs. M_{video} = 2.49, SD = 1.96; t(46) = 9.77, p < .001$), on 9-point scales, but considered the video more leisure-like compared to the math ($M_{math} = 3.89, SD = 2.63$ $vs. M_{video} = 6.78, SD = 2.37; t(46) = 6.24, p < .001$). The math task was also judged as relatively more effortful ($M_{math} = 5.59, SD = 2.44$ $vs. M_{video} = 1.95, SD = 1.52; t(46) = 9.05, p < .001$) and less entertaining compared to the video ($M_{math} = 4.64, SD = 2.32$ $vs. M_{video} = 6.43, SD = 2.10; t(46) = 4.26, p < .001$).

Participants felt that the math task had more long-term benefits whereas the videos had higher immediate benefits($M_{math} = 5.38, SD = 1.97$ $vs. M_{video} = 3.97, SD = 1.65; t(46) = 3.65, p < .001$). Participants also felt that more justification (on a 1 = *Less* to 9 = *More* scale) was needed for choosing to watch the video task over doing the math tasks, than for the opposite choice($M_{choose\ math} = 2.70, SD = 2.28$ $vs. M_{choose\ video} = 4.72, SD = 2.97; t(46) = 3.53, p < .001$).

Most importantly, both the tasks satisfied the pre-condition required of tasks that can be deemed appropriate for testing theories of intrinsic motivation (Deci, Koestner, & Ryan, 1999). Both task had a rating of higher than the mid-point on a scale measuring how "interesting and enjoyable" the task is (1=*Low*, 9=*High*): math task ($M_{math} = 6.02, SD = 2.19$) and the video task ($M_{video} = 6.64, SD = 2.08$).

The question texts and the scales used are shown below:

Questions Texts	Scale
To what extent did you find the cognitive task interesting and enjoyable?	Extremely uninteresting and unenjoyable (1), (2), (3), (4), Neutral (5), (6), (7), (8), Extremely interesting and enjoyable (9)
To what extent did you find the video interesting and enjoyable?	
To what extent would you describe doing the cognitive task as a work activity?	NOT at all (1), (2), (3), (4), Neutral (5), (6), (7), (8), VERY MUCH like a work activity (9)
To what extent would you describe watching the video as a work activity?	
To what extent would you describe doing the cognitive task as a leisure activity?	
To what extent would you describe watching the video as a leisure activity?	
To what extent did you find the cognitive task effortful and tiresome?	NOT effortful and tiresome at all (1), (2), (3), (4), Neutral (5), (6), (7), (8), EXTREMELY

To what extent did you find the video effortful and tiresome?	effortful and tiresome (9)
To what extent did you find the cognitive task entertaining and relaxing?	Extremely unentertaining and unrelaxing (1), (2), (3), (4), Neutral (5), (6), (7), (8), Extremely entertaining and relaxing (9)
To what extent did you find the video entertaining and relaxing?	
To what extent do you believe that doing the cognitive task has short-term versus long-term benefits?	Has immediate benefits but little long-term benefits (1), (2), (3), (4), Neutral (5), (6), (7), (8), Has long-term benefits but little immediate benefits (9)
To what extent do you believe that watching the video has short-term versus long-term benefits?	
Imagine you had to make a choice between doing the cognitive tasks that will train your mental reasoning skills or watching the videos of interesting television advertisements. To what extent would you feel the need to justify to yourself choosing to do the cognitive tasks?	I will NOT need any justification at all (1), (2), (3), (4), Neutral (5), (6), (7), (8), I will need a LOT of justification (9)
Imagine you had to make a choice between doing the cognitive tasks that will train your mental reasoning skills or watching the videos of interesting television advertisements. To what extent would you feel the need to justify to yourself choosing to watch the videos?	

Breaks Pretest (Study 2)

Fifty-two participants from the same population as Study 2 participated in the online pretest. Participants were either shown the two versions of the writing task (with and without choice on the topic of writing), or the two versions of the logo-matching task (with and without choice on the product category for which the brand names and the logos were required to be matched).

After reading about these tasks, participants indicated which of the versions they thought required more thinking, required more work, was more difficult, or provided more autonomy. For each of the questions, participants were also given the option to indicate if they could not say whether the two versions were different on the attribute in question.

Across these various attributes, the proportion of participants who chose the "Can't Say" option varied from 7% ("required more thinking") to 19% ("granted more autonomy"). We coded this data as not-available. Among participants who gave an answer, 73% indicated that the version with choices required more thinking ($\chi^2(1) = 10.08$, $p = .001$), 68% indicated that the version with choices required more work ($\chi^2(1) = 6.72$, $p = .009$) and 66% indicated that the version with choices was more difficult ($\chi^2(1) = 5.00$, $p = .025$). However, 64% of the participants indicated that the task version with choices gave them more autonomy than the version without choices ($\chi^2(1) = 3.43$, $p=.064$). Therefore, the participant population found the task versions with choices more effortful, even though they felt it granted them marginally more autonomy.

The question text and the measurement instrument used are shown below:

How would you compare the two tasks on the following dimensions? Please remember these are relative judgments. That is you are indicating, of these two tasks, which description fits one of them better. However, if you are completely unsure please indicate CAN'T SAY.			
	Task without Choice	**Task with Choice**	**Can't Say**
Requires more thinking			
Requires more work			
Is more difficult			
Grants more autonomy			

APPENDIX C: DATA CLEANING PROTOCOL

Every study started with an initial sampling of both types of tasks (math, video) after which participants were asked if they faced any technical problems. If a problem was reported, the study was aborted and data from these participants were discarded from further analysis.

Our experimental paradigm was specifically designed to capture dynamic changes in behavior over time, and could distinguish between temporary and permanent disengagements. A temporary decrease in motivation to do the target task would be reflected in the participant choosing to watch the video for a few trials before choosing to do the math task again. A more persistent decrease in motivation to do the target activity could be reflected in two ways. Participants could "quit" within the study, by repeatedly deciding to only watch the videos for the remaining duration. Alternatively, participants could quit by ending the study part way through and not completing the remaining trials. We tracked all dropouts, and included participants who dropped-out of the study after completing the pre-reward baseline period coding their participation as zero for the target task. The reward for the math task to the treatment group was announced at the end of the pre-reward baseline period, and therefore this analysis strategy ensured that we included anyone whose behavior could have been impacted by the incentives, whether they finished the study or not.

Participant's data containing duplicate IP addresses were removed prior to analysis. Finally, an attention check question was administered at the end of the survey, and data from participants who reached till the end of the survey but failed the attention check were discarded prior to analysis. Participants who quit part way through and therefore did not answer the attention check question were given the benefit of doubt and were included in the analysis.

APPENDIX D: FULL DETAILS OF HIERARCHICAL REGRESSIONS

Model for Momentary post-reward engagement reduction

We capture total momentary post-reward engagement reduction using a functional form assumption about how effort returns to baseline over time in the post-reward period. Assuming a non-linear return of effort (i.e., likelihood of choosing the math task) to baseline over time (the number of periods *t* since the incentive ended), we parameterize momentary post-reward engagement reduction (MOMENTARY) as:

$$MOMENTARY_t = \frac{1}{t} \tag{1}$$

Using this parameterization[1], the probability of individual i choosing to do the math task in post-reward (Round 3) during trial *t* can be written as:

$$P(Y_{ti} = 1) = \phi\ (\beta_{0i} + \beta_{Mi}\ MOMENTARY_t) \tag{2}$$

In our model we set ϕ to the logit link function and β_{0i} is a person-specific intercept and β_{Mi} is a person-specific momentary post-reward engagement reduction behavior. In the hierarchical regression the parameters in Equation (2) are a function of time-invariant individual-level covariates, to account for the repeated observations per person.

$$\beta_{0i} = \beta_{00} + \beta_{01}\ C_i + \beta_{02}\ X_i + u_{0i} \tag{3}$$

The person-specific baseline parameter β_{0i} is a function of the condition that individual *i* has been randomly assigned to experimental condition C_i, the total number of math task choices by individual i in the pre-incentive Round 1 X_i, as well as the population baseline β_{00} and time-invariant person-specific error term u_{0i}.

$$\beta_{Mi} = \beta_{10} + \beta_{11}\ C_i + u_{1i} \tag{4}$$

The person-specific momentary post-reward engagement reduction behavior β_{Mi} is estimated as a function of experimental condition C_i, as well as the baseline β_{10} and the individual-specific error term u_{1i}. The random effects for the intercept and the slope for every individual i, u_{0i}, u_{i1}, are assumed to be bi-variate normal with zero-mean, variances τ_{00}, τ_{11} and common co-variance τ_{01}. This error structure accounts for the potentially correlated repeated-measures for each individual. Combining equations, (2), (3), and (4) yields an "intercepts and slopes-as-outcomes" model (Raudenbush and Bryk, 2002).

The expected proportion of math tasks chosen in each trial *t* of Round 3 is:

---

[1] We report robustness analysis with other parameterization in Appendix D of the Supplemental Materials. We also report results using a more flexible non-parametric approach in Appendix E.

$$P(Y_{ti}=1)=\phi\,(\beta_{00}+\ \beta_{10}MOMENTARY_t+\ \beta_{01}C_i+\beta_{02}X_i+\ \beta_{MOMENTARY}C_i * MOMENTARY_t) \quad (5)$$

The coefficient β_{01} is renamed to β_{LT} in equation (5). The coefficient $\beta_{MOMENTARY}$ in equation (5), which is β_{11} from equation (4) renamed for ease of exposition, tests for a difference in the extent of momentary post-reward engagement reduction in the experimental condition ($C_i = 1$), compared to the corresponding time periods in the control condition ($C_i = 0$). A significant and negative $\beta_{MOMENTARY}$ generally indicates momentary post-reward engagement reduction after the incentive ended, compared to the corresponding trials in the control condition, controlling for individual differences in baseline effort X_i. An important exception to such an interpretation arises when there is an overall *increase* in effort of the reward group relative to the control group in the post-reward period as indicated by a significant β_{LT}. In this case a significant negative $\beta_{MOMENTARY}$ might indicate an immediate decrease in the effort of the reward group relative to its longer-run steady-state level, but not a momentary post-reward engagement reduction relative to the contemporaneous control group.

In a similar vein, an estimate of $\beta_{MOMENTARY}$ that is not statistically distinguishable from zero represents a consistent level of effort throughout Round 3, with two very different potential interpretations. A non-significant $\beta_{MOMENTARY}$ could indicate that no post-reward reduction in engagement has occurred or it could represent a consistent overall increase or decrease in engagement in the reward group in Round 3. Hence, it will be important to also estimate the overall effects of incentives on choices in Round 3, in addition to momentary reduction in engagement behavior and interpret these parameters jointly. Next, we describe the tests we use to estimate the overall effects of incentives.

Difference-in-Difference Model for Overall Effects

We use a hierarchical non-linear difference-in-difference model to estimate differences in the overall probability of choosing the math task between two experimental conditions $C_i = 0$ or $C_i = 1$ and between two experimental rounds $R_t = 0$ or $R_t = 1$. The general specification can be written as follows:

$$P(Y_{ti}=1)=\phi\,(\beta_0+\ \beta_1 *\ R_t+\ \beta_2 * C_i+\beta_3 * R_t * C_i\,) \quad (6)$$

The interpretation of the key coefficient β_3 depends on how the rounds (R_t) and conditions (C_i) are coded. To estimate the effect of incentives on during-reward performance β_{REWARD} we compare during-incentive Round 2 ($R_t = 1$) to baseline Round 1 ($R_t = 0$) and exclude Round 3 data. To estimate the overall post-reward engagement level β_{POST} we compare post-reward Round 3 ($R_t = 1$) to baseline Round 1 ($R_t = 0$), and exclude Round 2. The net effect of incentives (e.g., during-reward and post-reward behavior) β_{NET} is estimated by comparing the combined during-reward Round 2 and post-reward Round 3 trials ($R_t = 1$) to baseline Round 1 ($R_t = 0$).

APPENDIX E: STUDIES USED IN META-ANALYSIS

Studies	Remarks	Sample Size	$\beta_{MOMENTARY}$	SE	z	p
Study 1	Study 1 in paper	C = 39; R = 38	-2.6	1.0	-2.6	.009**
Study 2	Study 2 in paper	C = 41; R = 46	-2.5	0.9	-2.5	.01*
Study 3	Study 3 in paper	C = 68; R = 56	-3.5	1.1	-3.1	.002**
Other Studies with Control and Replication Treatment conditions included in meta-analysis						
Study A	Replication of momentary reduction with advance notice about temporary nature of rewards at the end of Round 2	C = 31; R = 33	-2.2	0.9	-2.3	.02*
Study B	Replication of momentary reduction in engagement	C = 72; R = 74	-1.6	0.9	-1.7	.09
Study C	Replication of Study 2	C = 42; R = 41	-2.3	1.6	-1.4	.15
Study D	Replication of Study 3	C = 35; R = 36	-1.5	1.2	-1.3	.20
Study E	Study with and without a pre-reward round	C = 56; R = 96	-2.6	1.3	-2.0	.038*
Study F	Study with framing math vs both math and video as important (refer Appendix M)	C = 59; R = 52	-4.7	1.2	-3.8	<. 001***
Study G	Replication of Study with incentives for Math vs Leisure (refer Appendix L)	C = 87; R = 96	-3.2	0.7	-4.6	<. 001***

*p<.05; **p<.01; ***p<.001; C = control group, R = Reward group

Notes:

1. Study 4 in the main paper was not included in the meta-analysis because it used a different math task. The results will not change (in fact, will become stronger) if that study is also included. Study 5 was not included because the participants made a single choice for all the trials in Round 3 (as opposed to a series of choices, on per trial).
2. Studies B and E are part of a different project which also had the regular control and replication cells.

APPENDIX F: ROBUSTNESS CHECKS USING META-ANALYSIS DATA

Estimates for Momentary Reduction in Engagement ($\beta_{MOMENTARY}$) in meta-analysis with various specifications

	(1)	(2)	(3)	(4)
	Model with $MOMENTARY_t = \frac{1}{t}$	Model with $MOMENTARY_t = \frac{1}{t^2}$	Model with $MOMENTARY_t = \frac{1}{\sqrt{t}}$	Model with $MOMENTARY_t = \frac{1}{t}$ and excluding dropouts
Constant	-4.8 (0.3)***	-4.8 (0.3)***	-5.1 (0.3)***	-4.4 (0.3)***
$MOMENTARY_t$	1.3 (0.2)***	1.5 (0.3)***	1.5 (0.3)***	1.2 (0.2)***
Condition = Reward	1.3 (0.2)***	0.9 (0.2)***	2.2 (0.3)***	1.3 (0.2)***
Total Attempts in Round 1	1.0 (0.05)***	1.1 (0.05)***	1.0 (0.05)***	1.0 (0.04)***
$MOMENTARY_t$ * Condition	-3.0 (0.3)***	-2.8 (0.4)***	-3.7 (0.4)***	-2.9 (0.3)***
Dropouts Included	Yes	Yes	Yes	No
N	1098	1098	1098	1046
BIC	9881	9899	9881	9526

*p<.05; **p<.01; ***p<.001

Notes:

1. For the sake of parsimony and simplicity, we model $\boldsymbol{MOMENTARY_t}$ as per specification 1 in this paper which also has the lowest BIC.

2. Columns 1-3 compare different functional form specifications for the $\boldsymbol{MOMENTARY_t}$ parameter, and column 4 excludes the dropouts. Dropouts are participants who dropped-out of the study after completing Round 1.The parameter $\boldsymbol{t}$ indicates the post-reward trial or choice number.

APPENDIX G: NON-PARAMETRIC MODEL FOR ESTIMATIMNG MOMENTARY POST-REWARD ENGAGEMENT REDUCTION

Specification 1

We used the following flexible specification for predicting probability of choosing the math task in the post-reward round by individual i at trial t:

$$P(Y_{ti} = 1) = \phi\ (T * C_i + X_i + S)$$

$\phi =$ Logit link function

$T =$ Cubic splines $(df = 3)$ for post-reward trial$t$, $t \in$ [1..12] i.e., 12 post-reward choices

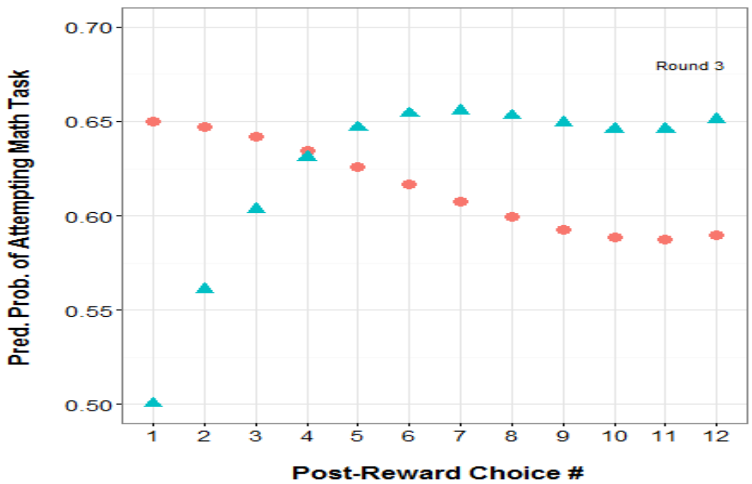
C_i = Experimental condition to which individual i is randomly assigned

$X_i =$ Total number of math attempts by individual i in the pre-incentive round. This is a measure of individual level differences in ability or interest in the target-task of the experiment

$S =$ Study Fixed Effect

The model specification did not use any assumption about how effort (e.g. attempting the math task) returns to baseline after incentives are stopped. Using flexible cubic-splines a piece-wise third order polynomial is used to fit individual post-reward behavior after the incentive ends.

Also, instead of using distributional assumptions to draw statistical inferences, we predicted the probability of attempting the math task for each of the post-reward trials in both the control and the treatment group. Using a 1000-sample bootstrap test, we examined if this difference did not contain zero to infer statistical difference in behavior between these two experimental groups.



Post-reward Choice #	Bootstrapped 95% CI	
Choice 1	-19.6%	-10.6%
Choice 2	-12.6%	-4.8%
Choice 3	-7.8%	0.0%
Choice 4	-4.3%	3.6%
Choice 5	-1.9%	6.1%
Choice 6	-0.2%	7.7%
Choice 7	0.9%	8.7%
Choice 8	1.4%	9.3%
Choice 9	1.7%	9.6%
Choice 10	1.8%	9.7%
Choice 11	2.0%	9.8%
Choice 12	1.7%	10.7%

Figure in the left-panel shows the predicted probability of choosing the math task in the reward group (blue triangular dot) and the control group (red circular dot) in Round 3 based on a typical run of the non-parametric model. The table in the right-panel shows the bootstrapped 95% CI from the non-parametric model for the *difference* in the predicted probability of choosing the math task between the reward and the control condition.

Specification 2

We use an alternative specification that too did not make parametric assumptions about how post-reward behavior varied during Round 3. The specification was as follows:

$$P(Y_{ti} = 1) = \phi\ (TD * C_i + Round2 * C_i + S)$$

ϕ = Logit link function

TD = Trial dummies for each of the post-reward trials $t \in [1..12]$

$Round2$ = Reward Round dummy

C_i = Experimental condition to which individual i is randomly assigned

S = Study Fixed Effect

Therefore, the interaction terms of post-reward trial dummies with the experimental condition measured the differences in the relative level of post-reward engagement in the reward group versus the control group, relative to the pre-reward baseline level (i.e., Round 1).

	β	SE	z	p	
Constant	0.45	0.06	7.87	<.001	***
Reward Round	-0.08	0.04	-1.74	0.082	
Trial 1	0.01	0.10	0.10	0.923	
Trial 2	-0.02	0.10	-0.25	0.804	
Trial 3	-0.14	0.10	-1.45	0.146	
Trial 4	-0.12	0.10	-1.28	0.200	
Trial 5	-0.10	0.10	-1.02	0.306	
Trial 6	-0.02	0.10	-0.16	0.871	
Trial 7	-0.24	0.09	-2.56	0.010	*
Trial 8	-0.29	0.09	-3.07	0.002	**
Trial 9	-0.24	0.09	-2.56	0.010	*
Trial 10	-0.35	0.09	-3.75	<.001	***
Trial 11	-0.23	0.09	-2.48	0.013	*
Trial 12	-0.28	0.09	-2.99	0.003	**
Condition = Reward	-0.10	0.04	-2.23	0.026	*
Round 2 x Reward	1.42	0.07	21.52	<.001	***
Trial 1 x Reward	-0.52	0.13	-3.96	<.001	***
Trial 2 x Reward	-0.24	0.13	-1.82	0.069	
Trial 3 x Reward	-0.03	0.13	-0.24	0.811	
Trial 4 x Reward	0.05	0.13	0.36	0.716	
Trial 5 x Reward	0.16	0.13	1.20	0.230	
Trial 6 x Reward	0.27	0.14	1.97	0.049	*
Trial 7 x Reward	0.37	0.13	2.75	0.006	**
Trial 8 x Reward	0.39	0.13	2.94	0.003	**
Trial 9 x Reward	0.34	0.13	2.58	0.010	**
Trial 10 x Reward	0.27	0.13	2.06	0.039	*
Trial 11 x Reward	0.37	0.13	2.75	0.006	**
Trial 12 x Reward	0.39	0.13	2.94	0.003	**

*p<.05; **p<.01; ***p<.001

Note: The regression includes study fixed effects whose estimates are not shown.

APPENDIX H: EFFECT OF TEMPORARY INCENTIVES ON ACCURACY AND NET OUTCOME

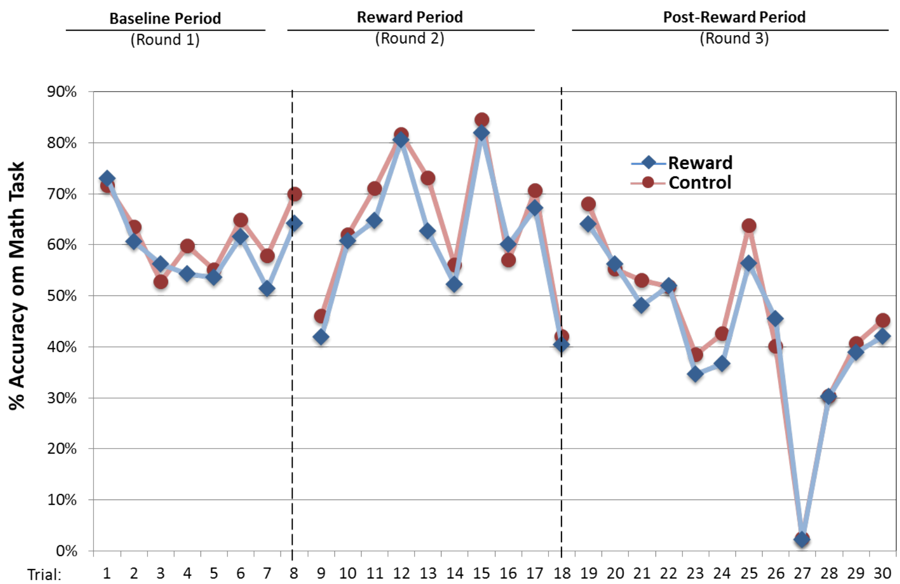


Figure H.1: Raw data showing the percentage of participants accurately solving the math task, conditional on choosing to attempt it. Incentives did not affect accuracy in our experiments.

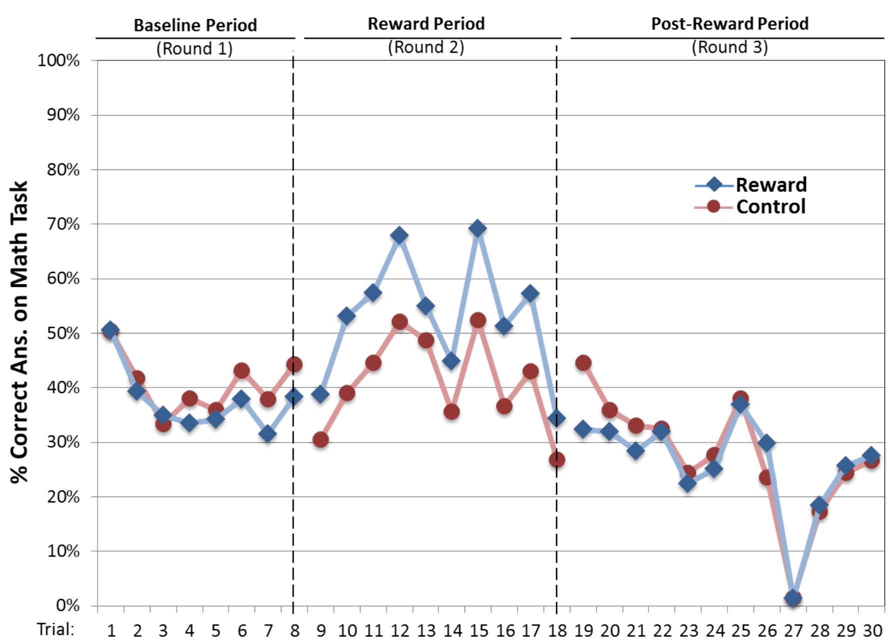


Figure H.2: Raw data showing proportion of correct answers for every trial in each round. Incentives had a net positive effect on the total number of correct answers, driven by the reward period, despite a significant post-reward decrease after incentives ended.

We primarily focused on effort (e.g., choosing to do the target task) as the key variable of interest, consistent with the approach used in the intrinsic motivation literature because effort is a behavioral outcome variable that represents a person's motivation level. Our flexible experimental paradigm also allowed us to also examine the effect of temporary incentives on accuracy (probability of correctly answering the math task after deciding to attempt it) and net outcome (total number of correct answers). The incentive could have resulted in people choosing the math task without being able to answer correctly, resulting in a decrease in accuracy compared to the control condition. However, we did not observe any such effects and temporary incentives did not affect accuracy at all. Therefore, as shown in figures H.1 and H.2, the same conclusions hold for effort and net outcome – a significant positive effect of incentives on the total number of math problems solved correctly ($\beta_{NET} = +0.46, z = +5.52, p < .001$).

APPENDIX I: MODERATION BY INITIAL MOTIVATION

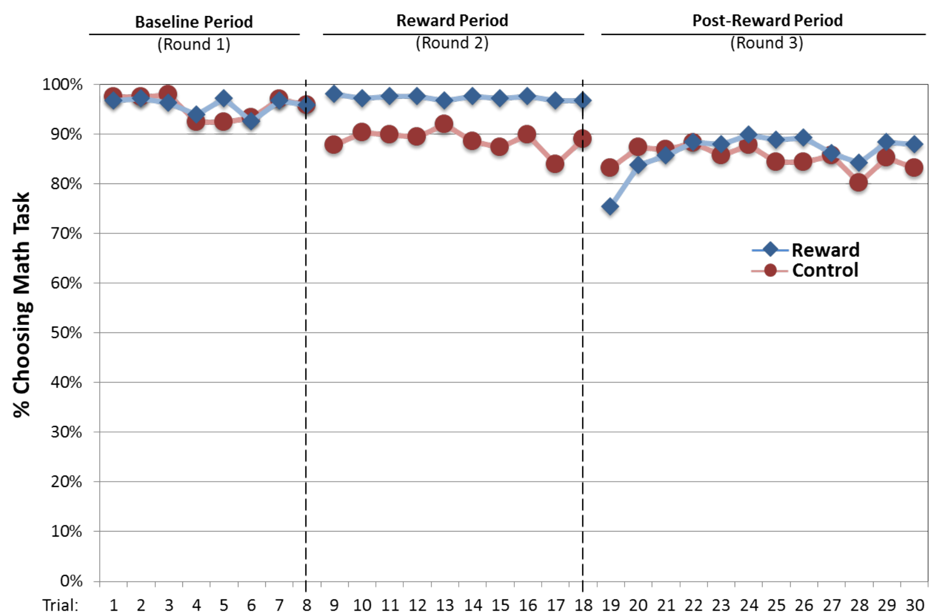


Figure I.1: Post-reward behavior of participants with high initial task interest using the internal meta-analysis data. The figure shows a significant post-reward decrease in engagement after incentives ended.

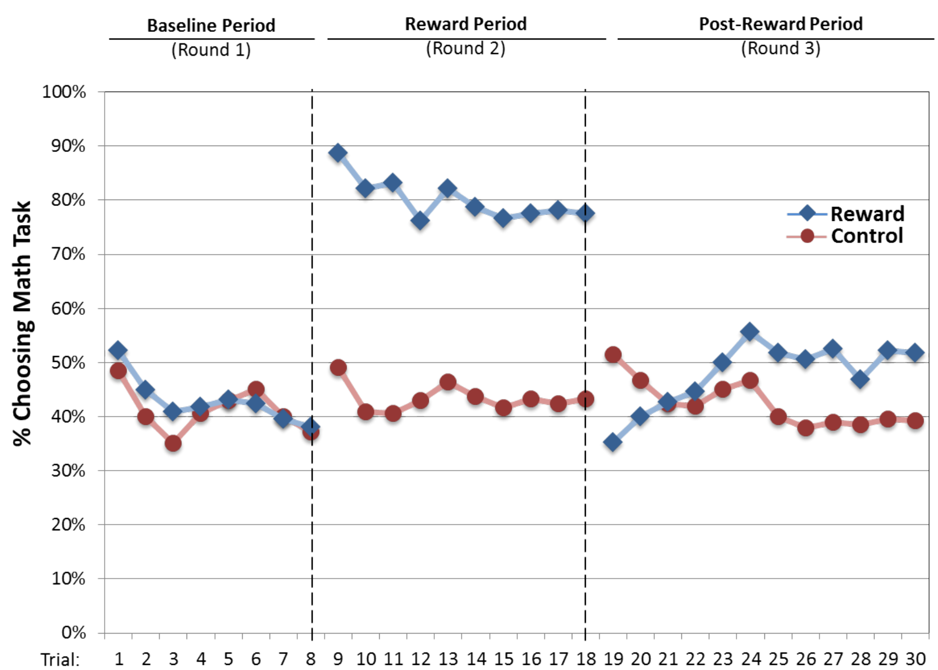


Figure I.2: Post-reward behavior of participants who low initial task interest using the internal meta-analysis data. The figure shows a significant post-reward decrease in engagement after incentives ended.

As shown in figures I.1 and I.2, the result of temporary incentives on post-reward behavior is very similar for the group of participants who exerted more versus less effort in the pre-reward round. This difference in initial effort represents difference in initial intrinsic motivation because in the pre-reward period participants did not know about any impending rewards. Both groups show an increase in effort when the rewards are available, followed by a decrease in the choice of the math task in the immediate post-reward period (low intrinsic motivation: $\beta_{MOMENTARY} = -3.22, z = -8.83, p < .001$; high intrinsic motivation ($\beta_{MOMENTARY} = -2.88, z = -2.94, p = .003$), and initial interest level does not moderate the momentary reduction in engagement behavior ($\beta_{MOMENTARY\ interaction} = .13, z < 1$).

APPENDIX J: HETEROGENEITY IN POST REWARD BEHAVIOR

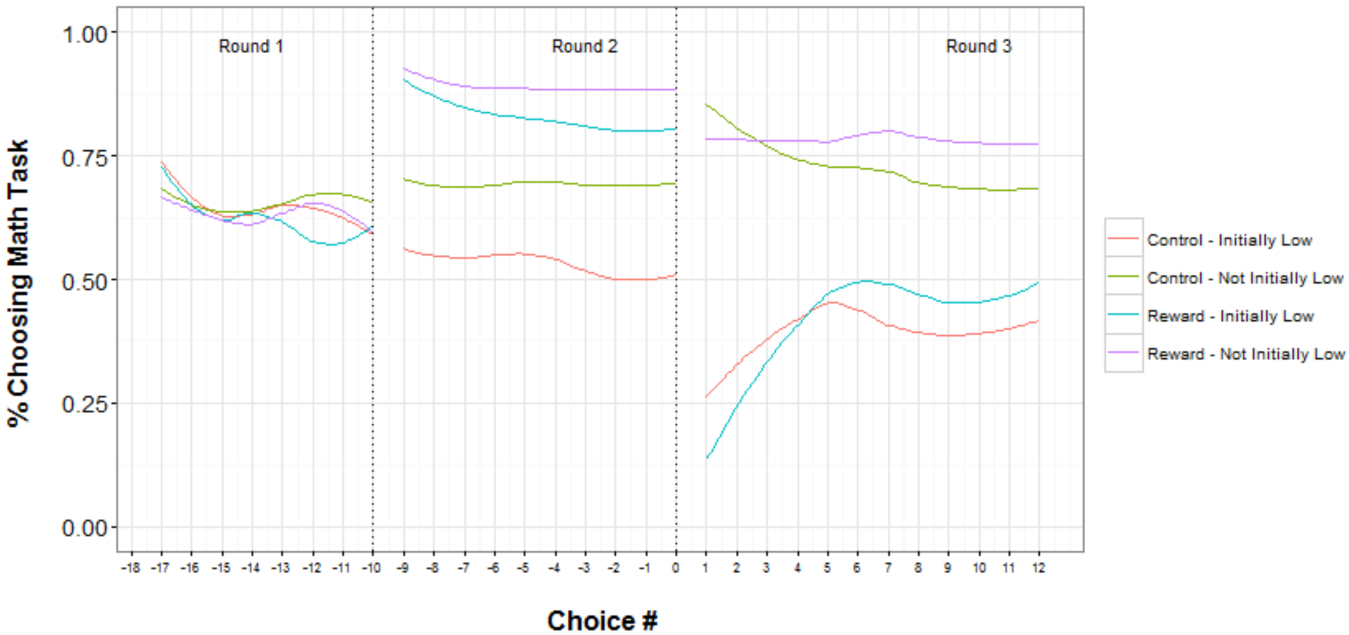


Figure J.1: Post-reward behavior (lowess lines) of participants with different initial behavior after the rewards ended. Both the reward and the control groups are further divided into two sub-groups each based on their initial post-reward behavior (initial post reward behavior lower or higher than their individual pre-reward baseline).

As shown in Figure J.1, the reward group that shows initial post-reward reduction in engagement eventually settles at a level higher than the corresponding control group. This suggests that the aggregate behavior was *not* driven by two types of reward group participants – one that showed a persistent post-reward reduction in engagement as predicted by prior theories (relative to corresponding control) and another that showed a persistent post-reward increase in engagement (relative to corresponding control). As shown in the table below, the final post-reward behavior does not differ between the control and the reward as a function on their initial post-reward behavior.

	DV = Final Post-reward Behavior (Normalized)	
Condition = Reward	0.121***	(-0.014)
Initial Reduction in Engagement (Normalized)	0.627***	(-0.03)
Condition = Reward x Initial Reduction in Engagement	0.016	(-0.039)
Constant	-0.048***	(-0.01)
Observations	1098	
R-Squared	0.503	

Note: *p<0.1; **p<0.05; ***p<0.01; Normalized = Average Initial Reduction - Pre-reward Baseline

Effect of High vs. Low Accuracy on Post-Reward Behavior

	β	SE	z	p	
Constant	-4.21	0.40	-10.48	<.001	***
$MOMENTARY_t$	1.54	0.55	2.82	0.005	**
Condition = Reward	0.90	0.46	1.94	0.053	
Proportion Correct in Round 2	1.70	0.52	3.27	0.001	**
Total Attempts in Round 1	0.81	0.05	17.40	<.001	***
$MOMENTARY_t$ * Condition	-2.50	0.76	-3.29	0.001	***
$MOMENTARY_t$ * Proportion Correct in Round 2	-0.63	0.86	-0.73	0.469	
Reward * Proportion Correct in Round 2	0.19	0.72	0.26	0.794	
$MOMENTARY_t$ * Reward * Proportion Correct in Round 2	-0.64	1.16	-0.55	0.582	

*p<.05; **p<.01; ***p<.001

Effect of Average Time Taken in Round 2 on Post-Reward Behavior

	β	SE	z	p	
Constant	-1.27	0.67	-1.90	0.057	
$MOMENTARY_t$	0.41	1.04	0.39	0.696	
Condition = Reward	1.40	0.82	1.70	0.090	
Average Time to do Math in Round 2	-0.10	0.03	-3.52	<.001	***
Total Attempts in Round 1	0.82	0.05	18.13	<.001	***
$MOMENTARY_t$ * Condition	-4.25	1.35	-3.15	0.002	**
$MOMENTARY_t$ * Average Time to do Math in Round 2	0.04	0.05	0.76	0.450	
Reward * Average Time to do Math in Round 2	-0.02	0.04	-0.44	0.660	
$MOMENTARY_t$ * Reward * Average Time to do Math in Round 2	0.07	0.06	1.07	0.286	

*p<.05; **p<.01; ***p<.001

Effect of Average Time Taken in Round 3 on Post-Reward Behavior

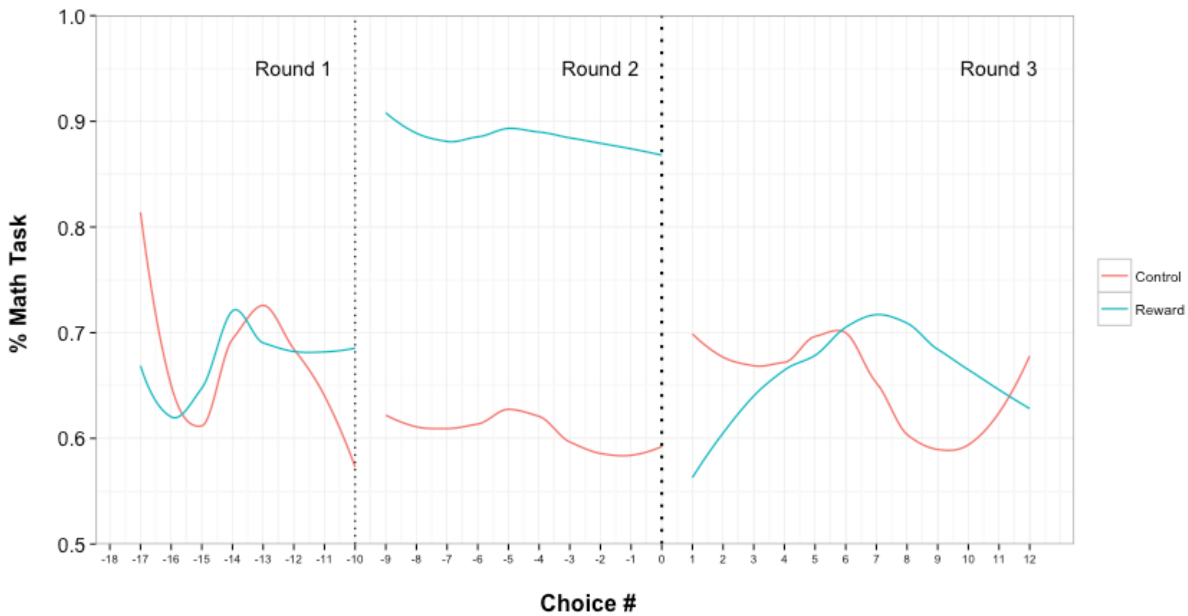
	β	SE	z	p	
Constant	-1.96	0.23	-8.39	<.001	***
$MOMENTARY_t$	1.69	0.38	4.44	<.001	***
Condition = Reward	1.42	0.43	3.29	<.001	**
Average Time to do Math in Round 3	0.00	0.00	-0.43	0.67	
Total Attempts in Round 1	0.59	0.03	17.25	<.001	***
$MOMENTARY_t$ * Condition	-4.33	0.88	-4.91	<.001	***
$MOMENTARY_t$ * Average Time to do Math in Round 3	-0.01	0.01	-0.90	0.37	
Reward * Average Time to do Math in Round 3	-0.01	0.02	-0.33	0.74	
$MOMENTARY_t$ * Reward * Average Time to do Math in Round 3	0.07	0.04	1.67	0.09	

*p<.05; **p<.01; ***p<.001

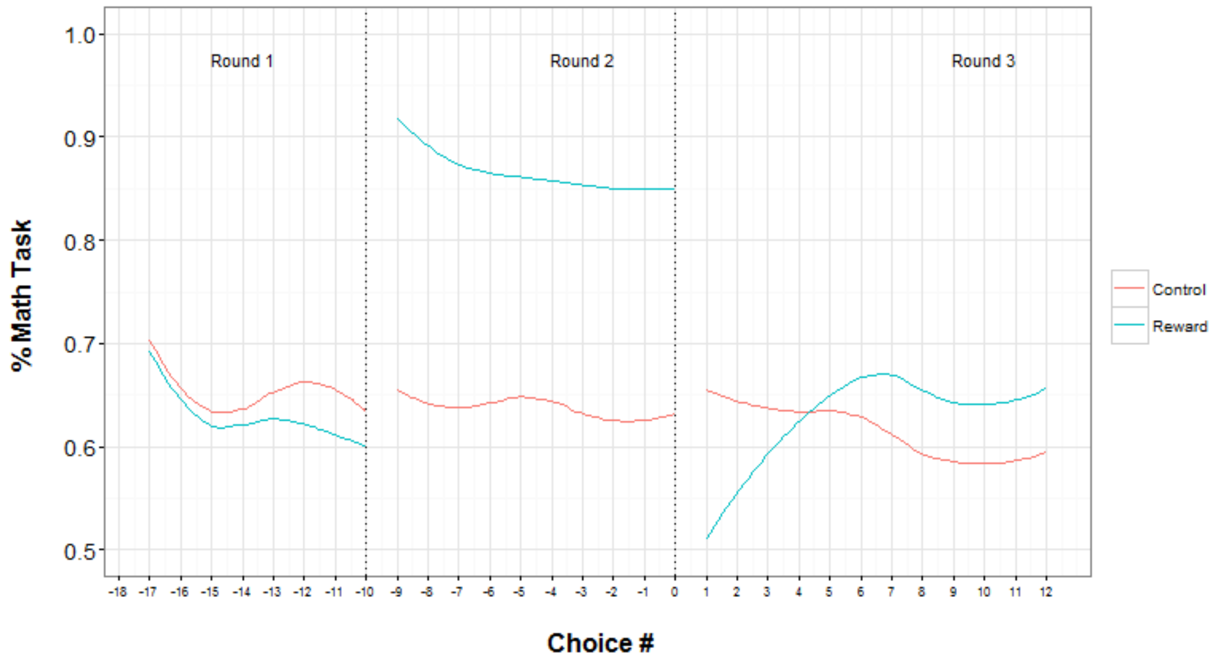
APPENDIX K: ADDITIONAL CHARTS FOR ALL STUDIES

Raw Choice Data of All Studies with Lowess Smoother

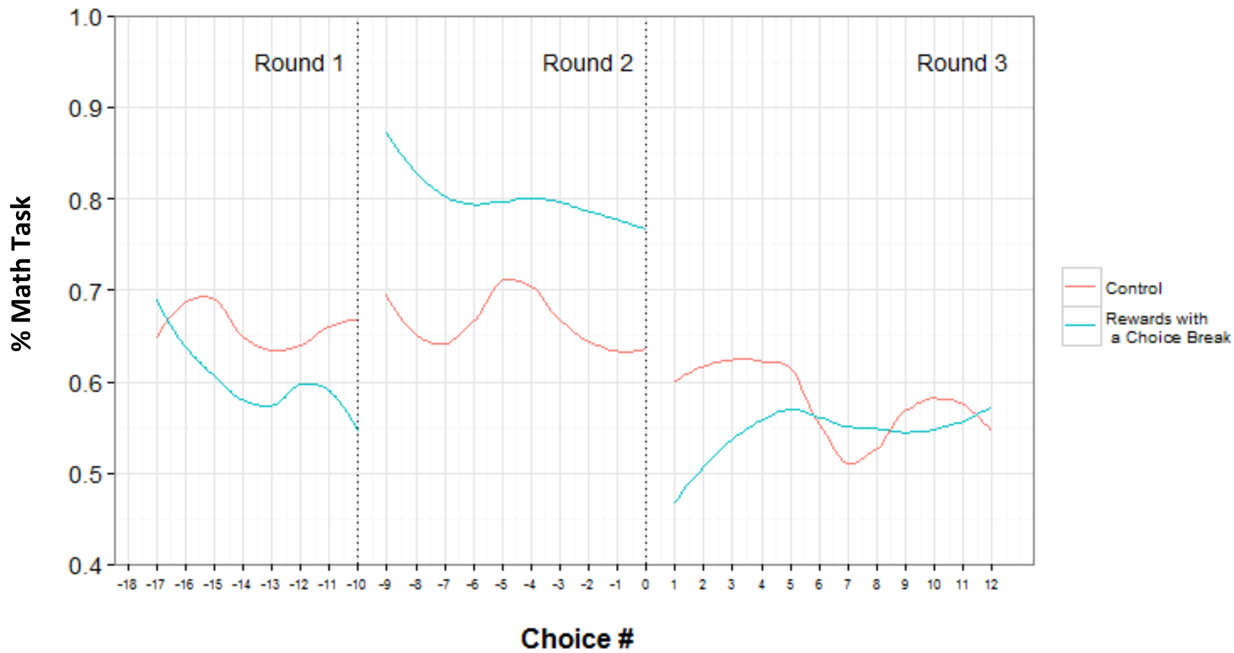
Study 1: The Dynamics of Post-Reward Task Engagement

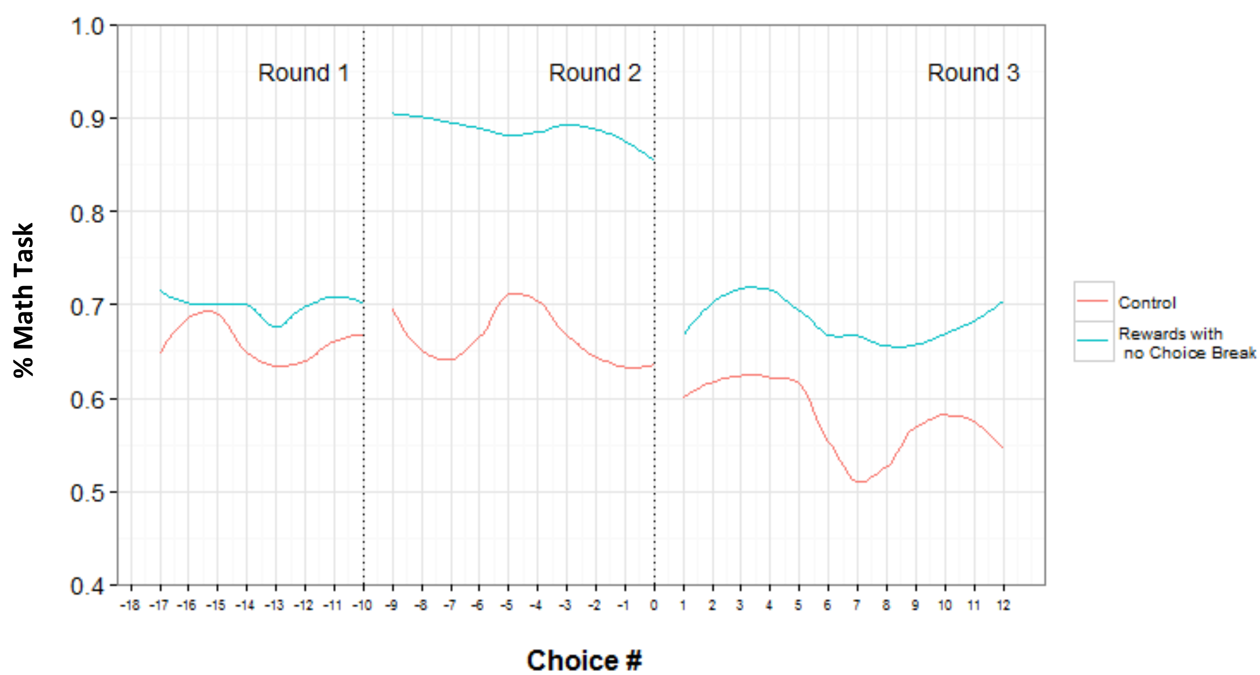


Meta-Analysis: The Dynamics of Post-Reward Task Engagement

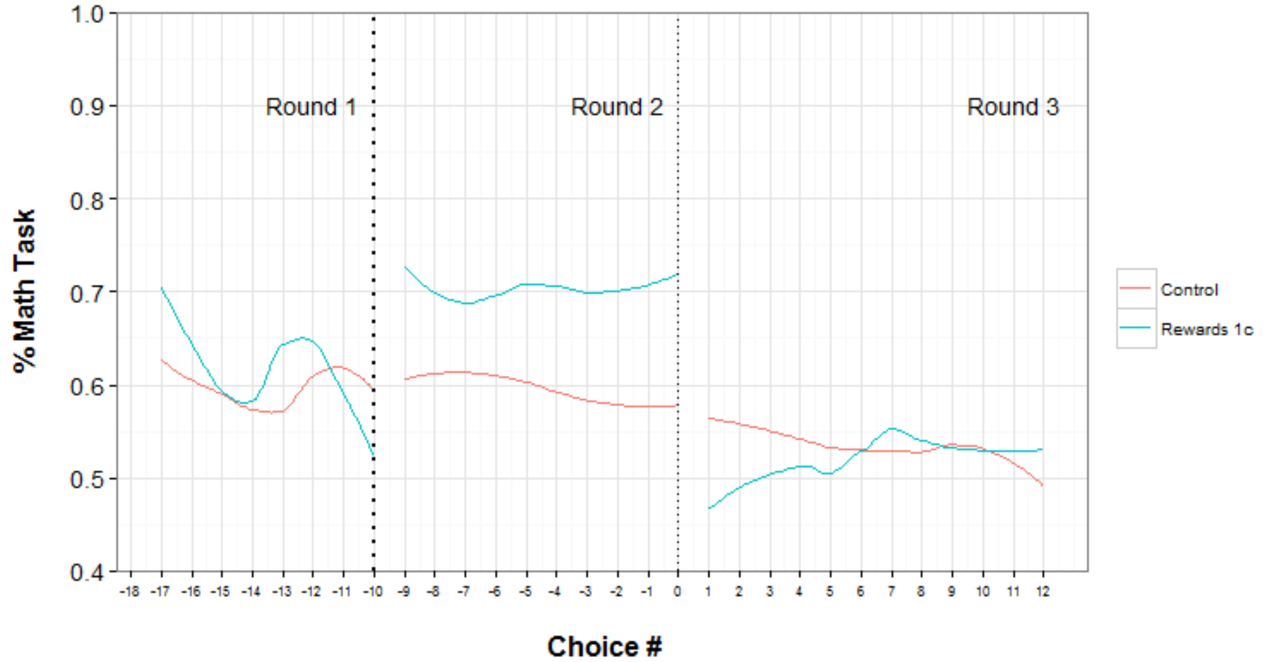


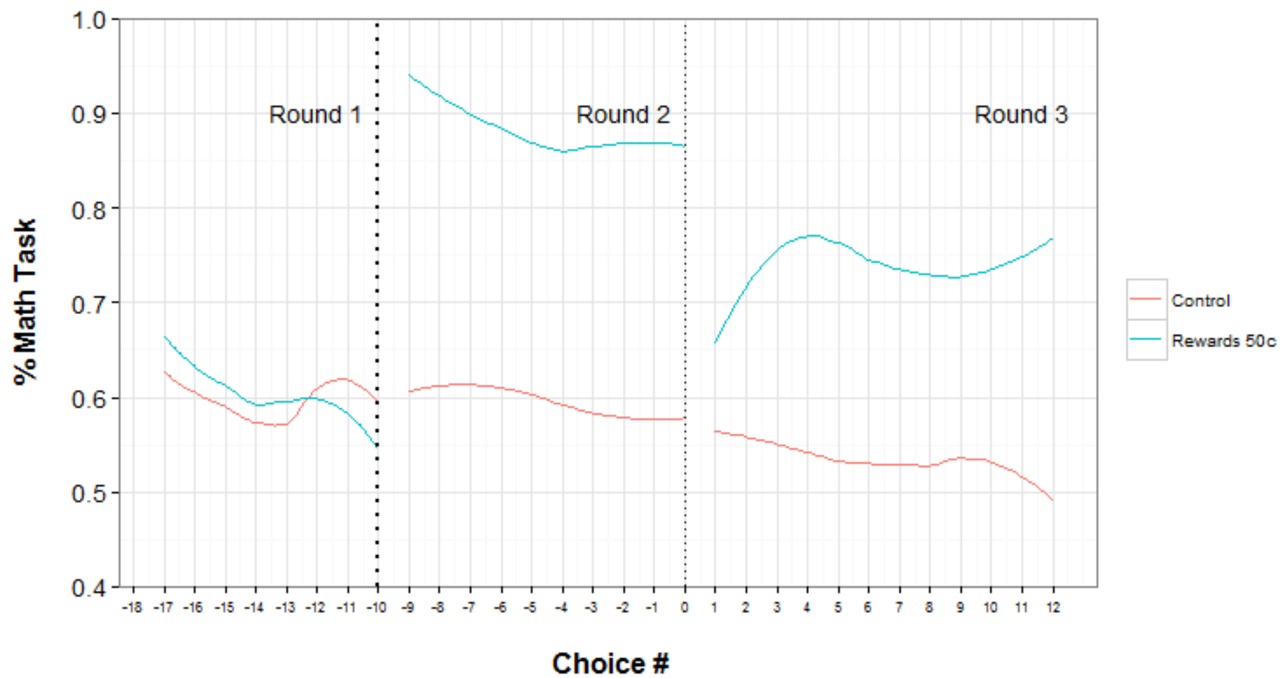
Study 2: Providing a Break Eliminates Engagement Reduction



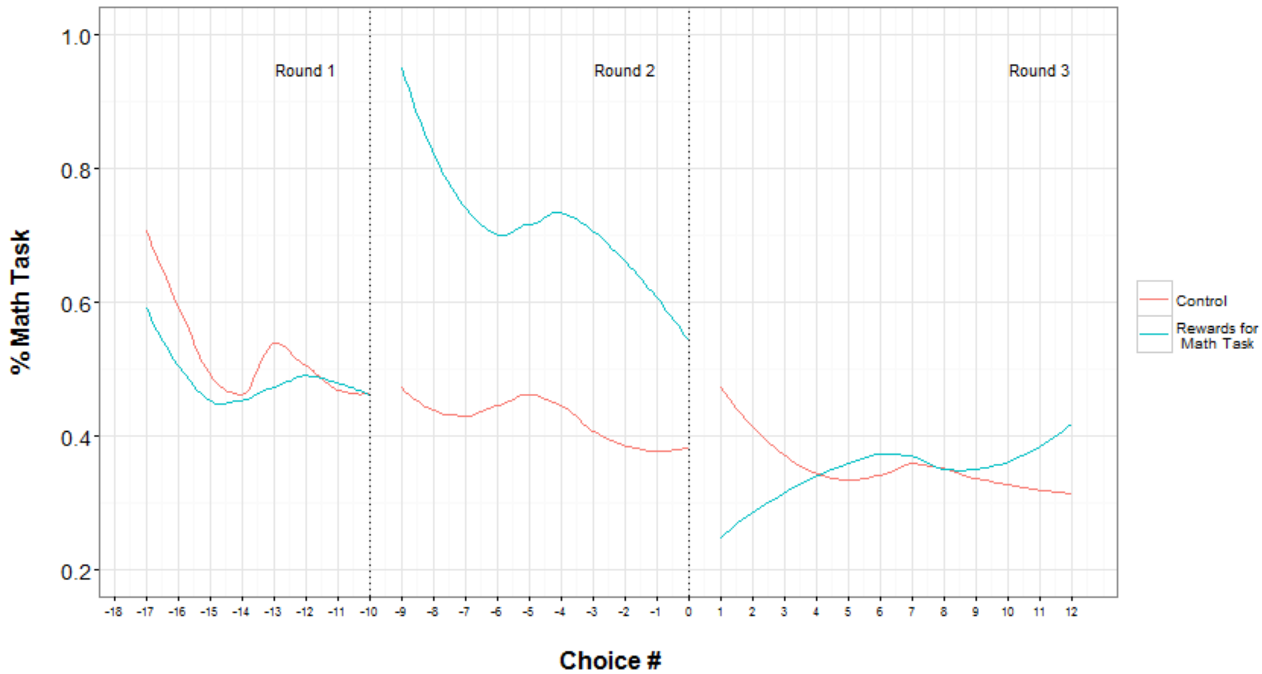


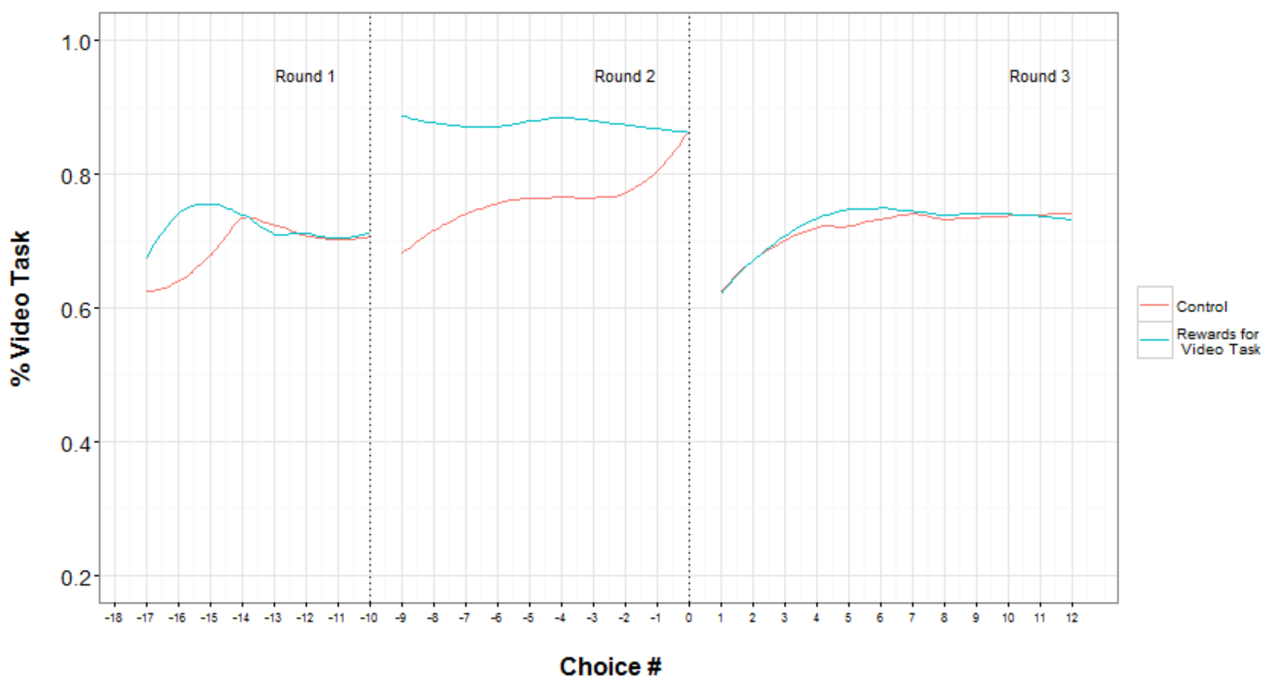
Study 3: Large Rewards Do Not Reduce Engagement





Study 4: Paying for a Leisure Task Does Not Reduce Engagement

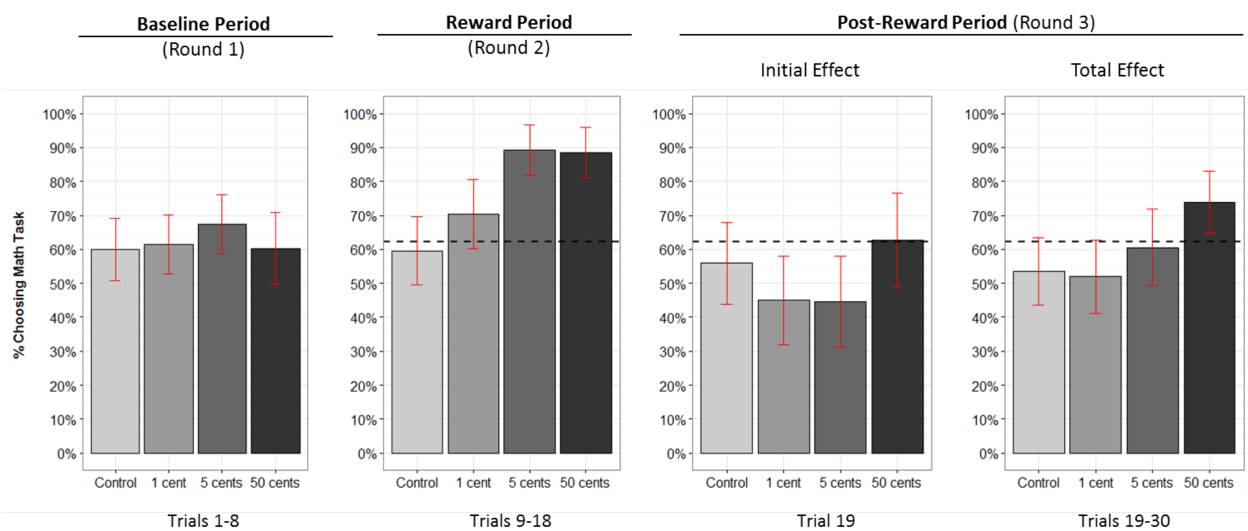




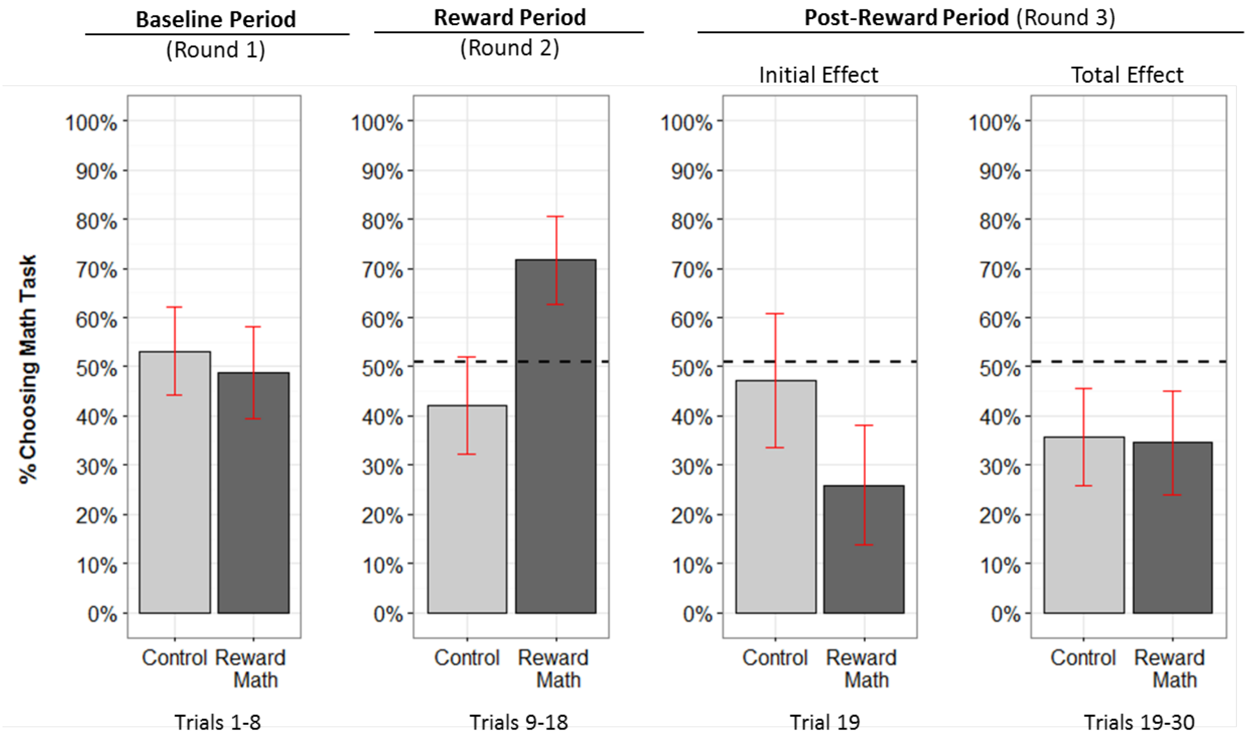
Average Effort by Rounds for All Rounds with 95% CI

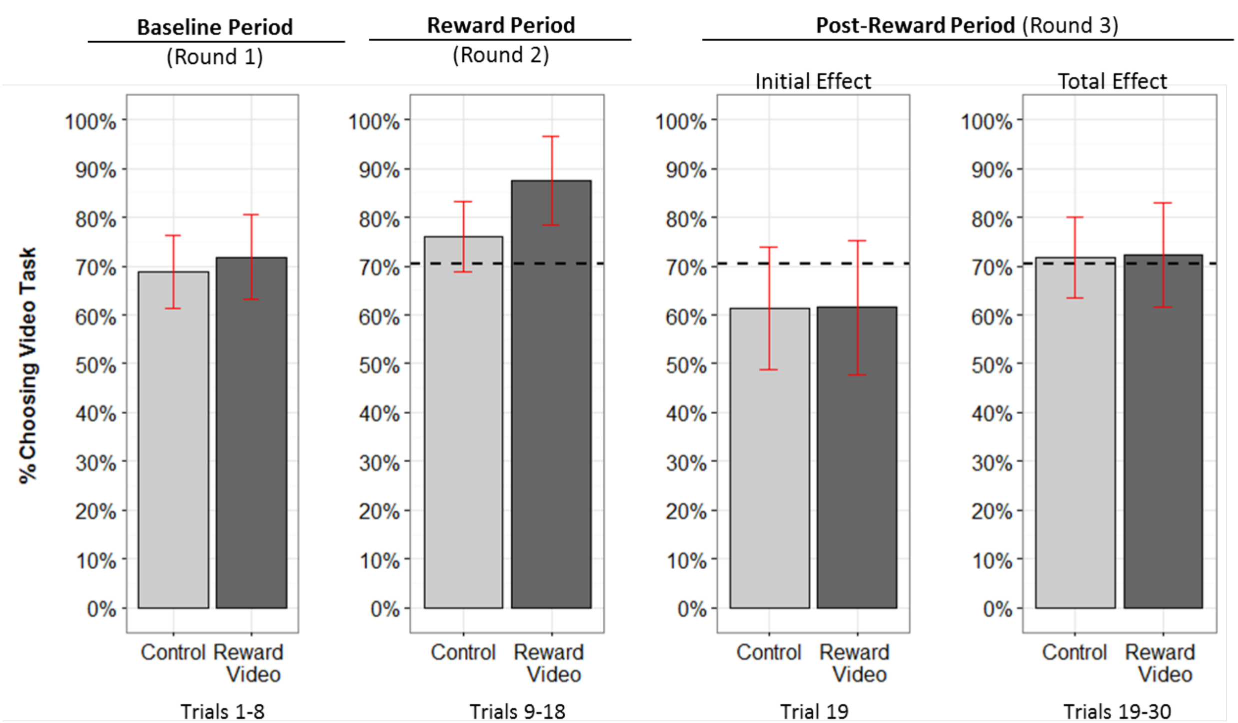
Study 2: Providing a Break Eliminates Engagement Reduction

Study 3: Large Rewards Do Not Reduce Engagement



Study 4: Paying for a Leisure Task Does Not Reduce Engagement





APPENDIX L: STUDY WITH PAYING FOR A LEISURE TASK

In the main paper, we reported a study where we varied whether the target task was a cognitive math task or a leisure task like watching and rating videos. In that study we used a different math task that entailed counting that number of 1s in a grid of 150 1s and zeros. Here we report a replication of the same experiment with the original math task that was used in Studies 1-3 and Study 5 of the main paper.

Method

Adult participants were recruited from Amazon MTurk. A target of 480 participants were requested, yielding 477 surveys. Unusable cases (duplicate IP addresses, technical problems, failed attention check) were removed prior to analysis, yielding 340 valid completes.[2] Participants who completed Round 1 but then dropped-out (4.9%) were coded as not doing the focal task and included in the analysis.

Participants were randomly assigned to one of four conditions in a 2 (Target task: Math vs. Video) x 2 (Control vs. Incentive) between-subjects design. In the two incentive conditions, they were either paid 5 cents for correctly completing math tasks, as in the prior studies, or paid 5 cents for each video they watched and rated (1- 5 stars). The two control conditions matched the two incentive conditions, highlighting the target task without any incentive.[3]

Results

We have two different control conditions in this study, so we compare each incentive condition to the corresponding control condition. We replicated the momentary reduction in engagement when people were paid for doing the math task. Fewer people chose the math task in the first trial of Round 3 in the incentive condition after the rewards had ended, compared to in the matching control condition (47% vs. 68%; $\beta = -0.18$, $t = -2.77$, $p = .006$). There was no long-term reduction of effort due to incentives, relative to control (62% in both). These results were further confirmed in the hierarchical regressions ($\beta_{MOMENTARY} = -3.20, z = -4.55, p < .001$; $\beta_{POST} = +0.06, z = +0.14, p > .250$).

When participants were paid for the leisure task instead, we did not find a reduction in engagement. The proportion of video choices in the first trial after the incentive ended was similar to the corresponding control condition (50% vs. 47% , $\beta = 0.02, t < 1$) and we found no long-term reduction of effort (57% vs. 56%; $\beta = 0.0004$, $t < 1$). These results were further confirmed by the hierarchical regression models ($\beta_{MOMENTARY} = +0.11, z = +0.11, p > .250$; $\beta_{POST} = +0.10, z = +0.24, p > .250$).

[2] The high number of unusable cases was due to duplicate IPs that happened due because the software did not filter on previous respondents since a few months had passed between this study and the previous studies.
[3] Two different framings of the math and video tasks were used in each condition, but since there were no differences, the results were merged. These framing details are provided at the end of this study.

In fact, the momentary reduction in engagement observed when incentivizing the math task was completely eliminated when the videos were incentivized instead ($\beta_{MOMENTARY\ interaction} = +3.33, z = +2.92, p = .003$). There was no difference between the two conditions in terms of the longer-term post-reward baseline level($\beta_{POST\ interaction} = +0.03, z = +0.05, p > .250$).

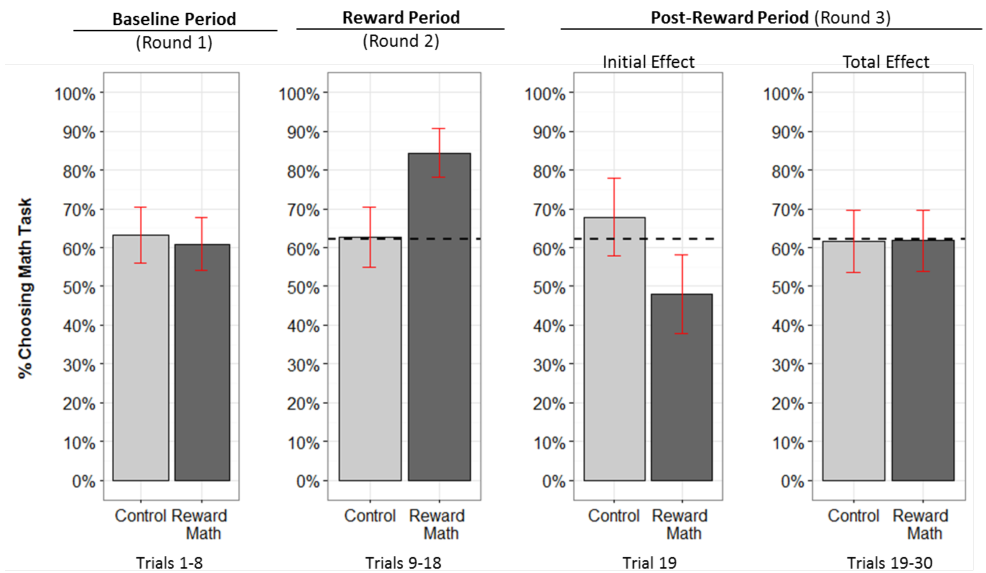


Figure L.1: Results of all rounds when the math task was incentivized. Dotted lines represent the baseline (average effort level Round 1), and the vertical lines are 95% CIs

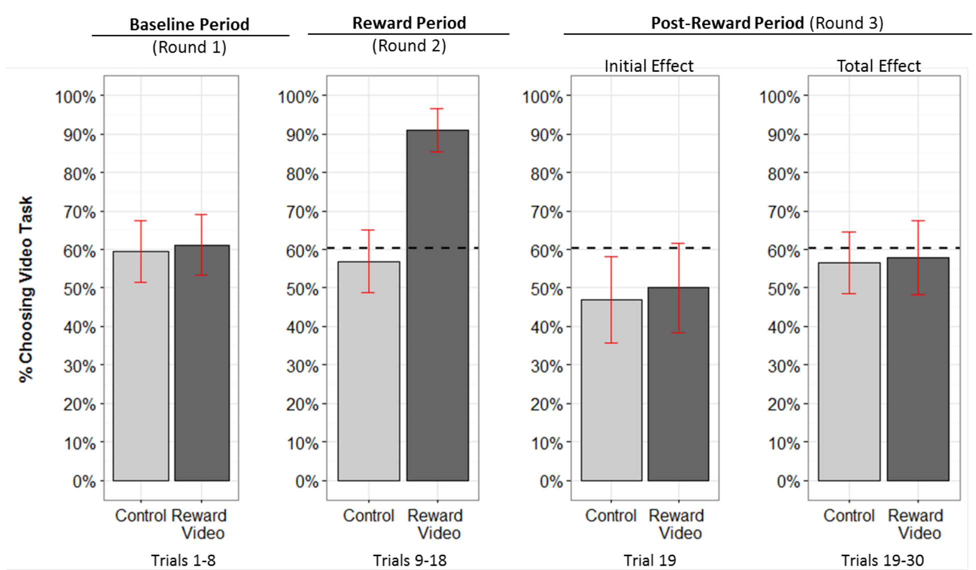


Figure L.2 Results of all rounds when the video task was incentivized. Dotted lines represent the baseline (average effort level Round 1), and the vertical lines are 95% CIs

Framing details used in this Study

Target Task = Math

PLEASE READ THE INSTRUCTIONS CAREFULLY.

In this survey you will be asked to do a task. The task is to **solving cognitive math tasks**. We will use your results to calibrate and standardize a training task for a reasoning study.

Since doing the task can be tiring, you will also have an option of a different task, **evaluating videos of television advertisements,** so that you can take a break.

It is completely up to you to choose which task you want to do in each round.

PLEASE READ THE INSTRUCTIONS CAREFULLY.

In this survey you will be asked to do a task. The task is to **solving cognitive math tasks**. We will use your results to calibrate and standardize a training task for a reasoning study.

Since doing the task can be tiring, you will also have an option of a different task, **evaluating videos of television advertisements,** so that you can take a break.

Although the task can feel tedious, you may **learn something useful** from solving the math problems, and **train your mental skills.**

It is completely up to you to choose which task you want to do in each round.

Target Task = Video

PLEASE READ THE INSTRUCTIONS CAREFULLY.

In this survey you will be asked to do a task. The task is to **evaluate videos of television advertisements**. We will use your responses to design experimental stimuli for an attention and perception task.

Since doing the task can be tiring, you will also have an option of a different task, **solving cognitive math problems,** so that you can take a break.

It is completely up to you to choose which task you want to do in each round.

PLEASE READ THE INSTRUCTIONS CAREFULLY.

In this survey you will be asked to do a task. The task is to **evaluate videos of television advertisements**. We will use your responses to design experimental stimuli for an attention and perception task.

Since doing the task can be tiring, you will also have an option of a different task, **solving cognitive math problems,** so that you can take a break.

Although the task can feel tedious, you may **learn something useful** from evaluating the videos, and **train your mental skills.**

It is completely up to you to choose which task you want to do in each round.

APPENDIX M: STUDY WITH FRAMING BOTH CHOICE OPTIONS AS IMPORTANT

In our studies reported in the paper, the math task was framed as important and potentially beneficial, in part to highlight the self-control tradeoff between goals with more immediate and delayed benefits. This raises the possibility, however, that the framing made participants feel obligated to work on the math task, rather than watch the videos, even after the reward ended. Could this have resulted in a short-lived reduction in engagement? In this study we investigate this potential concern.

Method

Adult participants were recruited from Amazon MTurk to complete an online survey. A target of 300 participants were requested, yielding 291 surveys. Records with duplicate IP addresses, or who reported having technical problems with viewing the videos or working on the math task, or who failed the basic attention check were removed prior to analysis, yielding 219 valid completes. The proportion of participants in this sample who reached until the end of Round 1, but then dropped-out part way through was 2.3%. Participants were randomly assigned to one of four conditions, in a 2 (Control, Reward) x 2 (Math Important, Both Important) between-subjects design. The two replication conditions (Math-Important control and reward) were similar to Study 1. The other two conditions (Both-Important control and reward) were the same, except that participants were told that their data, both from doing math and from the video task was important in the study. Participants were also told that, since the survey was being administered to many people, it was completely up to them to choose what they wanted to do. This framing was designed to remove any signal to the participants that were expected to do the math tasks, and to encourage participants to choose what they truly wanted in do in each round. As a result, if participants' sense of obligation to do the math task had arrested the post-reward reduction in engagement in our studies, we would observe a stronger decrease in engagement in the Both-Important condition.

Results

A manipulation check, collected at the end of the study, confirmed that participants in the Both-Important control condition expressed more agreement that the videos and math task were equally important (on a 9 point scale) than in the Math-Important control condition ($M_{control,both} = 4.43,\ SD = 1.61\ vs. M_{control,math} = 3.79, SD = 1.75;\ t(110) = 1.98, p =$ .05).

Since this study included two differently-framed control conditions, we compared each reward condition to the corresponding control condition. We replicated the momentary reduction in engagement behavior when using the same instructions, in the Math-Important incentive and control conditions. Fewer people chose the math task in the first trial of Round 3 in the reward condition after the incentives had ended, compared with the same trial in the control condition

(38% vs. 64%; β= -0.18, *t*= -2.13, *p*=.04; Figure M.1). There was no longer-term reduction in engagement due to incentives, relative to control (57% vs. 63%; β= - 0.03, *t* <1). These results were further confirmed in the hierarchical regression models ($\beta_{MOMENTARY} = -4.69, z = -4.00,\ p < .001;\ \beta_{POST} = -0.04, z = -0.07, p > .250$).

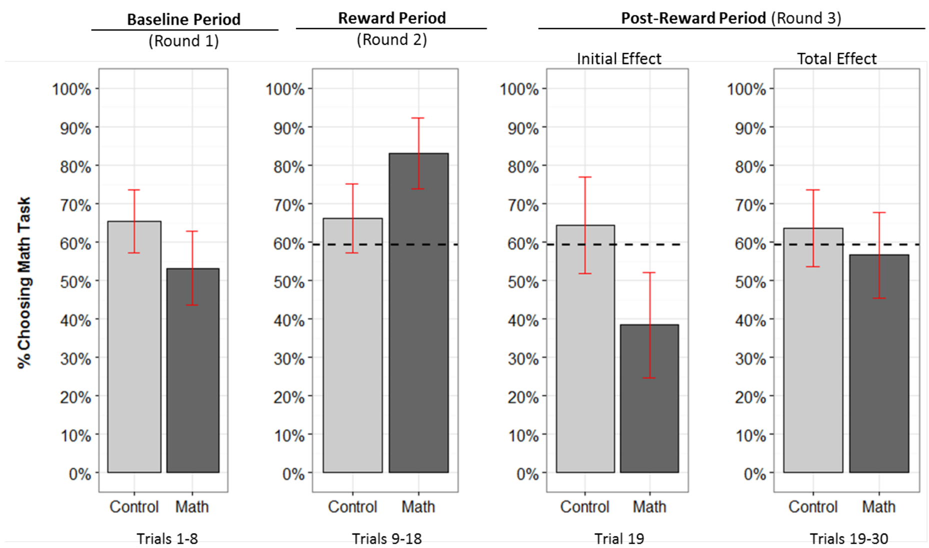


Figure M.1: Results of all rounds in the replication condition when math was the focal task. Dotted lines represent the baseline (average effort level Round 1), and the vertical lines are 95% CIs

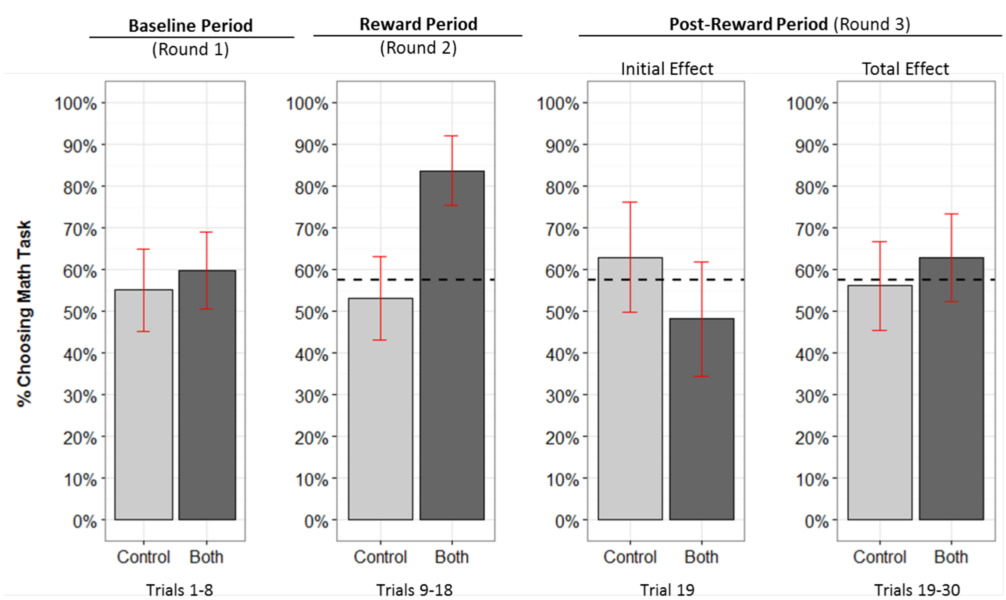


Figure M.2: Results of all rounds when both tasks were framed as equally important to the experimenter. Dotted lines represent the baseline (average effort level Round 1), and the vertical lines are 95% CIs

Likewise, when we instead told participants that both the task options (math and video) are equally important, we again replicate the findings in the Both-Importance incentive and control conditions. Fewer people chose the math task in the first trial of Round 3 in the incentive condition after the incentives had ended, compared to in the control condition (48% vs. 63%; β = -0.18, *t*=-2.42, *p*=.02). There was no longer-term reduction in engagement due to incentives, relative to control (63% vs. 57%; β=0.03, *t* <1). These results were further confirmed in the hierarchical regression ($\beta_{MOMENTARY} = -2.52, z = -3.33, p < .001$; $\beta_{POST} = +0.25, z = +0.51, p > .250$).

A hierarchical regression model confirmed that there was no difference in the extent of momentary reduction in engagement between the two task-framing conditions ($\beta_{MOMENTARY:\, MATH\, VS.\, BOTH} = +1.47, z = +1.08, p > .250$). Likewise, there was no difference in the extent of overall post-reward reduction between the two conditions ($\beta_{POST:\, MATH\, VS.\, BOTH} = +0.27, z = +0.38, p > .250$).

The results of this study suggest that the momentary nature of post-reward reduction in engagement cannot be explained by the experimental instructions inducing a feeling of obligation to do math tasks among the participants. Furthermore, the findings are also inconsistent with a self-signaling account, in which the participants continued with the more challenging math tasks (after a short break) to feel good about themselves.

APPENDIX N: ALTERNATIVE PARAMETERIZATION TO ESTIMATE INITIAL REDUCTION IN ENGAGEMENT IN THE POST-REWARD PERIOD

The following estimates are of regression models that use the same specification as Equation 5 in Appendix D, with $MOMENTARY_t$ replaced by $(MOMENTARY_t - 1)$. $MOMENTARY_t$ is modeled in all equations as $\frac{1}{t}$ where t is the post-reward trial number. In this specification, the estimate of $Condition = Reward$ indicates the initial difference between the reward versus the control group at the start of the post-reward period relative to the long-run post-reward baseline level.

Study 1: Reward vs Control

	β	SE	z	p	
Constant	-3.77	1.01	-3.73	<.001	***
$(MOMENTARY_t - 1)$	1.06	0.75	1.40	.161	
Condition = Reward	-1.97	0.91	-2.18	.029	*
Total Attempts in Round 1	1.09	0.17	6.58	<.001	***
$(MOMENTARY_t - 1)$ * Reward	-2.59	1.00	-2.59	.010	**

*p<.05; **p<.01; ***p<.001

Internal Meta-analysis: Reward vs Control

	β	SE	z	p	
Constant	-3.54	0.33	-10.62	<.001	***
$MOMENTARY_t - 1$	1.30	0.25	5.23	<.001	***
Condition = Reward	-1.76	0.33	-5.34	<.001	***
Total Attempts in Round 1	1.05	0.05	21.87	<.001	***
$MOMENTARY_t - 1$ * Reward	-3.03	0.33	-9.14	<.001	***

*p<.05; **p<.01; ***p<.001

Study 2: Replication (5c) and High-effort Break vs. Low-effort Break

	β	SE	z	p	
Constant	-3.52	0.66	-5.34	<.001	***
$MOMENTARY_t - 1$	0.00	0.52	0.00	.990	
Condition = Reward	-1.20	0.59	-2.02	.043	*
Total Attempts in Round 1	0.95	0.09	10.05	<.001	***
$MOMENTARY_t - 1$ * Reward	-1.24	0.63	-1.97	.049	*

*p<.05; **p<.01; ***p<.001

Study 2: Replication (5c) vs. Control

	β	SE	z	p	
Constant	-3.18	0.94	-3.38	.001	***
$MOMENTARY_t - 1$	0.99	0.77	1.29	.198	
Condition = Reward	-0.59	0.92	-0.64	.523	
Total Attempts in Round 1	0.85	0.13	6.70	<.001	***
$MOMENTARY_t - 1$ * Reward	-2.49	0.98	-2.55	.011	*

*p<.05; **p<.01; ***p<.001

Study 2: High-effort Break vs. Control

	β	SE	z	p	
Constant	-3.34	1.01	-3.30	.001	***
$MOMENTARY_t - 1$	1.23	0.87	1.42	.155	
Condition = Reward	-1.38	0.99	-1.39	.163	
Total Attempts in Round 1	0.93	0.12	7.63	<.001	***
$MOMENTARY_t - 1$ * Reward	-2.15	1.03	-2.10	.036	*

*p<.05; **p<.01; ***p<.001

Study 2: Low-effort Break vs. Control

	β	SE	z	p	
Constant	-4.77	0.98	-4.87	<.001	***
$MOMENTARY_t - 1$	1.06	0.93	1.14	.254	
Condition = Reward	0.15	0.87	0.18	.859	
Total Attempts in Round 1	1.17	0.15	7.77	<.001	***
$MOMENTARY_t - 1$ * Reward	-1.14	1.09	-1.04	.297	

*p<.05; **p<.01; ***p<.001

Study 2: High-effort Break vs. Low-effort Break

	β	SE	z	p	
Constant	-3.78	0.74	-5.14	<.001	***
$MOMENTARY_t - 1$	0.00	0.56	-0.01	.996	
Condition = Reward	-1.45	0.68	-2.13	.033	*
Total Attempts in Round 1	1.01	0.11	8.87	<.001	***
$MOMENTARY_t - 1$ * Reward	-1.04	0.75	-1.39	.165	

*p<.05; **p<.01; ***p<.001

Study 3: Replication (5c) vs. Control

	β	SE	z	p	
Constant	-5.61	1.13	-4.98	<.001	***
$MOMENTARY_t - 1$	0.72	0.74	0.97	.332	
Condition = Reward	-2.33	1.14	-2.04	.042	*
Total Attempts in Round 1	1.35	0.19	6.97	<.001	***
$MOMENTARY_t - 1$ * Reward	-3.52	1.15	-3.07	.002	**

*p<.05; **p<.01; ***p<.001

Study 3: Low Reward (1c) vs. Control

	β	SE	z	p	
Constant	-5.10	1.09	-4.67	<.001	***
$MOMENTARY_t - 1$	0.76	0.64	1.19	.232	
Condition = Reward	-1.71	1.11	-1.54	.123	
Total Attempts in Round 1	1.25	0.18	6.84	<.001	***
$MOMENTARY_t - 1$ * Reward	-1.94	0.95	-2.04	.042	*

*p<.05; **p<.01; ***p<.001

Study 3: High Reward (50c) vs. Control

	β	SE	z	p	
Constant	-4.01	0.96	-4.17	<.001	***
$MOMENTARY_t - 1$	0.67	0.76	0.87	.382	
Condition = Reward	1.31	1.07	1.22	.222	
Total Attempts in Round 1	1.00	0.15	6.76	<.001	***
$MOMENTARY_t - 1$ * Reward	-1.48	1.19	-1.24	.215	

*p<.05; **p<.01; ***p<.001

Study 4: Replication (5c) vs. Control

	β	SE	z	p	
Constant	-4.51	0.92	-4.89	<.001	***
$MOMENTARY_t - 1$	1.52	0.69	2.20	0.028	*
Condition = Reward	-3.21	1.11	-2.89	0.004	**
Total Attempts in Round 1	0.97	0.13	7.38	<.001	***
$MOMENTARY_t - 1$ * Reward	-4.15	1.10	-3.77	<.001	***

*p<.05; **p<.01; ***p<.001

Study 4: Incentives for Video vs. Control

	β	SE	z	p	
Constant	-4.96	1.02	-4.88	<.001	***
$MOMENTARY_t - 1$	-1.42	0.60	-2.36	0.018	*
Condition = Reward	-0.47	0.87	-0.54	0.589	
Total Attempts in Round 1	1.14	0.17	6.59	<.001	***
$MOMENTARY_t - 1$ * Reward	-0.52	0.90	-0.57	0.568	

*p<.05; **p<.01; ***p<.001

APPENDIX O: RESULTS OF FIELD STUDIES THAT HAVE MEASURED BEHAVIOR AFTER CONTINGENT INCENTIVES ENDED

	Domain	Target Group	Incentive size / type	Time-point(s) when post-reward behavior was measured	Finding(s)
Garbarino & Slonim, 2005	Education	University students at a private school	$10 for passing a test	Immediate post-reward behavior (e.g, the subsequent test)	On Average, 2 fewer questions attempted in incentive group vs control (p<.05)
Volpp et al., 2006	Smoking Cessation	Smokers at Philadelphia Veteran Affairs Medical Center	$100 for quitting to smoke	6 months after incentive to quit	Quit rates in incentive group (6.5%) not different from control (4.6%, p>.2)
Jackson, 2010	Evaluation of Advanced Placement Incentive Program (APIP) in Education	11th and 12th grade students (and teachers) in Texas schools serving underprivileged populations	Between $100 and $500 for getting a score of 3 or more in each eligible test subject	Future test scores and college graduation	13% increase in number of students scoring about 1100/24 on SAT/ACT (p<.05) and 5% increase in students matriculating in college (p<.10)
Volpp et al., 2008	Warfarin Adherence	Warfarin patients at the UPenn Anticoagulation Management Center	Lottery with daily expected value of $5(Study 1) or $3(Study 2)	Not-reported	Regulation of anticoagulation levels changed from 35% (pre) to 42% (post, w/S, Study 1; ns) and from 65% (pre) to 60% (post, w/S, Study 2; ns)
Volpp et al., 2008	Weight Loss	Healthy adults age 30-70 with a BMI of 30-40 from the Philadelphia VA Medical Center	Lottery incentive (expected value $3/day), or deposit contract with matching incentives (max $8.4/day)	7 months after end of intervention	Both in Lottery ($\Delta \approx$ -9 lbs., p=.01) and in deposit contract ($\Delta \approx$ -6 lbs., p=.03) participants weighed less than the beginning of the study
Angrist et al., 2009	Education	Entering first-year undergraduates at a primarily commuter school	Merit scholarship or merit scholarship with support service	1 year after end of intervention	0.28 percentage points (p<.01) increase in grade points for women; longer-term effect on men non-significant
Cawley & Price, 2009	Weight Loss	Employees from employer that has contract with ‘Company X’	Various quarterly monetary rewards or lottery plus refundable bonds	1 year after end of intervention	No significant difference with quarterly rewards w.r.t baseline. For lottery + bonds 3.6 lbs. (p<.05) loss w.r.t baseline

Charness & Gneezy, 2009	Gym Attendance	University of Chicago undergrad students	Low Reward: $25 to attend gym once in week. High Reward: $100 to attend gym 8 times in 4 weeks	One attendance measure per week, 7 weeks after intervention	Higher post-intervention gym attendance in high-reward vs control (0.67 visits/week) and vs. low-reward group (0.58 visits/week)
Acland & Levy, 2015	Gym Attendance	Self-reported non-regular gym attenders	$25 to attend gym once in week and then $100 to attend gym 2 times every week for 4 weeks	One attendance measure per week, 5 weeks post-treatment and following weeks into next semester	Higher post-intervention gym attendance in reward vs control (0.256 visits/week)
John et. al, 2011	Weight Loss	Patients at the Philadelphia Veterans Affairs Medical Center with BMIs of 30–40, age 30-70	Deposit contracts in which participants put $0-3 daily of their own money at risk (matched 1:1)	Weigh in 36 weeks after end of intervention	No longer-term difference in weight loss between treatment (1.2 Lbs.) and control (0.27 Lbs; p=.76)
Kimmel et. al, 2012	Warfarin Adherence	Warfarin patients at the UPenn Anticoagulation Management Center	Lottery with daily expected value of $3	6-months after end of intervention	No difference on anticoagulation levels between reward (23%) and control (25.9%; ns)
Royer et. al, 2015	Gym Attendance	Employees from Midwest Fortune 500 company	$10 for visiting company gym (up to 3x per wk.) over 4 wks., free membership, and $20 for new members; w/ or w/o self-funded commitment contracts	Gym use via login records 5-13 weeks and 14-52 weeks after end of incentives	Significantly higher post-intervention gym attendance in incentives vs control (0.11 visits/week; p <.05) in weeks 5-13 after incentives end. The results are directionally positive but *ns* in weeks 14-52 after incentives end
Bareket-Bojmel et. al 2014	Work productivity	Technicians at a global high-tech semiconductor company working at a fabrication plant in Israel.	$25, family pizza meal voucher, verbal reward, or own choice if performance level exceeded base productivity.	Productivity on the first day, second day, and third day after rewards were stopped.	Removal of the cash bonus significantly reduced performance by 13.2% relative to base productivity on day1. However, with verbal praise productivity was 4.2% higher than baseline on day 1.
Sen et. al, 2014	Adherence to medical regimen among diabetes patients	Patients of a Primary Care Medical Home at UPenn	Lottery incentive with expected daily value of $2.80 or $1.40 for daily monitoring	Every month for three months after end of incentives	After three months, adherence rate was 62 % in low, 35 % in high (p=.015) and 27 % in control (p=.002vs. low incentives).
Halpern et. al, 2015	Smoking Cessation	CVS Caremark employees	$800 or refundable deposit of $150	6 months and 12 months after end	Compared to the abstinence rate in control (6.0%), abstinence with individual

			with an opportunity to with $650 in rewards	of incentives	rewards was significantly higher after 6 months (≈15%, p<.01) and after 12 months (≈7.5%, p<.05).
Huffman & Bognanno, 2015	Work Productivity	Workers hired to register people for a company database during street festival	Hourly wage ($18) with a per sign-up monetary bonus ($5/sign-up) for 1 hour	Every hour for three hours after incentive ended	Immediately after incentives ended Reward group recruited more than Control (7% higher)
Mochan et. al, 2015	Purchase of healthy grocery items	Households participating in Points-based Healthy Food program	Forfeiting Heathy Food discount for failing to increase healthy food items purchased by 5% for the month.	Supermarket shopping data in following 6 months after intervention ended	0.49 percentage-points (p<.1) average increase in healthy items purchased in the first three months, and 0.79 percentage-points (p<.01) average increase in healthy items purchased in the next three months
Wang et. al, 2016	Number of hotel nights	Loyalty program members at a major international hotel chain	Bonus points during an 8-month period	Hotel stays in 8 months after the intervention	On Average, compared to the control group, the treatment group stayed one-night more in the post-reward period (p<.01)

Note: Only studies that (a) have measured post-reward behavior (and not self-report), (b) have studied adults (including a Study with 11th and 12th graders), and (c) where the target task is not a pro-social activity (e.g., contribution to charity, blood donation etc.) are included.

APPENDIX O Supplemental Reference

Acland, D., & Levy, M. R. (2015). Naiveté, projection bias, and habit formation in gym attendance. Management Science, 61(1), 146-160.

Angrist, J., Lang, D., & Oreopoulos, P. (2009). Incentives and services for college achievement: Evidence from a randomized trial. American Economic Journal: Applied Economics, 1(1), 136-163.

Bareket-Bojmel, L., Hochman, G., & Ariely, D. (2014). It's (Not) All About the Jacksons Testing Different Types of Short-Term Bonuses in the Field. Journal of Management, doi:0149206314535441.

Cawley, J., & Price, J. A. (2009). Outcomes in a program that offers financial rewards for weight loss (No. w14987). National Bureau of Economic Research.

Charness, G., & Gneezy, U. (2009). Incentives to exercise. Econometrica, 77(3), 909-931.

Garbarino, E., & Slonim, R. (2005). Pay-To-Perform Educational Policy: An Experiment Examining a Crowding Out Effect. Working Paper, University of Sydney.

Halpern, S. D., French, B., Small, D. S., Saulsgiver, K., Harhay, M. O., Audrain-McGovern, J., ... & Volpp, K. G. (2015). Randomized trial of four financial-incentive programs for smoking cessation. New England Journal of Medicine,372(22), 2108-2117.

Huffman, D., & Bognanno, M. (2015). Performance Pay and Workers' Non-Monetary Motivations: Evidence from a Natural Field Experiment. Working Paper, University of Pittsburgh

Jackson, C. K. (2010). A little now for a lot later a look at a texas advanced placement incentive program. Journal of Human Resources, 45(3), 591-639.

John, L. K., Loewenstein, G., Troxel, A. B., Norton, L., Fassbender, J. E., & Volpp, K. G. (2011). Financial incentives for extended weight loss: a randomized, controlled trial. Journal of General Internal Medicine, 26(6), 621-626.

Kimmel, S. E., Troxel, A. B., Loewenstein, G., Brensinger, C. M., Jaskowiak, J., Doshi, J. A., ... & Volpp, K. (2012). Randomized trial of lottery-based incentives to improve warfarin adherence.American Heart Journal, 164(2), 268-274.

Mochon, D., Schwartz, J., Maroba, J., Patel, D., & Ariely, D. (2016). Gain without pain: The extended effects of a behavioral health intervention. Management Science, forthcoming.

Royer, H., Stehr, M., & Sydnor, J. (2015). Incentives, commitments, and habit formation in exercise: evidence from a field experiment with workers at a Fortune-500 company. American Economic Journal: Applied Economics, 7(3), 51-84.

Sen, A. P., Sewell, T. B., Riley, E. B., Stearman, B., Bellamy, S. L., Hu, M. F., ... & Asch, D. A. (2014). Financial incentives for home-based health monitoring: a randomized controlled trial. Journal of General Internal Medicine, 29(5), 770-777.

Volpp, K. G., John, L. K., Troxel, A. B., Norton, L., Fassbender, J., & Loewenstein, G. (2008). Financial incentive–based approaches for weight loss: a randomized trial. Jama, 300(22), 2631-2637.

Volpp, K. G., Loewenstein, G., Troxel, A. B., Doshi, J., Price, M., Laskin, M., & Kimmel, S. E. (2008). A test of financial incentives to improve warfarin adherence. BMC Health Services Research, 8(1), 1.

Volpp, K. G., Levy, A. G., Asch, D. A., Berlin, J. A., Murphy, J. J., Gomez, A., ... & Lerman, C. (2006). A randomized controlled trial of financial incentives for smoking cessation. Cancer Epidemiology Biomarkers & Prevention, 15(1), 12-18.

Wang, Y., Lewis, M., Cryder, C., & Sprigg, J. (2016). Enduring effects of goal achievement and failure within customer loyalty programs: A large-scale field experiment. Marketing Science.