Supplemental Materials for

Rewarding cognitive effort increases the intrinsic value of mental labor

Table of Contents

**S1. Experiment 1 Additional Analyses** ................................................................. 2
  S1.1 Baseline Phase .......................................................................................... 2
  S1.2 Learning Phase .......................................................................................... 2
  S1.3 Subsequent Task Performance ................................................................... 4
  S1.4 Across-task analyses ............................................................................... 5
  S1.5 Additional Measures ............................................................................... 5

**S2. Experiment 2 Additional Analyses** ............................................................. 8
  S2.1 Sample Characteristics ........................................................................... 8
  S2.2 Learning Phase ....................................................................................... 8
  S2.3 Subsequent Task Performance ................................................................ 9
  S2.4 Additional Measures ............................................................................. 14
S1. Experiment 1 Additional Analyses

S1.1 Baseline Phase

No significant differences were found between the two experimental groups in any baseline CV-measure, $p > .410$, $\eta^2 < .007$, or subscale of the MDBF completed at baseline, $p > .627$, $\eta^2 < .003$.

S1.2 Learning Phase

Relationship between effort and reward

To determine the accuracy of our manipulation we tested whether mobilized effort during an N-back block predicted the amount of reward received for said block. To this end, we conducted a hierarchical linear model with random slopes and intercepts, treating the average PEP change score from each N-back block as nested within participants. Reward received for each block was the dependent measure, experimental group was the Level 2 predictor, and PEP change was the Level 1 predictor. Results indicated a significant interaction effect for PEP change score by group, $F(1, 1803) = 4.41, p = .036$, $\eta_p^2 = .002$, with PEP change scores predicting reward only in the experimental condition.

Reward by group

Results indicated from an ANOVA indicated that reward did significantly differ between our two groups, $F(1, 119) = 48.32, p < .001$, $\eta^2 = .289$, with the experimental condition ($M = 366.03, SD = 35.87$) receiving more reward on average than their control counterparts ($M = 329.62, SD = 18.18$). This effect, while not intended, is due to the overall high level of engagement seen from the participants in study 1. If participants chose to engage in the experimental condition they received the maximum reward value possible for every block of the learning phase. Since rewards were given randomly in the control condition, they cannot receive the maximum reward for every block. Importantly, Experiments 2a-e showed no significant group difference in rewards, $p > .126$, $\eta^2 < .010$. 
Table 1

**Pre-Ejection Period, Blood Pressure, and Heart Rate During the Baseline period, N-back Task, and Math Effort Task**

| Group    | Baseline       | N-Back Task    | Math Effort Task |
|----------|----------------|----------------|-----------------|
|          | Experimental   | Control        | Experimental    | Control        | Experimental | Control    |
|          | (N = 63)       | (N = 58)       | (N = 63)        | (N = 58)       | (N = 63)     | (N = 58)    |
| PEP      | M (SD)         |                |                 |                |              |            |
|          | 130.21 (14.53) | 129.43 (12.66) | 123.59 (15.60)  | 123.04 (13.91) | 124.94 (15.80) | 125.68 (13.09) |
| SBP      | M (SD)         |                |                 |                |              |            |
|          | 109.77 (10.84) | 111.45 (8.31)  | 116.58 (13.84)  | 117.38 (10.90) | 117.52 (14.44) | 117.83 (10.92) |
| DBP      | M (SD)         |                |                 |                |              |            |
|          | 64.25 (7.27)   | 65.43 (6.38)   | 68.45 (8.25)    | 69.64 (7.23)   | 68.77 (8.31)  | 69.99 (7.72) |
| MAP      | M (SD)         |                |                 |                |              |            |
|          | 81.18 (8.02)   | 82.41 (6.40)   | 86.59 (9.86)    | 87.62 (8.30)   | 87.13 (10.17) | 87.72 (8.84) |
| HR       | M (SD)         |                |                 |                |              |            |
|          | 75.70 (11.43)  | 78.99 (11.54)  | 78.81 (12.07)   | 83.13 (12.28)  | 77.99 (12.10) | 82.12 (11.51) |

**Note.** PEP = pre-ejection period; SBP = systolic blood pressure; DBP = diastolic blood pressure; MAP = mean arterial pressure; HR = heart rate; M = mean; SD = standard deviation; n = cell n. PEP is quantified in millisecond (ms). Blood pressure is quantified in millimeters of mercury (mmHg). HR is quantified in beats per minute (bpm). Lower PEP change values indicate greater arousal.
S1.3 Subsequent Task Performance

We analyzed accuracy on single trials of the Math Effort Task. Since performance on this task is primarily affected by level of difficulty, which participants choose themselves, it was important to control for level of difficulty on the level of each trial. To that means we ran a logistic hierarchical linear model with random slopes and intercepts, treating single math trials as nested within participants. Accuracy on each trial was the dependent measure, experimental group was the Level 2 predictor, and level of difficulty the Level 1 control variable. As expected, the simple effect of trial difficulty was significant, $B = -.96, SE = .050, z = -19.25, p < .001$. Importantly, experimental condition had no effect on accuracy, $B = -.036, SE = .18, z = -.20, p = .84$, suggesting that participants in the control condition did not disengage after selecting the more demanding tasks as compared to participants in the control condition. The proportion of correct responses on each difficulty level of the MET for Experiment 1 can be seen in Table 2.

Effort mobilization

To test if the difference in effort mobilization between groups varied across the MET, we compared the PEP change score for each block of the MET using a hierarchical linear model with random slopes and intercepts, treating block as nested within participants. PEP change score for each block was the dependent measure, experimental group was the Level 2 predictor and time and difficulty were the Level 1 predictors. Aside from a main effect of block, $F(1, 115) = 41.48, p < .0001$, with effort decreasing across time, neither the effect of group nor difficulty reached significance, $F < 1.18, p > .27$. 
Table 2

*Performance by group at each difficulty level of the MET*

| Difficulty Level | Experimental Mean (SD) | Control Mean (SD) |
|------------------|------------------------|-------------------|
| 1                | 0.93 (0.13)            | 0.98 (0.05)       |
| 2                | 0.88 (0.16)            | 0.91 (0.10)       |
| 3                | 0.78 (21 )             | 0.81 (0.15)       |
| 4                | 0.58 (0.27)            | 0.57 (0.26)       |
| 5                | 0.49 (0.30)            | 0.43 (0.32)       |

*Note.* Performance operationalized as percent of math problems answered correctly; $M =$ mean; $SD=$ standard deviation.

**S1.4 Across-task analyses**

**N-back performance predicting difficulty choice**

Exploratory analyzes indicated that participants performance on the N-back task predicted difficulty choice on the MET irrespective of group, $F(1, 117) = 5.53, p = .020, \eta^2 = .074$. This result while not predicted makes theoretical sense given the influence ability has on both performance and difficulty level selection.

**S1.5 Additional Measures**

The post-task questionnaire asked participants to rate their enjoyment of the MET overall, their enjoyment of each difficulty level of the MET, and the overall difficulty of the MET. Responses for all questions of the post-task questionnaire and post-task MDBF revealed no statistical significant main effects for group, $p > .162, \eta^2 < .03$. We also included Theories of Willpower (Job et al., 2010) and found a marginal effect of group, $F(1, 117) = 3.15, p = .079, \eta^2 = .026$. 

| Variable                  | 1    | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10   | 11   | 12   | 13   | 14   | 15   | 16   | 17   |
|--------------------------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 1. Age                   |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 2. Gender                | -0.16|      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 3. Math SC               | -0.08|      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 4. Diff choice           | -0.08|      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 5. N back perf           | -0.33*** |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 6. MET perf              | -0.13|      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 7. SBP N-back            | 0.05 |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 8. DBP N-back            | -0.14|      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 9. MAP N-back            | -0.06|      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 10. HR N-back            | -0.07|      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 11. PEP N-back           | 0.05 |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 12. SBP MET              | 0.05 |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 13. DBP MET              | -0.02|      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 14. MAP MET              | 0.01 |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 15. HR MET               | -0.14|      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 16. PEP MET              | -0.06|      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 17. WP                   | 0.05 |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 18. AMS                  | 0.10 |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |

Note: p < .001: ***; p < .01: **; p < .05: *. Gender: Female = 1, Male = 0. Math SC = math self concept; Diff choice = Average difficulty choice on MET; N back perf = N-back performance; MET perf = MET performance; SBP = systolic blood pressure; DBP = diastolic blood pressure; MAP = mean arterial pressure; HR = heart rate; WP = willpower beliefs; AMS = achievement motivation.
S2. Experiment 2 Additional Analyses

S2.1 Sample Characteristics

All participants in Experiments 2a-e lived in the US and were recruited from Amazon Mechanical Turk. Experiment 2a comprised a sample of 249 participants of which 228 fulfilled the inclusion criteria (112 male, 115 female), of mean age 38.4 (SD = 11.98). Experiment 2b comprised a sample of 303 participants of which 255 fulfilled the inclusion criteria (128 female, 125 male, 1 other) of mean age 38.88 (SD = 11.25). Experiment 2c comprised a sample of 281 participants, of which 241 fulfilled the inclusion criteria (116 female, 121 male, 3 other and 1 preferred not to say), of mean age 40.00 (SD = 11.75). Experiment 2d comprised a sample of 306, of which 233 fulfilled the inclusion criteria (106 female, 126 male) of mean age 37.07 (SD = 11.25). Experiment 2e comprised a sample of 530 participants, of which 500 fulfilled the inclusion criteria (184 males, 309 females, 7 other/no gender indicated), with a mean age of 41.10 (SD = 14.00).

S2.2 Learning Phase

Reward by group

To determine if our groups received significantly different rewards we performed an ANOVA. Results indicated that reward did not significantly differ between our two groups in any of the samples, $p > .125$, $\eta^2 < .011$.

Performance

We performed an ANOVA to determine if the groups performed differently on the N-back task. Performance did not differ significantly in any of the experiments, $p > .078$, $\eta^2 < .0079$. 
S2.3 Subsequent Task Performance

Difficulty levels

Mean difficulty levels selected by the experimental and control conditions, as well as test statistics for the effect of group on difficulty choice controlling for math self-concept in each of Experiments 2a-e can be seen in Table 4.

Table 4

*Difficulty choice in Experiments 2a-e*

| Study | N  | Mean difficulty choice (SD) | F   | p     | ηp²   | 90% CI          |
|-------|----|-----------------------------|-----|-------|-------|-----------------|
|       | Experimental | Control                     | Experimental | Control |       |                 |
| 2a    | 107 | 121                         | 2.38 (1.16) | 2.14 (1.06) | 3.04  | .083 | .013 | [0.00, 0.05]    |
| 2b    | 138 | 117                         | 2.43 (1.18) | 2.42 (1.05) | 0.099 | .754 | <.001 | [0.00, 0.01]    |
| 2c    | 129 | 112                         | 2.49 (1.06) | 2.30 (1.00) | 5.02  | .026 | .021 | [0.001, 0.06]   |
| 2d    | 113 | 120                         | 2.57 (1.10) | 2.32 (0.88) | 2.67  | .104 | .011 | [0.0, 0.04]     |
| 2e    | 258 | 242                         | 2.54 (1.06) | 2.29 (1.01) | 10.25 | .001 | .020 | [0.005, 0.05]   |

Difficulty choice across time

We used a hierarchical linear model to predict difficulty choice by group across time controlling for math self-concept (see Table 5).
Table 5

Coefficients in quadratic hierarchical linear model predicting difficulty choice across trials on the MET

| Study | Variable | B     | SE  | t    | p       | 95% CI       |
|-------|----------|-------|-----|------|---------|--------------|
| 2a    | Intercept| 1.06  | 0.22| 4.85 | <.0001  | [0.63, 1.49] |
|       | MSC      | 0.07  | 0.01| 6.13 | <.0001  | [0.05, 0.10] |
|       | Group    | -0.30 | 0.15| -1.95| 0.05    | [-0.59, 0.002] |
|       | Time     | 0.62  | 0.18| 3.48 | .0005   | [0.27, 0.97] |
|       | Time²    | -0.43 | 0.13| -3.26| .001    | [-0.69, -0.17] |
|       | Group:Time| 0.64  | 0.25| 2.61 | .009    | [0.16, 1.12] |
|       | Group:Time²| -0.79 | 0.18| -4.38| <.0001  | [-1.14, -0.44] |
| 2b    | Intercept| 1.34  | 0.24| 5.64 | <.0001  | [0.88, 1.81] |
|       | MSC      | 0.06  | 0.01| 4.23 | <.0001  | [0.03, 0.09] |
|       | Group    | -0.13 | 0.15| -0.86| 0.39    | [-0.42, 0.16] |
|       | Time     | 0.98  | 0.17| 5.70 | <.0001  | [0.65, 1.32] |
|       | Time²    | -0.96 | 0.13| -7.25| <.0001  | [-1.22, -0.70] |
|       | Group:Time| 0.15  | 0.25| 0.58 | 0.56    | [-0.35, 0.65] |
|       | Group:Time²| 0.03  | 0.20| 0.13 | 0.90    | [-0.36, 0.41] |
| 2c    | Intercept| 1.06  | 0.21| 5.08 | <.0001  | [0.65, 1.47] |
|       | MSC      | 0.08  | 0.01| 6.32 | <.0001  | [0.06, 0.11] |
|       | Group    | -0.19 | 0.14| -1.38| 0.17    | [-0.47, 0.08] |
|       | Time     | 1.03  | 0.18| 5.73 | <.0001  | [0.68, 1.38] |
|       | Time²    | -0.80 | 0.14| -5.87| <.0001  | [-1.07, -0.54] |
| Variable          | Estimate | Std. Error | t value | Pr(>|t|) | Lower 95% | Upper 95% |
|-------------------|----------|------------|---------|---------|-----------|-----------|
| Group:Time        | 0.13     | 0.26       | 0.51    | 0.61    | [-0.38, 0.65] |
| Group:Time²       | -0.44    | 0.20       | -2.20   | 0.03    | [-0.84, -0.05] |
| Intercept         | 1.33     | 0.26       | 5.18    | <.0001  | [0.83, 1.83] |
| MSC               | 0.06     | 0.01       | 4.28    | <.0001  | [0.03, 0.09] |
| Group             | -0.27    | 0.15       | -1.86   | 0.06    | [-0.56, 0.01] |
| Time              | 1.31     | 0.21       | 6.11    | <.0001  | [0.89, 1.73] |
| Time²             | -1.24    | 0.17       | -7.41   | <.0001  | [-1.57, -0.91] |
| Group:Time        | 0.44     | 0.30       | 1.47    | 0.14    | [-0.15, 1.02] |
| Group:Time²       | -0.47    | 0.23       | -1.99   | 0.05    | [-0.92, -0.01] |
| Intercept         | 1.19     | 0.13       | 8.89    | <.0001  | [0.93, 1.45] |
| Math self-concept | 0.07     | 0.01       | 9.00    | <.0001  | [0.06, 0.09] |
| Group             | -0.23    | 0.10       | -2.36   | 0.02    | [-0.43, -0.04] |
| Time              | 1.71     | 0.13       | 13.23   | <.0001  | [1.46, 1.96] |
| Time²             | -1.6     | 0.10       | -16.25  | <.0001  | [-1.79, -1.40] |
| Group:Time        | -0.18    | 0.19       | -0.97   | 0.33    | [-0.55, 0.18] |
| Group:Time²       | 0.14     | 0.14       | 1.02    | 0.31    | [-0.13, 0.42] |

**Note.** Group: 0 = experimental condition, 1 = control condition. Time = number of trials completed/total number of trials. Variables with 95% CI excluding zero shown in bold. MSC = math self-concept.
MET performance

Proportion of correct responses on the MET for each wave as well as test statistics for the effect of group on overall MET accuracy controlling for math self-concept in Experiments 2a-e can be seen in Table 6.

Table 6

*MET accuracy in Experiments 2a-e*

| Study | Mean accuracy (SD) | F   | p     | $\eta^2_p$ | 90% CI     |
|-------|--------------------|-----|-------|------------|------------|
|       | Experimental       | Control       |       |            |            |
| 2a    | 0.89 (0.16)        | 0.85 (0.24)   | 2.95  | .087       | .013 [0, 0.05] |
| 2b    | 0.84 (0.22)        | 0.81 (0.28)   | 1.28  | .258       | .005 [0, 0.03] |
| 2c    | 0.85 (0.20)        | 0.83 (0.24)   | 0.56  | .456       | .002 [0, 0.02] |
| 2d    | 0.80 (0.25)        | 0.76 (0.31)   | 0.82  | .367       | .004 [0, 0.03] |
| 2e    | 0.86 (0.14)        | 0.89 (0.15)   | 5.15  | .024       | .010 [0.0007, 0.03] |

Performance at each difficulty level

Proportion of correct responses on each difficulty level of the MET for Experiments 2a-e can be seen in Table 7.
Table 7  
Performance by group at each difficulty level of the MET

| Study | Difficulty Level | Mean (SD) | | | | |
|-------|------------------|----------|----------|----------|----------|
|       |                  | Experimental | Control |
| 2a    | 1                | 0.95 (0.23) | 0.86 (0.35) |
|       | 2                | 0.89 (0.31) | 0.88 (0.33) |
|       | 3                | 0.85 (0.36) | 0.84 (0.37) |
|       | 4                | 0.79 (0.41) | 0.77 (0.42) |
|       | 5                | 0.87 (0.34) | 0.77 (0.42) |
| 2b    | 1                | 0.92 (0.27) | 0.86 (0.34) |
|       | 2                | 0.89 (0.31) | 0.86 (0.35) |
|       | 3                | 0.80 (0.40) | 0.80 (0.40) |
|       | 4                | 0.68 (0.47) | 0.64 (0.48) |
|       | 5                | 0.70 (0.46) | 0.69 (0.46) |
| 2c    | 1                | 0.89 (0.31) | 0.87 (0.34) |
|       | 2                | 0.89 (0.31) | 0.84 (0.37) |
|       | 3                | 0.80 (0.40) | 0.80 (0.40) |
|       | 4                | 0.76 (0.43) | 0.78 (0.42) |
|       | 5                | 0.78 (0.42) | 0.73 (0.44) |
| 2d    | 1                | 0.83 (0.38) | 0.83 (0.37) |
|       | 2                | 0.85 (0.36) | 0.83 (0.38) |
|       | 3                | 0.80 (0.40) | 0.74 (0.44) |
|       | 4                | 0.71 (0.45) | 0.61 (0.49) |
|       | 5                | 0.76 (0.43) | 0.56 (0.50) |
| 2e    | 1                | 0.95 (0.21) | 0.93 (0.25) |
|       | 2                | 0.87 (0.33) | 0.91 (0.29) |
|       | 3                | 0.80 (0.40) | 0.84 (0.36) |
|       | 4                | 0.74 (0.44) | 0.78 (0.41) |
|       | 5                | 0.78 (0.41) | 0.82 (0.39) |

Logistic hierarchical linear model on MET performance

In order to control for the influence task difficulty has on performance we conducted a logistic hierarchical linear model on Experiments 2a-2e with random slopes and intercepts, treating single math trials as nested within participants. Accuracy on each trial was the dependent
measure, experimental group was the Level 2 predictor, and level of difficulty the Level 1 control variable. Findings indicated that the experimental group performed better than their control counterparts in Experiment 2a, that no statistical difference in performance occurred in Experiments 2b-2d, and that the control group performed better than their experimental counterparts to a marginally significant level, in Experiment 2e (see Table 8).

Table 8

*Coefficients of group in logistic hierarchical linear model predicting accuracy on the MET with difficulty level as level 1 predictor. Experimental condition dummy coded as 0, control condition as 1.*

| Study | B   | 95% CI       | z    | p   |
|-------|-----|--------------|------|-----|
| 2a    | -0.63 | [-1.21, -0.05] | -2.13 | .033 |
| 2b    | -0.12 | [-0.71, 0.47]  | -0.39 | .70  |
| 2c    | -0.19 | [-0.74, 0.36]  | -0.68 | .50  |
| 2d    | -0.33 | [-0.98, 0.31]  | -1.01 | .31  |
| 2e    | 0.26  | [0.00, 0.53]   | 1.94 | .053 |

S2.4 Additional Measures

**Study 2a.** We included the Need for Cognition scale (Cacioppo, Petty & Kao, 1982). Controlling for math self-concept, need for cognition was not significantly associated with difficulty choice, $F(1, 225) = 1.104, p = .295, \eta^2 = 0.004$, nor was there a significant interaction with group, $F(1, 233) = 0.395, p = .53, \eta^2 = 0.0015$. 
Table 9

Pearson correlation matrix for Experiment 2a (N = 228)

| Variable | 1 | 2 | 3 | 4 | 5 | 6 |
|----------|---|---|---|---|---|---|
| 1. Age   |   |   |   |   |   |   |
| 2. Gender| 0.04 |   |   |   |   |   |
| 3. Math SC | 0.07 | -0.21** |   |   |   |   |
| 4. Diff choice | 0.02 | -0.21** | 0.36*** |   |   |   |
| 5. N-back perf | 0.06 | 0.00 | 0.16* | 0.11 |   |   |
| 6. MET perf | 0.10 | 0.09 | 0.10 | 0.01 | 0.43*** |   |
| 7. NFC | 0.02 | -0.12 | 0.43*** | 0.21** | 0.22*** | 0.11 |

Note: p < .001: ***; p < .01: **; p < .05: *. Gender: Female = 1, Male = 0. Math SC = math self concept; Diff choice = Average difficulty choice on MET; N-back perf = N-back performance; MET perf = MET performance; NFC = need for cognition.

Study 2b. We included Theories of Willpower (Job et al., 2010) and action-state orientation (Kuhl & Beckmann, 1994), specifically the subscales relating to failure-related action orientation vs. preoccupation (AOF) and decision-related action orientation vs. hesitation (AOD). Neither AOF nor AOD moderated the relationship between condition and difficulty choice, p > .287, η² < 0.005. Willpower beliefs did moderate the effect of condition, F(1, 250) = 4.70, p = .031, η² = 0.017, such that those with non-limited willpower beliefs selected more difficult levels in the MET, but only in the experimental condition.

Table 10

Pearson correlation matrix for Experiment 2b (N = 255)

| Variable | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|----------|---|---|---|---|---|---|---|---|
| 1. Age   |   |   |   |   |   |   |   |   |
| 2. Gender | 0.03 |   |   |   |   |   |   |   |
| 3. Math SC | -0.04 | -0.02 |   |   |   |   |   |   |
| 4. Diff choice | 0.01 | -0.17** | 0.25*** |   |   |   |   |   |
| 5. N-back perf | 0.21*** | 0.08 | 0.18** | -0.03 |   |   |   |   |
| 6. MET perf | 0.15* | 0.04 | 0.07 | -0.21*** | 0.45*** |   |   |   |
| 7. WP | - | -0.05 | 0.29*** | 0.23*** | -0.12 | -0.14* |   |   |
| 8. AOD | 0.08 | -0.03 | 0.02 | 0.07 | 0.18** | 0.24*** | 0.03 |   |
| 9. AOF | 0.03 | -0.10 | 0.07 | 0.14* | 0.04 | 0.09 | 0.11 | 0.21*** |

Note: p < .001: ***; p < .01: **; p < .05: *. Gender: Female = 1, Male = 0. Math SC = math self concept; Diff choice = Average difficulty choice on MET; N-back perf = N-back performance; MET perf = MET performance; WP = willpower beliefs.
**Study 2c.** We included the Multidimensional State Boredom Scale (Fahlman et al., 2013) and the Boredom Proneness Scale (Farmer & Sundberg, 1986) and implicit theories about willpower (Job, Dweck, & Walton, 2010) as potential moderators. There was a main effect of willpower beliefs, $F(1, 236) = 4.38, p = .037, \eta^2 = 0.016$, with those with non-limited willpower beliefs selecting more difficult levels than those with limited beliefs, however the interaction with condition was not significant, $F(1, 236) = 0.012, p = .912, \eta^2 < 0.001$. State boredom did not moderate the effect of condition, $F(1, 236) = 0.557, p = .456, \eta^2 = 0.002$, nor did trait boredom $F(1, 232) = 1.421, p = .235, \eta^2 = 0.005$.

**Table 11**
*Pearson correlation matrix for Experiment 2c (N = 241)*

| Variable  | 1 | 2  | 3  | 4  | 5  | 6  | 7  | 8  |
|-----------|---|----|----|----|----|----|----|----|
| 1. Age    |   |    |    |    |    |    |    |    |
| 2. Gender |   | 0.02 |    |    |    |    |    |    |
| 3. Math SC|   | -0.10 | -0.18** |    |    |    |    |    |
| 4. Diff choice |   | -0.07 | -0.23*** | 0.36*** |    |    |    |    |
| 5. N-back perf |   | 0.15* | 0.18** | 0.10 | 0.13* |    |    |    |
| 6. MET perf |   | 0.19** | 0.16* | 0.02 | 0.00 | 0.38*** |    |    |
| 7. WP     |   | 0.08 | -0.15* | 0.18** | 0.20** | -0.02 | -0.06 |    |
| 8. MDBS   |   | -0.21** | -0.06 | -0.19** | -0.05 | -0.34*** | -0.26*** | -0.18** |
| 9. BPS    |   | -0.25*** | -0.11 | -0.13* | -0.05 | -0.30*** | -0.28*** | -0.17** | 0.74*** |

*Note:* p < .001: ***; p < .01: **; p < .05: *. Gender: Female = 1, Male = 0. Math SC = math self-concept; Diff choice = Average difficulty choice on MET; N-back perf = N-back performance; MET perf = MET performance; WP = willpower beliefs; MDBS = state boredom; BPS = trait boredom.

**Study 2d.** We included the revised Achievement Motives Scale (AMS-R; Lang & Fries, 2006) and found no group difference controlling for math self-concept, $F(1, 230) = .38, p = .54, \eta^2 = .0013$. 
Table 12

Pearson correlation matrix for Experiment 2d (N = 233)

| Variable     | 1    | 2    | 3    | 4    | 5    | 6    |
|--------------|------|------|------|------|------|------|
| 1. Age       |      |      |      |      |      |      |
| 2. Gender    | 0.13*|      |      |      |      |      |
| 3. Math SC   | -0.16*| -0.16*|      |      |      |      |
| 4. Diff choice| -0.08| -0.24***| 0.28***|      |      |      |
| 5. N-back perf| 0.06 | 0.14*| 0.18**| 0.05 |      |      |
| 6. MET perf  | 0.11 | 0.09 | 0.17**| -0.10| 0.47***|      |
| 7. AMS       | -0.11| -0.07| 0.46***| 0.27***| -0.02| 0.13*|

Note: p < .001: ***; p < .01: **; p < .05: *. Gender: Female = 1, Male = 0. Math SC = math self concept; Diff choice = Average difficulty choice on MET; N-back perf = N-back performance; MET perf = MET performance; AMS = achievement motivation.

Study 2e. The post-task questionnaire asked participants to rate their enjoyment of each difficulty level of the MET. ANCOVAs on the enjoyment rating of each level controlling for math self-concept produced a significant result for Level 2, $F(1, 434) = 5.382, p = 0.021, \eta^2 = 0.012$. Responses for all remaining questions of the post-task questionnaire produced no results of statistical significance, $p > .248, \eta_p^2 < .003$.

Table 13

Pearson correlation matrix for Experiment 2e (N = 500)

| Variable     | 1    | 2    | 3    | 4    | 5    | 6    |
|--------------|------|------|------|------|------|------|
| 1. Age       |      |      |      |      |      |      |
| 2. Gender    | 0.13*|      |      |      |      |      |
| 3. Math SC   | -0.16*| -0.16*|      |      |      |      |
| 4. Diff choice| -0.08| -0.24***| 0.28***|      |      |      |
| 5. N-back perf| 0.06 | 0.14*| 0.18**| 0.05 |      |      |
| 6. MET perf  | 0.11 | 0.09 | 0.17**| -0.10| 0.47***|      |
| 7. AMS       | -0.11| -0.07| 0.46***| 0.27***| -0.02| 0.13*|

Note: p < .001: ***; p < .01: **; p < .05: *. Gender: Female = 1, Male = 0. Math SC = math self concept; Diff choice = Average difficulty choice on MET; N-back perf = N-back performance; MET perf = MET performance; AMS = achievement motivation.