**Supplementary Information**

**for**

# When and why people perform mindless math

M. Asher Lawson[1]*, Richard P. Larrick[2], and Jack B. Soll[2]

[1]Decision Sciences Area, INSEAD, Boulevard de Constance, Fontainebleau, 77300, France

[2]The Fuqua School of Business, Duke University

**Author Note:**

*Corresponding author: asher.lawson@insead.edu

Data and materials available here: https://osf.io/mhe5q/

**Table of Contents**

*Studies S1, S2a, and S2b* ........ *4*

**Study S1** ........ **4**

Table S1. Responses to Study S1's items. ........ 4

**Study S2a** ........ **5**

Table S2. Responses to examples of the 'chips problem' ........ 5

Table S3. Response to verbal formulations of 'chips problem' ........ 5

Table S4. Responses to sheep item. ........ 5

**Study S2b** ........ **6**

Table S5. Response to verbal formulations of 'chips problem' (confound removed) ........ 6

*Study S3* ........ *7*

**Methods** ........ **7**

Table S6. Study items for Study S3 ........ 8

**Non-trick math filler items** ........ **9**

Table S7. Coding scheme. ........ 10

**Results** ........ **10**

Figure S1: Average number of questions answered with the *correct* and *MM* answers across numeric demand conditions *easier* and *harder* in Study S3. ........ 11

Table S8: Item level responses across numeric demand conditions *easier* and *harder* in Study S3. ........ 11

Table S9. Geometric mean response times for each condition and type of response. ........ 12

**Analytical details** ........ **12**

Table S10. Logistic regression models for Study S3. ........ 14

Table S11. Multinomial logistic regression models predicting 'correct' and 'MM' responses. ........ 15

Table S12. Goodness of fit indices for multinomial regression models. ........ 16

*Study S4* ........ *16*

**Methods** ........ **16**

**Results** ........ **17**

Table S13. Responses to Study S4 items. ........ 18

*Study 3* ........ *19*

**Table S14. Items for Study 3 – Solvers.** ........ **19**

**Session 1: Anticipated cognitive effort and problem prototypicality** ........ **20**

Table S15. The items Raters answered. ........ 21

**Session 2: Satisfaction and Feeling Impressive** ........ **21**

**Results** ........ **21**

Figure S2. Ratings of mechanism variables across different levels of numeric demands with math difficulty measured by geometric mean response time ........ 22

Figure S3. Ratings of anticipated cognitive effort and feeling of impressiveness plotted against the rate of correct responding. ........ 22

**Table S16. Mean, SD, CIs, & correlations between study variables in Study 3 (including Raters) ........ 24**

***Study S5 ........ 25***

**Table S17. Testing for inversion of effect with 'chips problem'. ........ 25**

**Table S18. Testing for inversion of effect for 'sheep problem' ........ 25**

**Table S19. Verbal formulations of 'chips problem'. ........ 26**

**Table S20. Examples of mindless math calculations. ........ 27**

***References ........ 28***

## Studies S1, S2a, and S2b

In addition to the primary studies in the manuscript, we ran several supplementary studies to provide data to substantiate various assertions in the arguments we proposed. We include details of three of these studies below. Each study followed the same experimental procedure and thus we will economize by writing the procedure up once.

We recruited participants from MTurk using Cloud Research. We set the restriction that participants had to have over 90% of their HITs accepted and restricted our sample to the Cloud Research approved participant pool. Upon entering the survey, participants answered the comprehension check item, “To let us know that you are not a robot, please type the letters of the word (not the numbers) you see in the following image into the response box below.”

If participants did not respond “printer” or “Printer”, they were ejected from the survey prior to random assignment. All participants were paid $0.10 for completing a single item, and we did not collect any additional information.


### Study S1

We recruited 302 participants with the aforementioned procedure. Participants were randomly assigned to answer one of 3 questions.

Table S1. Responses to Study S1’s items.

| | Percentage of responses… | | |
|---|---|---|---|
| Problem | 27 | 0 | Other |
| How many cubic feet of air are in an empty box that is 3’ deep x 3’ wide x 3’ long? (N = 99) | 70% | 12% | 18% (3, 9, 11, 18, 24, 26) |
| How many cubic feet of dirt are in an empty hole that is 3’ deep x 3’ wide x 3’ long? (N = 101) | 42% | 34% | 24% (3, 9, 72) |
| How many cubic feet of dirt are in an empty box that is 3’ deep x 3’ wide x 3’ long? (N = 102) | 54% | 26% | 20% (1, 3, 9, 12, 20, 24, 36) |

**Study S2a**

We recruited 1,209 participants with the aforementioned procedure. Participants were randomly assigned to answer one of 6 questions.

Table S2. Responses to examples of the 'chips problem'

| | Answer coding | | Percentage of answers… | | |
|---|---|---|---|---|---|
| Question | Correct | Mindless | Correct | Mindless | Other |
| Joey went to the store and bought a pack of chips. A bottle of water costs \$3.00, a pack of chips costs \$1.00 and a pack of gum costs \$2.00. How much did he spend in total? (N = 196) | 1.00 | 6.00 | 74% | 24% | 2% |
| Joey went to the store and bought a pack of chips. A bottle of water costs \$1.05, a pack of chips costs \$0.75 and a pack of gum costs \$1.70. How much did he spend in total? (N = 200) | 0.75 | 3.50 | 61% | 35% | 4% |

Table S3. Response to verbal formulations of 'chips problem'

| | Answer coding | | Percentage of answers… | | |
|---|---|---|---|---|---|
| Question | Correct | Mindless | Correct | Mindless | Other |
| Joey went to the store and bought a pack of chips. A pack of chips has a C on the front, a box of altoids has an A on the front, and a twix has a T on the front. At home, he looks down. What does he see? (N = 203) | C | CAT | 45% | 34% | 20% |
| Joey went to the store and bought a pack of chips. A twix has a T on the front, a pack of chips has a C on the front, and a box of altoids has an A on the front. At home, he looks down. What does he see? (N = 196) | C | TCA | 52% | 6% | 43% |

Table S4. Responses to sheep item.

| | Answer coding | | Percentage of answers… | | |
|---|---|---|---|---|---|
| Question | Correct | Mindless | Correct | Mindless | Other |
| At the end of fall, Gary had 6 sheep. All but 2 were eaten by wolves during the winter. | 2 | 4 | 88% | 10% | 2% |

| | | | | | |
|---|---|---|---|---|---|
| How many sheep did he have in the spring? (N = 208) | | | | | |
| At the end of fall, Gary had 106 sheep. All but 27 were eaten by wolves during the winter.<br><br>How many sheep did he have in the spring? (N = 206) | 27 | 79 | 86% | 14% | 0% |

## Study S2b

We recruited 407 participants with the aforementioned procedure. Participants were randomly assigned to answer one of 2 questions.

Table S5. Response to verbal formulations of 'chips problem' (confound removed)

| | Answer coding | | Percentage of answers… | | |
|---|---|---|---|---|---|
| Question | Correct | Mindless | Correct | Mindless | Other |
| Joey went to the store and bought a pack of chips. A pack of chips has a C on the front, a box of altoids has an A on the front, and a twix has a T on the front. At home, he looks down. What does he see? (N = 203) | C | CAT | 41% | 34% | 25% |
| Joey went to the store and bought a pack of chips. A pack of chips has a C on the front, a twix has a T on the front, and a box of altoids has an A on the front. At home, he looks down. What does he see? (N = 204) | C | CTA | 49% | 3% | 49% |

## Study S3

In Study S3, we replicated the results of Study 1 from the manuscript using a different design and a different set of items. Study S3 was performed before the present Study 1. In Study S3, we found support for all of our pre-registered hypotheses. We chose to substitute it out because we had devised items that we felt were cleaner demonstrations of role of problem difficulty in determining prevalence of mindless math. We present it here for transparency purposes to give readers the fullest view of which items are and are not likely to display the effects. In Study S3, all participants answered four study items. In addition to asking participants our study items, we measured respondents' disposition to reflect (Frederick, 2005) and ability to perform calculations (Cokely et al., 2012), as potential moderators of the effect of the numeric demands manipulation. Notably, items from the CRT (e.g., the bat and ball) are quite similar in nature to our study items. However, our primary reason for including the CRT in our analysis was to test whether high CRT people exhibited smaller hard-easy differences in their propensity to perform mindless math than other groups. For this aim, the similarity is not a concern.

### Methods

***Participants.*** Participants (N = 516) were recruited from Amazon's Mechanical Turk platform (registration: https://aspredicted.org/rf7wk.pdf). We excluded 66 participants prior to random assignment based on a pre-registered comprehension check, where participants were asked to remember a number from one page to the next. The final sample of 450 participants (Mean age = 36.9) contained 252 males and 195 females (3 people reported a gender of 'Other'). Participants were paid $2.00 for participation in the study and granted an additional $0.20 incentive for providing the correct answer to a randomly determined study question.

***Procedure.*** See Table S6 for a list of our items. All of the items had tempting math. For Question 2, for example, many people are tempted to add the costs of the three items, but the correct answer is simply the cost of chips.

Table S6. Study items for Study S3

| Q | *easier* tempting math | *harder* tempting math |
|---|---|---|
| 1 | On my way to the Himalayas, imagine I met a man with his wife going the opposite direction. His wife was carrying a sack and the sack had a cat in it. How many living creatures, in total, were going to the Himalayas? | On my way to the Himalayas, imagine I met a man with four wives going the opposite direction. Each wife was carrying a sack and each sack had two cats in it. How many living creatures, in total, were going to the Himalayas? |
| 2 | Imagine Joey is going to the store to buy a pack of chips. A bottle of water costs \$3.00, a pack of chips costs \$1.00 and a pack of gum costs \$2.00. How much does he spend in total? **(in dollars)** | Imagine Joey is going to the store to buy a pack of chips. A bottle of water costs \$1.05, a pack of chips costs \$0.75 and a pack of gum costs \$1.70. How much does he spend in total? **(in dollars)** |
| 3 | Imagine 5 candles stand burning in a dining room. A strong breeze blows in through an open window and extinguishes 2 of them. Assuming the wind doesn't extinguish any more candles, how many candles do you have left in the end? | Imagine 17 candles stand burning in a dining room. A strong breeze blows in through an open window and extinguishes 8 of them. Assuming the wind doesn't extinguish any more candles, how many candles do you have left in the end? |
| 4 | Imagine I have a red box. In red boxes there are 2 oranges and 3 apples. In blue boxes there are 2 bananas and 2 peaches. How many bananas do I have? | Imagine I have 3 red boxes. In red boxes there are 8 oranges and 2 apples. In blue boxes there are 6 bananas and 4 peaches. How many bananas do I have? |

We manipulated numeric demands by varying the numbers featured in an item. The wording and structure of the problem otherwise remained identical. For Question 2, the easier numbers on the left included prices of \$3.00, \$1.00, and \$2.00. The harder numbers on the right included prices of \$1.05, \$0.75, and \$1.70. If one were tempted to add numbers within a problem, the left-hand numbers are easy to add, whereas the right-hand numbers take more effort (since they involve holding places and carrying over dollars).

After providing consent, participants answered the four study items. These items had three blocks of four randomized filler questions between them. These filler questions were non-trick math problems, and are included in full in the supplementary materials. By including these filler questions, we meant to reduce the likelihood that participants perceived a trick in our study items. If they spotted a trick in an earlier item, it could lead to them discarding their intuitions in subsequent items. However, given that these were numeric questions, the filler items could also have primed the use of mathematical strategies.

## Non-trick math filler items

f1 - Imagine you are competing in a darts competition. You start off requiring a score of 501 to win. The points you register will lead to an amount being subtracted from 501 until you reach 0. In order to win ('check out') you must hit exactly the score required to reduce the total to 0 from your final 3 darts. What is the maximum score you can achieve from throwing 3 darts?

f2 - Imagine you are going to the store to buy the lunches for your class for a school trip. The school trip is 7 days long. You need to buy lunches for 9 students for the entire trip. In addition, on the first day you need to buy lunch for yourself and another teacher. How many lunches should you buy total?

f3 - Imagine that you need to buy some supplies for your 5 cars. If a car requires seat covers, you must buy 4 seat covers. 2 of your cars already have seat covers. Your friend agrees to give you 3 spare seat covers she already has. How many car seat covers do you need to buy?

f4 - Imagine you are a school teacher. You have to give an assembly to a class of 500 students. Some of the students are entitled to a free juice box as part of a recent scheme. This only applies to 60% of the students. How many juice boxes do you need to buy?

f5 - Imagine you need to divide 56 sweets between a group of children. There are 3 girls and 4 boys. You should divide the sweets equally. How many sweets does each child get?

f6 - Imagine you are organising a convoy to travel to see a concert. It is a favourite band of your local church and a large group of you want to go. 5 people with 5 seater sedans volunteer to drive. In addition, you can take the church's minibus which seats 8 people, including the driver. How many spaces, including drivers, do you have?

f7 - Imagine you are in charge of the budget of your high school prom. You are told that no more than 20% of the budget should be used to purchase refreshments. The total budget is $3000. How much does that leave you (minimum) left to spend on venue hire and other expenses?

f8 - A pen and a pencil cost $2.00 together. The pen costs $0.50. How much does the pencil cost? (in dollars)

f9 - For the purposes of mathematical calculations, we sometimes have to use approximations. Imagine you want to calculate the diameter of a circle, but only have pen and paper. Accordingly, you decide to round pi to 2 decimal places. The diameter of the circle is 2cm. What is its circumference?

f10 - Imagine you have to plan how many rations to bring on a camping trip. There are 10 children and 6 adults. Children only require 1/2 of a ration allocation, owing to being smaller. How many ration allocations do you need to bring, in total?

f11 - Imagine you are the mayor of a small town of 6000 people. As part of a recent initiative, you hope to supply a free orange to some groups of residents, at a cost of $1 a week. The residents entitled to this free orange are the local unemployed (10%) and those in the bottom paying jobs (another 10%). How much will this initiative cost, per week?

f12 - Imagine you are hosting a dinner party. You need to work out how many of different items to purchase. Each person needs 2 appetizers, 1 entree and 1 glass of wine. Each of these costs $3. You invited 20 people, but 4 could not attend. How much will your supplies cost?

The order of presentation of study items was varied using a Latin square design, with random assignment to one of four order conditions. After completing the study items, participants were presented with a randomized block containing the three items from the Cognitive Reflection Test (Frederick, 2005) and the four items from the Berlin Numeracy Test (Cokely et al., 2012). Our manipulated variable was the numeric demands of the tempting math in the study items. All study items in a condition were presented at the same level of math difficulty (either *easier* or *harder*).

Table S7. Coding scheme.

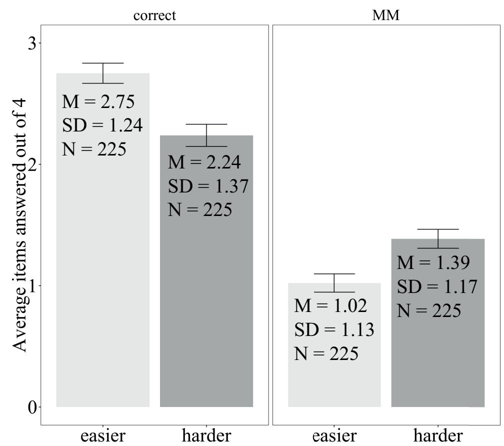
| Q | condition | correct | MM |
|---|---|---|---|
| 1 | easier | 1 | 3 or 4 |
| 1 | harder | 1 | 13 or 14 |
| 2 | easier | 1 | 6 |
| 2 | harder | 0.75 | 3.5 |
| 3 | easier | 5 | 3 |
| 3 | harder | 17 | 9 |
| 4 | easier | 0 | 2 |
| 4 | harder | 0 | 6 or 18 |

## Results

As shown in Figure S1, increasing the numeric demands reduces solution rates (2.24 vs. 2.75) and increases the number of errors that involve mindless math (1.39 vs. 1.02).[1] The proportion of errors that involved mindless math was approximately equal across conditions (79% vs. 82%, see Table S8). Participants who were higher in dispositional reflection and numeracy returned more correct answers and fewer mindless math answers, but neither individual difference score moderated the effect of our manipulation.

[1] We analyze these data formally in the SI using Generalized Estimation Equations (see Table S10). Both effects are statistically significant.

Figure S1: Average number of questions answered with the *correct* and *MM* answers across numeric demand conditions *easier* and *harder* in Study S3.




*Note*. The error bars indicate plus or minus 1 standard error.

Table S8: Item level responses across numeric demand conditions *easier* and *harder* in Study S3.

| | | | Proportion of responses… | | |
|---|---|---|---|---|---|
| **Problem** | **Numeric Demands** | **N** | **Correct** | **MM** | **Neither** |
| 1 | easy | 225 | 0.79 | 0.16 | 0.05 |
| 1 | hard | 225 | 0.56 | 0.26 | 0.18 |
| 2 | easy | 225 | 0.78 | 0.18 | 0.04 |
| 2 | hard | 225 | 0.60 | 0.31 | 0.09 |
| 3 | easy | 225 | 0.47 | 0.42 | 0.11 |
| 3 | hard | 225 | 0.38 | 0.55 | 0.07 |
| 4 | easy | 225 | 0.72 | 0.26 | 0.02 |
| 4 | hard | 225 | 0.70 | 0.27 | 0.03 |

We also inspected the response times of participants in each condition providing each type of answer (see Table S9). Participants took similar amounts of time to return *correct* responses regardless of condition (easier = 22s; hard = 25s), but longer to return *MM* responses when the numeric demands were *harder* (27 seconds) than when they were *easier* (19s).[2] We suggest that respondents might engage in mindless math due to allocating attention away from noticing the correct problem representation. This claim is supported by the response times—

[2] We tested this by estimating a random intercept linear regression model predicting log response times, with items nested within participant. We removed participants who returned neither MM nor correct responses, and used the type of response, the condition, and their two-way interaction as independent variables, with fixed effects for participants' scores in the CRT and BNT to mitigate selection concerns. We found a significant negative effect of MM responding ($b$ = -0.190, $p$ < 0.001), a significant positive effect of harder numeric demands ($b$ = 0.108, $p$ = 0.027), and a significant interaction between MM and harder condition ($b$ = 0.205, $p$ = 0.002).

those who returned incorrect, mindless answers did take longer when the numbers were larger, whereas those who reached correct answers did not. By resisting the urge to allocate time away from representation, effective problem solvers avoided the need to complete onerous computations, thus minimizing the effect of numeric demands on response time.

Table S9. Geometric mean response times for each condition and type of response.

| Condition | Response Type | N | Geometric Mean RT |
|---|---|---|---|
| Easier | Correct | 619 | 22 |
| | MM | 230 | 19 |
| | Neither | 51 | 21 |
| Harder | Correct | 504 | 25 |
| | MM | 312 | 27 |
| | Neither | 84 | 16 |

## Analytical details

We tested our hypothesis formally using logistic regression models estimated with Generalized Estimation Equations. GEEs are used to estimate the parameters of generalized linear models when outcomes may be correlated. Using this modelling approach also allowed us to incorporate both question and position effects, as well as individual difference measures, in the prediction of a participant's response to each item. To conduct this analysis, we used the 'geepack' package (Hojsgaard, Halekoh and Yan, 2016) for R, clustering outcomes by participant. We specified a binomial distribution family with a logit link, an 'exchangeable' correlation between responses, following guidelines to minimize the Quasi Information Criterion (Hardin and Hilbe, 2003; Pan, 2001; Hin and Wang, 2009).

The numeric demands condition was coded as a dummy-variable where *harder* = 1 indicated the *harder* numeric demands condition. We mean centered participants' scores in both the CRT and the BNT, and included fixed effects for participants' education, age, and gender. The six logit models can be seen in Table S10. We estimated three models predicting the rate of *correct* responding (Models 1-3) and three models predicting the rate of *MM* responding (Models 4-6). Although these two sets of models are not independent (note that *MM* and correct responding are mutually exclusive), distinguishing between the two types of responses adds value in that it allows us to show that when participants are incorrect they tend to actually perform the calculations dictated by mindless math, as opposed to answering randomly. Harder numeric demands significantly predicted a lower rate of *correct* responses ($b$ = -0.614; $p$ < 0.001; *Model 1*) and a higher rate of *MM* responses ($b$ = 0.480; $p$ < 0.001; *Model 4*). Additionally, we found that CRT score significantly predicted a higher rate of *correct* responses ($b$ = 0.479; $p$ < 0.001; *Model 1*) and a lower rate of *MM* responses ($b$ = -0.380; $p$ < 0.001; *Model 4*). Similarly, BNT score was associated with a higher rate of *correct* responses ($b$ = 0.141; $p$ = 0.031; *Model 1*) and a lower rate of *MM* responses ($b$ = -0.153; $p$ = 0.018; *Model 4*). We found some evidence that participants answered items that were presented later in the block correctly at a higher rate.

Notably, we did not find evidence that cognitive reflection or numeracy moderated the effects of our numeric demands manipulations on the rate of *correct* responding. In Models 2 and 3, the interactions between *harder* and the individual difference measures did not attain statistical significance; neither the interaction with CRT (*b* = -0.165; *p = 0.207*) nor the BNT (*b* = 0.017; *p = 0.879*) was significant. In the prediction of *MM*, the interactions between *harder* numeric demands and both CRT (*b* = 0.201; *p = 0.096*) and BNT (*b* = 0.127; *p = 0.254*) were both positive, but not significant.

***Perfectly counterbalanced samples*** Five randomly drawn samples were taken from the complete study sample with Latin square order exactly counterbalanced across cells (n = 46 per cell, total N = 368). The main effect of our manipulation on 'correct' and 'MM' was statistically significant in all sub-samples. We use the full sample for all of our primary analyses, but pre-registered our intention to perform this robustness check.

***Model specification*** *A priori*, we considered using a multinomial logit model to analyze our data. Our multinomial logit model estimates and their goodness-of-fit metrics are included below (Tables S11 and S12). In our analyses above, we focus on the results of logistic regression models estimated using Generalized Estimation Equations. We believe the GEE models to be superior to multinomial logit in this instance for three main reasons. First, GEE models allow for us to model a 'position' effect for when an item is presented to a respondent. Secondly, GEE models are less prone to overfitting the sample data, allowing for greater generalizability. Finally, GEE models yield a single coefficient for each independent variable, which facilitates easier interpretation of effects.

Table S10. Logistic regression models for Study S3.

| Model | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| (Intercept) | 0.867*** | 0.884*** | 0.793*** | -1.368*** | -1.390*** | -1.304*** |
| | (0.163) | (0.165) | (0.158) | (0.176) | (0.178) | (0.173) |
| Harder (H) | -0.614*** | -0.657*** | -0.537*** | 0.480*** | 0.541*** | 0.450** |
| | (0.142) | (0.141) | (0.141) | (0.139) | (0.140) | (0.139) |
| CRT | 0.479*** | 0.629*** | | -0.380*** | -0.556*** | |
| | (0.074) | (0.090) | | (0.070) | (0.087) | |
| BNT | 0.141* | | 0.320*** | -0.153* | | -0.373*** |
| | (0.065) | | (0.079) | (0.065) | | (0.082) |
| H x CRT | | -0.165 | | | 0.201† | |
| | | (0.131) | | | (0.121) | |
| H x BNT | | | 0.017 | | | 0.127 |
| | | | (0.113) | | | (0.111) |
| Education | -0.029 | -0.010 | -0.052 | 0.004 | -0.016 | 0.022 |
| | (0.068) | (0.067) | (0.069) | (0.064) | (0.062) | (0.065) |
| Age | 0.012† | 0.012† | 0.019** | -0.005 | -0.097 | -0.112 |
| | (0.007) | (0.007) | (0.006) | (0.006) | (0.140) | (0.140) |
| Male | 0.022 | 0.042 | 0.058 | -0.077 | -0.005 | -0.012† |
| | (0.145) | (0.146) | (0.145) | (0.140) | (0.006) | (0.006) |
| Q2 | 0.084 | 0.084 | 0.078 | 0.213 | 0.212 | 0.208 |
| | (0.132) | (0.132) | (0.125) | (0.152) | (0.151) | (0.147) |
| Q3 | -1.168*** | -1.166*** | -1.110*** | 1.385*** | 1.381*** | 1.341*** |
| | (0.132) | (0.131) | (0.126) | (0.148) | (0.147) | (0.144) |
| Q4 | 0.211 | 0.210 | 0.195 | 0.320* | 0.318* | 0.314* |
| | (0.134) | (0.133) | (0.126) | (0.156) | (0.155) | (0.151) |
| P2 | 0.171 | 0.171 | 0.161 | -0.248† | -0.246† | -0.242† |
| | (0.126) | (0.126) | (0.120) | (0.139) | (0.139) | (0.135) |
| P3 | 0.263* | 0.262* | 0.253* | -0.311* | -0.309* | -0.306* |
| | (0.132) | (0.132) | (0.127) | (0.144) | (0.144) | (0.140) |
| P4 | 0.459*** | 0.460*** | 0.428*** | -0.494*** | -0.493*** | -0.476*** |
| | (0.135) | (0.134) | (0.128) | (0.148) | (0.147) | (0.143) |

N = 450, n = 1800
Note. *** p < 0.001, ** p < 0.01, * p < 0.05, † p < 0.10

Table S11. Multinomial logistic regression models predicting 'correct' and 'MM' responses.

| | | | Coefficients and p-values from Wald Tests | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Model | DV | Marginal | (Intercept) | Harder (H) | CRT | BNT | Education | H X CRT | H X BNT |
| 1 | correct | 1 | 1.70 | -0.89* | 0.46* | 0.07 | -0.28 | | |
| | | 2 | 1.32 | -0.54 | 0.76*** | -0.02 | -0.30 | | |
| | | 3 | 0.73 | -1.39*** | 0.86*** | 0.34 | -0.19 | | |
| | | 4 | 0.30 | -1.62*** | 1.21*** | 0.27 | -0.21 | | |
| 2 | correct | 1 | 1.55 | -0.84 | 0.69* | 0.08 | -0.27 | -0.24 | |
| | | 2 | 1.08 | -0.28 | 1.04*** | 0.00 | -0.30 | -0.38 | |
| | | 3 | 0.16 | -0.44 | 1.29*** | 0.38* | -0.18 | -0.69 | |
| | | 4 | -0.26 | -0.63 | 1.63*** | 0.3 | -0.21 | -0.68 | |
| 3 | correct | 1 | 1.37 | -0.48 | 0.28 | 0.48* | -0.28 | | -0.30 |
| | | 2 | 1.14 | -0.33 | 0.12 | 0.76*** | -0.30 | | -0.18 |
| | | 3 | 0.22 | -0.64 | 0.62 | 0.88*** | -0.18 | | -0.46 |
| | | 4 | -0.11 | -1.07 | 0.51 | 1.23*** | -0.21 | | -0.37 |
| 4 | MM | 1 | 1.70 | 1.08 | 1.57 | 0.12 | -0.89 | | |
| | | 2 | 1.32 | 1.08 | 1.22 | 0.22 | -0.54 | | |
| | | 3 | 0.73 | 1.05 | 0.69 | 0.49 | -1.39 | | |
| | | 4 | 0.30 | 1.08 | 0.28 | 0.78 | -1.62 | | |
| 5 | MM | 1 | -0.15 | 0.56 | -0.33 | -0.15 | 0.14 | 0.07 | |
| | | 2 | 0.35 | 0.02 | -0.59*** | -0.36* | 0.09 | 0.39 | |
| | | 3 | 1.21 | 0.01 | -0.98*** | -0.35* | 0.01 | 0.45 | |
| | | 4 | -0.84 | 1.08 | -1.17* | -0.13 | 0.05 | 0.15 | |
| 6 | MM | 1 | -0.36 | 1.05* | -0.07 | -0.31* | 0.14 | | -0.14 |
| | | 2 | 0.27 | 0.43 | -0.47* | -0.43*** | 0.09 | | 0.25 |
| | | 3 | 1.02 | 0.53 | -0.43* | -0.78*** | 0.01 | | 0.20 |
| | | 4 | -0.43 | 0.77 | -0.43 | -1.17*** | 0.04 | | 0.47 |

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, † $p < 0.10$

Table S12. Goodness of fit indices for multinomial regression models.

| Model | DV | Nagelkerke Pseudo R2 | AIC | BIC | loglik(m) | loglik(null) |
|---|---|---|---|---|---|---|
| 1 | correct | 0.237 | 1311.679 | 1393.864 | -635.840 | -693.450 |
| 2 | correct | 0.246 | 1314.259 | 1412.881 | -633.129 | -693.450 |
| 3 | correct | 0.240 | 1317.658 | 1416.279 | -634.829 | -693.450 |
| 4 | MM | 0.181 | 1236.384 | 1318.569 | -598.192 | -640.248 |
| 5 | MM | 0.188 | 1240.676 | 1339.298 | -596.338 | -640.248 |
| 6 | MM | 0.188 | 1240.834 | 1339.456 | -596.417 | -640.248 |

***Discussion.*** Study S3 largely corroborated the results of Study 1, with a few additional findings. Notably, CRT and BNT were associated with higher performance in the task, but did not moderate the effect of the numeric demands manipulation, and the similar response times for *correct* responses in the harder and easier numeric demands conditions suggested that people do not do math when they get the problems right, but we needed to test this directly in Study 2. Moreover, the results of Study S3 beg the question, is the higher rate of mindless math when numeric demands are higher a product of priming participants with a mathematical strategy? In Study S3 we used filler items that were math problems to reduce participants' perception of being tricked, but these may also have contributed to participants' strategy selection process . We addressed these concerns in Study 1, Study 2, and Study S4.

## Study S4

Study S4 was a first iteration of what is now Study 1. In a pre-registered design, we tested whether 5 items would display the hypothesized hard-easy effect when manipulating the numeric demands of normatively equivalent problems (https://osf.io/2y6vb). Each participant answered all 5 problems, either in the *easier* or *harder* numeric demands condition. In the end, three items displayed the expected pattern, and these effects were replicated in Study 1 in a between-subjects design. We include this study to avoid overrepresenting the likelihood of items displaying an effect where increased difficulty leads to people performing mindless math at a higher rate.

### Methods

***Participants.*** We recruited 620 participants from Amazon's Mechanical Turk, using CloudResearch approved participants and requiring an acceptance rate of over 90% of HITs. 19 participants were excluded prior to random assignment for failing a comprehension check—leaving a final sample of 601 (average age = 39.6y; 58% men).

***Procedure.*** Participants were randomly assigned to a numeric demands condition (*easier* versus *harder*) and a Latin square order condition. All participants answered five items in the same numeric demands condition, with the order of presentation varied. After completing the five problems, participants answered questions from the CRT (Frederick, 2005) and the BNT

(Cokely et al., 2012), before providing demographic information. The problems and their coding schemes are listed in full below:

**P1**

Joey went to the store and bought a pack of chips. A bottle of water costs $3.00, a pack of chips costs $1.00 and a pack of gum costs $2.00. How much did he spend in total? (in dollars)

- Correct = 1; mm = 6

Joey went to the store and bought a pack of chips. A bottle of water costs $1.05, a pack of chips costs $0.75 and a pack of gum costs $1.70. How much did he spend in total? (in dollars)

- Correct = 0.75; mm = 3.50

**P2**

A man is packing away 10 rolls of toilet paper into crates. A box contains 5 rolls. A crate contains 2 boxes. How many boxes does he need?

- Correct = 2; mm = 1, 10, 5

A man is packing away 288 rolls of toilet paper into crates. A box contains 12 rolls. A crate contains 6 boxes. How many boxes does he need?

- Correct = 24; mm = 4, 288, 48, 72, 144, 6

**P3**

Billy has 4 pens. He gives me 2 pencils. How many pens does he have left?

- Correct = 4; mm = 2, 6

Billy has 28 pens. He gives me 9 pencils. How many pens does he have left?

- Correct = 28, mm = 19, 37

**P4**

At the end of fall, Gary had 4 sheep. All but 3 were eaten by wolves during the winter. How many sheep did he have left in the spring?

- Correct = 3, Mindless = 1, 7

At the end of fall, Gary had 23 sheep. All but 8 were eaten by wolves during the winter. How many sheep did he have left in the spring?

- Correct = 8, Mindless = 23, 15

**P5**

How many cubic feet of dirt are in an empty box that is 2’ deep x 1’ wide x 3’ long?

- Correct = 0, Mindless = 6

How many cubic feet of dirt are in an empty box that is 3’ deep x 3’ wide x 3’ long?

- Correct = 0, Mindless = 27

## Results

The results are shown in Table S13. The first three items exhibited the predicted hard-easy effect, where increasing numeric complexity reduces the rate of *correct* responding and increases the rate of *MM* responding. For problems 4 and 5, this pattern was to some extent reversed (though not significantly).

To formally analyze these data, we used logistic regression with cluster robust standard errors to predict whether a participant got each of the five answers *correct*, and whether each response was an *MM* response. Harder numeric demands had a statistically significant negative effect on the rate of *correct* responding ($b$ = -0.238, $p$ = .010). The effect on *mindless math* responding was positive but not significant ($b$ = 0.065, $p$ = .489). We conducted further pre-registered tests to verify whether participants' CRT scores or BNT scores moderated the effect of the numeric demands manipulation on either *correct* or *MM* responding. None of the interactions came close to achieving statistical significance. When just analyzing the first three items, there was a significant negative effect of numeric demands on accuracy ($b$ = -0.505, $p$ = < .001) and a significant positive effect of numeric demands on *mindless math* responding ($b$ = 0.331, $p$ = .005).

Table S13. Responses to Study S4 items.

| | | | Proportion of responses… | | |
|---|---|---|---|---|---|
| **Problem** | **Numeric Demands** | **N** | **Correct** | **MM** | **Neither** |
| 1 | easy | 302 | 0.64 | 0.34 | 0.02 |
| 1 | hard | 299 | 0.58 | 0.35 | 0.07 |
| 2 | easy | 302 | 0.86 | 0.09 | 0.04 |
| 2 | hard | 299 | 0.70 | 0.20 | 0.10 |
| 3 | easy | 302 | 0.83 | 0.17 | 0.00 |
| 3 | hard | 299 | 0.76 | 0.23 | 0.01 |
| 4 | easy | 302 | 0.84 | 0.14 | 0.02 |
| 4 | hard | 299 | 0.89 | 0.11 | 0.00 |
| 5 | easy | 302 | 0.35 | 0.54 | 0.11 |
| 5 | hard | 299 | 0.35 | 0.45 | 0.20 |

***Discussion.*** We discuss the results of Studies S3 and S4 in the main manuscript as an aid to interpreting the results of Study 1. In short, we found evidence that higher numeric demands led to fewer correct answers and more *mindless* ones in 4/4 problems in Study S3 and 3/5 problems in Study S4. One problem featured in both studies ("chips problem"). The magnitude of the effect size varied quite substantially, and some effect sizes may be debilitatingly small: for example, for a difference between a 0.80 and 0.84 rate of *correct* answers to be statistically significant in a Fisher's exact t.test, you would need cell sizes of around 740. If the compared success rates were 0.50 and 0.54, this required cell size increases to around 1250. We present the most robust items in Study 1 in the manuscript, and offer the above analysis as a small step to combat the file drawer problem.

## Study 3

### Table S14. Items for Study 3 – Solvers.

| Question number | No trick version | Trick version |
|---|---|---|
| 1 | Imagine Joey is going to the store to buy a bottle of water, a pack of chips and a pack of gum. A bottle of water costs $3.00, a pack of chips costs $1.00 and a pack of gum costs $2.00. How much does he spend in total? (in dollars) | Imagine Joey is going to the store to buy a pack of chips. A bottle of water costs $3.00, a pack of chips costs $1.00 and a pack of gum costs $2.00. How much does he spend in total? (in dollars) |
| 2 | Imagine Joey is going to the store to buy a bottle of water, a pack of chips and a pack of gum. A bottle of water costs $4.00, a pack of chips costs $2.00 and a pack of gum costs $5.00. How much does he spend in total? (in dollars) | Imagine Joey is going to the store to buy a pack of chips. A bottle of water costs $4.00, a pack of chips costs $2.00 and a pack of gum costs $5.00. How much does he spend in total? (in dollars) |
| 3 | Imagine Joey is going to the store to buy a bottle of water, a pack of chips and a pack of gum. A bottle of water costs $5.00, a pack of chips costs $1.00 and a pack of gum costs $7.00. How much does he spend in total? (in dollars) | Imagine Joey is going to the store to buy a pack of chips. A bottle of water costs $5.00, a pack of chips costs $1.00 and a pack of gum costs $7.00. How much does he spend in total? (in dollars) |
| 4 | Imagine Joey is going to the store to buy a bottle of water, a pack of chips and a pack of gum. A bottle of water costs $3.10, a pack of chips costs $1.30 and a pack of gum costs $2.00. How much does he spend in total? (in dollars) | Imagine Joey is going to the store to buy a pack of chips. A bottle of water costs $3.10, a pack of chips costs $1.30 and a pack of gum costs $2.00. How much does he spend in total? (in dollars) |
| 5 | Imagine Joey is going to the store to buy a bottle of water, a pack of chips and a pack of gum. A bottle of water costs $3.30, a pack of chips costs $1.40 and a pack of gum costs $2.20. How much does he spend in total? (in dollars) | Imagine Joey is going to the store to buy a pack of chips. A bottle of water costs $3.30, a pack of chips costs $1.40 and a pack of gum costs $2.20. How much does he spend in total? (in dollars) |
| 6 | Imagine Joey is going to the store to buy a bottle of water, a pack of chips and a pack of gum. A bottle of water costs $3.25, a pack of chips costs $1.50 and a pack of gum costs $2.65. How much does he spend in total? (in dollars) | Imagine Joey is going to the store to buy a pack of chips. A bottle of water costs $3.25, a pack of chips costs $1.50 and a pack of gum costs $2.65. How much does he spend in total? (in dollars) |
| 7 | Imagine Joey is going to the store to buy a bottle of water, a pack of chips and a pack of gum. A bottle of water costs $3.15, a pack of chips costs $1.45 and a pack of gum costs $2.65. How much does he spend in total? (in dollars) | Imagine Joey is going to the store to buy a pack of chips. A bottle of water costs $3.15, a pack of chips costs $1.45 and a pack of gum costs $2.65. How much does he spend in total? (in dollars) |
| 8 | Imagine Joey is going to the store to buy a bottle of water, a pack of chips and a pack of gum. A bottle of water costs $3.15, a pack of chips costs $2.34 and a pack of gum costs $1.85. How much does he spend in total? (in dollars) | Imagine Joey is going to the store to buy a pack of chips. A bottle of water costs $3.15, a pack of chips costs $2.34 and a pack of gum costs $1.85. How much does he spend in total? (in dollars) |

We present the results of Study 3's solvers in the main manuscript. In addition to these, we ran two studies collecting *ratings* of the featured problems to offer suggestive evidence regarding the potential mechanisms that explain the link between math difficulty and doing mindless math / problem accuracy. We delineate the results of these additional studies below.

We call them "Raters" studies as part of Study 3—given that the data was analyzed jointly—but they were distinct experimental sessions.

Specifically, the Raters aspect of Study 3 aimed to test four competing explanations for why harder numeric demands might spur participants to spend less time framing the problem and instead jump into doing math. First, it could be the case that participants quickly determine how hard the math is and this determines how quickly they want to get underway with it. In this case, the difference in performance between conditions is explained by participants in the harder numeric demands condition non-consciously allocating attentional resources away from framing the problem and toward the computation. Another possible explanation is that the different levels of numeric demands differentially prime strategies in problem solvers. In this case, the difference between conditions is explained by the problem in the harder numeric demands condition more closely resembling a prototypical math problem, and thus more strongly evoking mathematical strategies. The third mechanism proposes that anticipating heightened personal satisfaction could explain people's elevated propensity to do math when the math was harder. The final mechanism focuses on how impressive solving the math would be: if respondents expect that solving harder math is more impressive, this might motivate them to jump into the math to attain this social capital.

We sought evidence to distinguish between these possible mechanisms in Study 3 using an experiment where we collected data from two groups of participants: Solvers and Raters (in 2 sessions). We expanded our range of numeric demands to eight different levels to study how people's perceptions of the problems varied across this wider range. The Raters rated one of the eight different chips problems on the relevant dimension for one of the four proposed mechanisms. This allowed us to evaluate what properties of the harder math led people to jumping into doing it. We include full details of these analyses below.

We conducted two different sessions with Raters. Participants in the first session rated the difficulty and prototypicality of the chips problem (first two columns of Table S15). Those in the second session rated how satisfying and impressive it would be to complete the math in the problem (second two columns of Table S15).

## Session 1: Anticipated cognitive effort and problem prototypicality

***Participants.*** We recruited 913 participants from MTurk (registration: https://osf.io/m27fy). We removed 112 participants prior to random assignment using a comprehension check, as per our other studies. In the final sample of 801 participants, the mean age was 39.6 years, with 49.7% of participants identifying as women, and 0.3% non-binary. We paid participants $0.31 for the completion of the study.

***Procedure.*** Participants saw a single non-conflict chips problem, and responded to items regarding their perceptions of it. We used the non-conflict version of the problem because Study 3 aimed to distinguish why participants are either satisfied that the numerical problem representation of the chips problem is correct (and thus return the MM answer), or scrutinize it further to reach the correct answer. Specifically, we randomly assigned participants in this study to two conditions: a level of numeric demands (from 1 to 8; conditions that corresponded to the same levels for the Solvers as shown in Table S14) and which items they rated the problem on (from Table S15). All participants saw their version of the chips problem for 10 seconds before being automatically advanced to the next page. Participants could not return to inspect the item for longer. We warned participants that they would only get to see the problem for 10 seconds before answering several questions on the problem. We imposed this time limit on participants so we could measure their initial perceptions of the problem, rather than their post-hoc rationalizations after having solved it. After seeing their chips problem, participants reported their response to each item on 7-point scales.

Table S15. The items Raters answered.

| Anticipated Cognitive Effort | Prototypicality of Math Problem | Personal satisfaction | Perceived impressiveness |
|---|---|---|---|
| How hard do you think it would be to solve? | How much did it look like a typical math problem? | How satisfying do you think solving it would be? | How impressive do you think solving it would be? |
| How much effort do you think it would take to solve? | How much did it resemble typical math problems? | How fulfilled would you feel from solving it? | How much would others be impressed by you solving it? |
| When you saw this problem, were you drawn to start adding the numbers? | When you saw this problem, were you drawn to start adding the numbers? | How proud would you feel having solved it? | How proud would you feel having solved it? |

In the first session, we measured two types of perception of the chips problem across the range of numeric demands levels—the extent to which they anticipated cognitive effort associated with the problem ($\alpha = 0.88$) and the extent to which the problem was prototypical of a math problem ($\alpha = 0.93$). We also measured participants' temptation to do the math over the range of numeric demands (the third item in Table S15).[3]

## Session 2: Satisfaction and Feeling Impressive

We recruited 837 participants from MTurk. 36 participants were excluded prior to random assignment for failing our comprehension check, and the final sample had a mean age of 42.2 with 55.6% men. The items we collected ratings of in session 2 of the Raters study are shown in the right two columns of Table S15. The measure of personal satisfaction had an alpha of 0.90, whereas the perceived impressiveness scale had an alpha of 0.93. Participants indicated significantly higher pride in the condition rating personal satisfaction ($t(799) = 3.84$, $p < .001$).

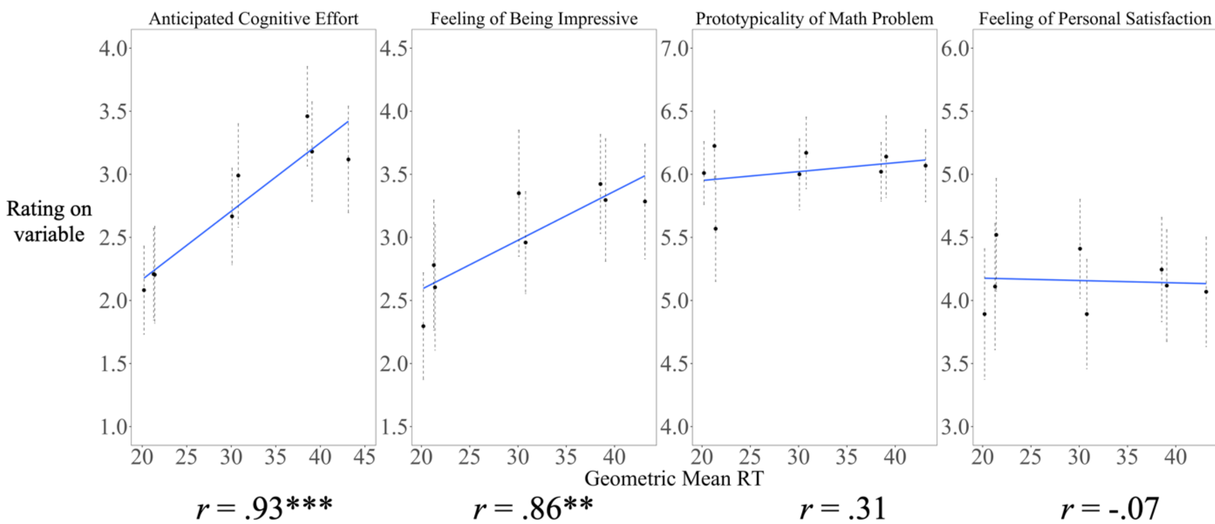
## Results

There was a clear positive relationship between the difficulty of the math (as measured by the geometric mean response time) and anticipated cognitive effort and respondents' feelings that

[3] There was a significant main effect of what rating we elicited in items 1-2 on this variable (participants asked about the prototypicality of the item returned higher ratings of their temptation to do the math), but no correlation between participants' responses and the numeric demands of the problem. This was an exploratory analysis.

solving the problem would be impressive (see Figure S2). The relationship between difficulty and problem prototypicality and feelings of satisfaction were much weaker. In one-tailed tests of the Pearson's product-moment correlation coefficient (with n = 8), we found a significant positive correlation between math difficulty and anticipated cognitive effort ($r$ = 0.93, $p$ < .001) and between math difficulty and perceived impressiveness ($r$ = 0.86, $p$ = .003). The relationship with math difficulty was not significant for problem prototypicality ($r$ = 0.31, $p$ = .224) or personal satisfaction ($r$ = -0.07, $p$ = .570).

<u>Figure S2. Ratings of mechanism variables across different levels of numeric demands with math difficulty measured by geometric mean response time.</u>

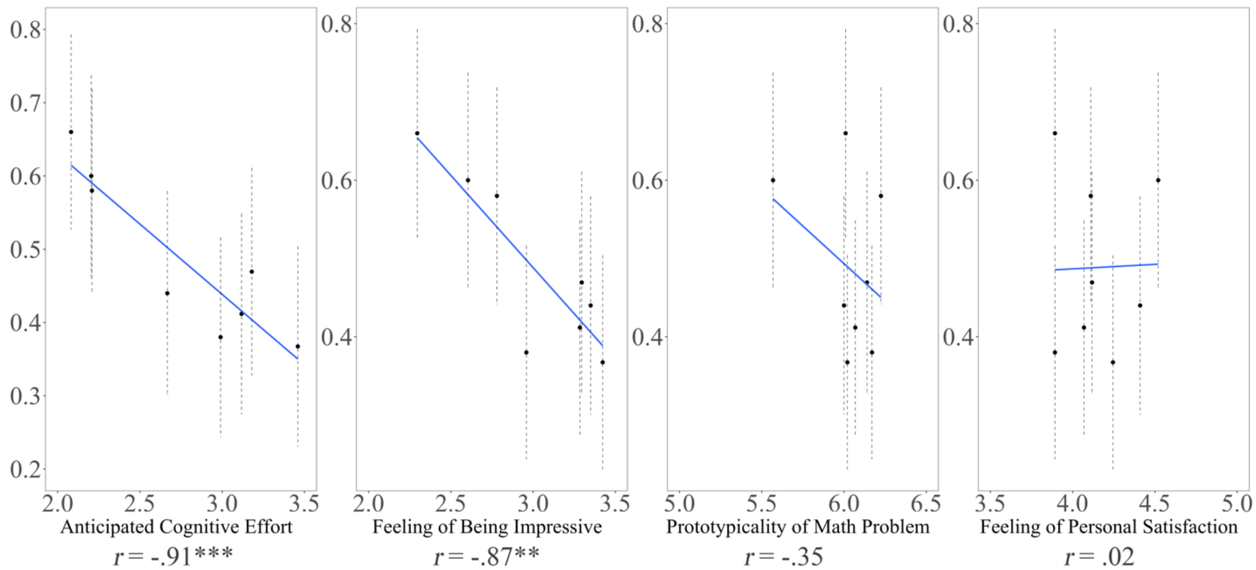



*Note.* Error bars indicate 95% confidence intervals.

This analysis showed that as the difficulty of tempting math increased, people perceived that there would be greater cognitive effort associated with it, and felt that it would be more impressive to do the math. We next tested whether these perceptions of the problems predicted whether people answered them correctly. To do so, we inspected correlations between these ratings and the rate of correct responses to a problem. Across the 8 items, participants' mean ratings of the anticipated cognitive effort associated with each item correlated negatively with the rate of correct responses ($r$ = -0.91, $p$ < .001), as did perceptions of how impressive it would be to solve the problem ($r$ = -0.87, $p$ = .002). The correlation between participants' ratings of the prototypicality of the items as math problems and the rate of correct responding was also negative but was not significant ($r$ = -0.35, $p$ = .197), and the correlation with perceived satisfaction was essentially 0 ($r$ = 0.02, $p$ = .522). We plot these relationships in Figure S3.

Overall, the evidence suggests that anticipated cognitive effort and/or perceptions of impressiveness could figure centrally in the mechanism explaining the relationship between numeric demands and problem solving accuracy.

<u>Figure S3. Ratings of anticipated cognitive effort and feeling of impressiveness plotted against the rate of correct responding.</u>




*Note.* Error bars indicate 95% confidence intervals.

***Discussion*** In Study 3's Raters section, we tested four possible explanations for why people might choose to start doing the math at a higher rate when the math was more difficult. We found that neither problem prototypicality nor the personal satisfaction associated with completing a problem significantly correlated with problem difficulty, suggesting that these are not compelling explanations. On the other hand, two potential explanations for the observed hard-easy effect emerged—anticipated cognitive effort and feeling impressive. However, these tests were correlational and offer suggestive, rather than conclusive, evidence of a mechanism.

## Table S16. Mean, SD, CIs, & correlations between study variables in Study 3 (including Raters)

| Variable | *M* | *SD* | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Anticipated Cognitive Effort (non-conflict) | 2.74 | 0.52 | | | | | | | | | | | |
| 2. Prototypicality of Math Problem (non-conflict) | 6.03 | 0.20 | .34<br>[-.48, .84] | | | | | | | | | | |
| 3. Temptation to do the Math (non-conflict) | 5.91 | 0.15 | -.10<br>[-.75, .65] | .60<br>[-.18, .92] | | | | | | | | | |
| 4. Feeling of Personal Satisfaction (non-conflict) | 4.16 | 0.23 | -.08<br>[-.74, .66] | -.71<br>[-.94, -.00] | -.76*<br>[-.95, -.12] | | | | | | | | |
| 5. Feeling of being Impressive (non-conflict) | 3.00 | 0.41 | .86**<br>[.41, .98] | .31<br>[-.51, .83] | -.27<br>[-.82, .54] | .24<br>[-.56, .81] | | | | | | | |
| 6. Geometric Mean RT (non-conflict) | 30.56 | 9.04 | .93**<br>[.67, .99] | .31<br>[-.50, .83] | .00<br>[-.70, .71] | -.07<br>[-.74, .66] | .86**<br>[.39, .97] | | | | | | |
| 7. Geometric Mean RT (conflict) | 27.54 | 4.10 | .89**<br>[.49, .98] | .27<br>[-.54, .82] | -.07<br>[-.74, .66] | -.06<br>[-.74, .67] | .71*<br>[.01, .94] | .71*<br>[.01, .94] | | | | | |
| 8. Geometric Mean RT of Correct Responses (conflict) | 24.51 | 3.86 | .55<br>[-.25, .91] | .22<br>[-.57, .80] | .11<br>[-.65, .75] | -.07<br>[-.74, .67] | .39<br>[-.44, .86] | .32<br>[-.50, .83] | .86**<br>[.38, .97] | | | | |
| 9. Geometric Mean RT of Incorrect Responses (conflict) | 30.52 | 4.77 | .94**<br>[.70, .99] | .15<br>[-.62, .77] | -.25<br>[-.81, .55] | .01<br>[-.70, .71] | .77*<br>[.14, .96] | .85**<br>[.35, .97] | .89**<br>[.51, .98] | .55<br>[-.26, .90] | | | |
| 10. Proportion of Correct Responses (non-conflict) | 0.93 | 0.05 | -.59<br>[-.91, .20] | -.59<br>[-.92, .19] | -.09<br>[-.75, .65] | .45<br>[-.37, .88] | -.50<br>[-.89, .32] | -.63<br>[-.93, .13] | -.24<br>[-.81, .56] | .14<br>[-.63, .77] | -.41<br>[-.87, .41] | | |
| 11. Proportion of Correct Responses (conflict) | 0.49 | 0.11 | -.91**<br>[-.98, -.58] | -.35<br>[-.85, .47] | .23<br>[-.57, .80] | .02<br>[-.69, .72] | -.87**<br>[-.98, -.44] | -.82*<br>[-.97, -.28] | -.80*<br>[-.96, -.23] | -.50<br>[-.89, .31] | -.79*<br>[-.96, -.20] | .64<br>[-.12, .93] | |
| 12. Proportion of MM Responses (conflict) | 0.45 | 0.08 | .49<br>[-.32, .89] | .37<br>[-.45, .85] | -.21<br>[-.80, .58] | -.22<br>[-.80, .57] | .50<br>[-.32, .89] | .47<br>[-.36, .88] | .27<br>[-.54, .82] | -.02<br>[-.72, .69] | .33<br>[-.49, .84] | -.81*<br>[-.96, -.24] | -.75*<br>[-.95, -.10] |

*Note.* *M* and *SD* are used to represent mean and standard deviation, respectively. Values in square brackets indicate the 95% confidence interval for each correlation. The confidence interval is a plausible range of population correlations that could have caused the sample correlation (Cumming, 2014). * indicates $p < .05$. ** indicates $p < .01$.

## Study S5

Study S5 followed the same procedure as Studies S1 and S2 but is placed later in the supplement to reflect when it was referenced in the manuscript. We recruited 914 participants with the aforementioned procedure. Participants were randomly assigned to answer one of 13 questions. Some of these questions will be familiar, as they were initial pilots of items later used in Study S2.

**Table S17. Testing for inversion of effect with 'chips problem'.**

| | Answer coding | | Percentage of answers… | | |
|---|---|---|---|---|---|
| Question | Correct | Mindless | Correct | Mindless | Other |
| Joey went to the store and bought a pack of chips. A bottle of water costs \$3.00, a pack of chips costs \$1.00 and a pack of gum costs \$2.00. How much did he spend in total? (in dollars) (N = 73) | 1.00 | 6.00 | 58% | 41% | 3% |
| Joey went to the store and bought a pack of chips. A bottle of water costs \$1.05, a pack of chips costs \$0.75 and a pack of gum costs \$1.70. How much did he spend in total? (in dollars) (N = 70) | 0.75 | 3.50 | 46% | 51% | 3% |
| Joey went to the store and bought a pack of chips. A bottle of water costs \$0.69, a pack of chips costs \$0.93 and a pack of gum costs \$3.43. How much does he spend in total? (in dollars) (N = 71) | 0.93 | 5.05 | 52% | 39% | 8% |
| Joey went to the store and bought a pack of chips. A bottle of water costs \$3.13, a pack of chips costs \$2.43 and a pack of gum costs \$7.26. How much does he spend in total? (in dollars) (N = 68) | 2.43 | 12.82 | 63% | 29% | 7% |

**Table S18. Testing for inversion of effect for 'sheep problem'.**

| | Answer coding | | Percentage of answers… | | |
|---|---|---|---|---|---|
| Question | Correct | Mindless | Correct | Mindless | Other |
| At the end of fall, Gary had 6 sheep. All but 2 were eaten by wolves during the winter. | 2 | 4 | 79% | 14% | 7% |

| | | | | | |
|---|---|---|---|---|---|
| How many sheep did he have in the spring? (N = 72) | | | | | |
| At the end of fall, Gary had 106 sheep. All but 27 were eaten by wolves during the winter.<br><br>How many sheep did he have in the spring? (N = 72) | 27 | 79 | 85% | 12% | 3% |

**Table S19. Verbal formulations of 'chips problem'.**

| | Answer coding | | Percentage of answers… | | |
|---|---|---|---|---|---|
| Question | Correct | Mindless | Correct | Mindless | Other |
| Joey went to the store and bought a pack of chips. A pack of chips has a C on the front, a box of Altoids has an A on the front, and a twix has a T on the front. He gets home and looks down. What does it spell? (N = 72) | C | CAT | 14% | 79% | 7% |
| Joey went to the store and bought a pack of chips. A twix has a T on the front, a pack of chips has a C on the front, and a box of Altoids has an A on the front. He gets home and looks down. What does it spell? (N = 71) | C | TCA | 24% | 6% | 70% |
| Joey went to the store and bought a pack of chips. A pack of chips has a C on the front, a box of altoids has an A on the front, and a twix has a T on the front. At home, he looks down. What does he see? (N = 66) | C | CAT | 42% | 41% | 17% |
| Joey went to the store and bought a pack of chips. A twix has a T on the front, a pack of chips has a C on the front, and a box of altoids has an A on the front. At home, he looks down. What does he see? (N = 69) | C | TCA | 58% | 3% | 39% |

Note that both of the verbal formulations of the 'chips problem' displayed the effect whereby the easily reached verbal answer "cat" was preferred to the less fluent "tca", but we opted for the second formulation as "what does he see" was less of a strong cue to responding with a word than was "what does it spell", which could be seen to imply a word was the desired responses.

**Table S20. Examples of mindless math calculations.**

| | Answer coding | | Percentage of answers… | | |
|---|---|---|---|---|---|
| Question | Correct | Mindless | Correct | Mindless | Other |
| What's the chance of flipping at least 1 head in 2 flips? (N = 69) | 0.75 | 0.50 | 12% | 32% | 57% |
| What's the chance of flipping at least 1 head in 3 flips? (N = 70) | 0.88 | 0.33 | 11% | 16% | 73% |
| What's the chance of flipping at least 3 heads in 5 flips? (N = 71) | 0.50 | 0.60 | 8% | 0% | 92% |
| Joey went to the store and bought a pack of chips. A bottle of water costs $3.00, a pack of chips costs $1.00 and a pack of gum costs $2.00. How much did he spend in total? (N = 73) | 1.00 | 6.00 | 58% | 41% | 3% |
| Joey went to the store and bought a pack of chips. A bottle of water costs $1.05, a pack of chips costs $0.75 and a pack of gum costs $1.70. How much did he spend in total? (N = 70) | 0.75 | 3.50 | 46% | 51% | 3% |

# References

Hardin, J. W., & Hilbe, J. M. (2003). *Generalized estimating equations.* Wiley.

Hin, L.-Y., & Wang, Y.-G. (2009). Working-correlation-structure identification in generalized estimating equations. *Statistics in Medicine*, *28*(4), 642–658. https://doi.org/10.1002/sim.3489

Højsgaard, S., Halekoh, U., & Yan, J. (2016). *Package 'geepack.'* http://cran.univ-paris1.fr/web/packages/geepack/geepack.pdf

Pan, W. (2001). Akaike's Information Criterion in Generalized Estimating Equations. *Biometrics*, *57*(1), 120–125. https://doi.org/10.1111/j.0006-341X.2001.00120.x