When and why people perform mindless math

M. Asher Lawson*   Richard P. Larrick†   Jack B. Soll‡

Abstract

In this paper, we show that the presence of numbers in a problem tempts people to perform mathematical operations even when the correct answer requires no math, which we term "mindless math". In three pre-registered studies across two survey platforms (total N = 3,193), we investigate how mindless math relates to perceived problem difficulty, problem representation, and accuracy. In Study 1, we show that increasing the numeric demands of problems leads to more mindless math (and fewer correct answers). Study 2 shows that this effect is not caused by people being wary of problems that seem too easy. In Study 3, we show that this effect is robust over a wider range of numeric demands, and in the discussion we offer two possible mechanisms that would explain this effect, and the caveat that at even harder levels of numeric demands the effect may invert such that *much* harder math increases accuracy relative to *moderately* hard math.

Keywords: mindless math, problem solving, dual-process theory, intuitive errors

1   Introduction

If asked *How many cubic feet of **air** are in an empty box that is 3' deep x 3' wide x 3' long*?[1] most (70%, N = 99) respondents correctly compute 27.[2] However, when asked the variant,

---

*INSEAD. Email: asher.lawson@duke.edu. https://orcid.org/0000-0001-5782-3583.

†The Fuqua School of Business, Duke University. https://orcid.org/0000-0002-0491-426X.

‡The Fuqua School of Business, Duke University. https://orcid.org/0000-0002-4496-1976.

[1]These problems are modified from a CRT-2 item (Thomson & Oppenheimer, 2016). The original wording is, "How many cubic feet of dirt are in an empty hole that is 3' deep x 3' wide x 3' long?" (N = 101, 42% respond 27, and 34% respond 0). For further details of these supplementary studies, refer to the supplement.

[2]An additional 12% of participants responded with 0, which could be due to familiarity with the original Thomson & Oppenheimer version (2016), or construing "empty" as also meaning empty of air.

*How many cubic feet of **dirt** are in an empty box that is 3' deep x 3' wide x 3' long?*, most (54%, N = 102) *still* answer 27, compared to just 26% who recognize that an *empty* box contains 0 cubic feet of *dirt*. The presence of a *possible* calculation leads respondents to perform it, even though it is unnecessary. Though it is not hard, it is hard enough to be rewarding. We call this process "mindless math".

In the present research, we find that mindless math is, somewhat ironically, *more* prevalent in problems involving slightly harder numeric operations. We believe that salient, simple but non-trivial math tempts people to perform it, and correspondingly distracts their attention from properly representing the problem, as is required for solution.

1.1   Problem representation and problem solving

To solve a problem, one must first *represent* it: to internally construct a problem space, with the problem's elements, its goal, and the permissible operations one can use to meet that goal (Greeno & Simon, 1988; Newell & Simon, 1972; Reimann & Chi, 1989, Johnson-Laird, 1983; McNamara, 1994). For example, in the *empty box* problem at the start of the paper "empty" is an important element of the problem to encode. If, however, a problem solver quickly represents the problem as a computation of spatial volume, they would focus on depth, width, and length as the key elements, and unnecessarily multiply them to reach an incorrect answer.

One of the barriers to solving the *empty box* problem is that one can easily produce an incorrect answer and never receive corrective feedback. It stands in contrast to many classic problems in creativity research (Wiley & Jarosz, 2012), such as the nine-dot (Chein, Weisberg, Streeter & Kwok, 2010) and matchstick problems (Ash & Wiley, 2006). For these problems it is often clear that a solution has not been reached; the respondent feels "stumped"; and is made to realize that their current representation is inaccurate. In contrast, many problems in everyday life are *opaque*: they map onto familiar structures and operations that allow a decision maker to reach a fast, seemingly satisfactory answer. They offer no obvious feedback that you have in fact failed. The ease of adopting familiar structures and operations reduces the tendency for problem solvers to scrutinize the accuracy of their approach, and so people don't spend sufficient time *representing* these problems to reach correct answers.

The problems we will use to demonstrate mindless math have similar properties to the bat and ball problem — the first item of the Cognitive Reflection Test (Frederick, 2005) — *A bat and a ball cost $110 in total. The bat costs $100 more than the ball. How much does the ball cost?* In this problem, there is no corrective feedback, and the presence of the erroneous intuition ($10) prevents the correct conclusion from being drawn even when checks are attempted (Meyer & Frederick, 2022). This is partly because even when checking this item, the problem solver can actively endorse $10 as a response (Meyer & Frederick, 2022) by explicitly affirming that the ball costs $10 and the bat costs $100, for a total of $110. In our problems, the presence of a possible calculation serves the role of

the $10 intuition: it can direct the checking of the answer to verify the calculation was completed successfully, thus leading the supervisory system to actively endorse that the correct answer has been reached. This is different from "stumpers" problems (Bar-Hillel et al., 2018, 2019), where there is an erroneous, intuitive interpretation of the problem stem — people form a mental model based on incomplete information that makes the problem *appear* unsolvable — but this model is recognized as inappropriate.

1.2 Why do people do mindless math?

There is a long history of research showing that familiarity can impede insight. For example, when someone typically uses an object in a specific way, they can become unable to conceive of alternative uses. This is called *functional fixedness* (Duncker, 1945; Adamson, 1952). A related phenomenon is the Einstellung effect, where people develop predispositions to solving problems in a particular manner even when better alternatives are available. As an example, Sheridan and Reingold (2013) used eye tracking data to show that the presence of a familiar solution to a chessboard situation blocks the discovery of better solutions, even among expert chess players.

Given that most people have significant experience with simple math problems, calculation strategies are likely to be a familiar approach, and thus may impede considering other approaches. This will especially be the case if the content of a problem evokes a calculation strategy. Though our problems don't require math, they take the form of problems that do. Problems that contain numbers evoke a script to perform mathematical operations (Bassok, 1996). This is generally an adaptive response. It aids people in solving everyday math problems, such as the cost of different combinations at a restaurant. But, in cases where the tempting and intuitively cued math is not the correct approach, this schematic response to numeric information will impede reasoners' ability to correctly answer the problem (Cooper & Sweller, 1987; Luchins & Luchins, 1950).

Though it is reasonable to construe problems mentioning numbers as math problems, it is less clear how varying the difficulty of the tempting but unnecessary calculations will affect people's propensity to perform the irrelevant math.

1.3 How does problem difficulty affect the rate of mindless math?

We have argued that the presence of numerical information can lead people to misrepresent a problem, leading them to return mathematical answers to fundamentally non-mathematical problems (mindless math). In the present research, we consider how these effects vary as the tempting but unnecessary math in a problem becomes harder.

Consider the following question: *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?* There are two primary answers that people reach when faced with this problem: the price of the pack of chips ($1), which is correct, and the sum of the

three items ($6). Performing the sum to reach $6 is incorrect, but in a specific way. It is not random noise. People who reach this answer have misrepresented the problem as requiring a sum, and perform mindless math to reach their answer. In this example where performing the tempting math was relatively straightforward, 74% of participants responded correctly with $1, and 24% with the mindless math answer of $6 (N = 196).

In contrast, consider the same problem with harder numbers: *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?* The numbers here are harder to add because it involves carrying digits to reach a final sum. The greater difficulty of this addition may steer people away from spending the time needed to correctly represent the problem. This will lead people into performing math when, in fact, none is required; Joey bought only a bag of chips and the correct answer is $0.75. In response to this question, 61% of participants gave the correct answer of $0.75 and 35% of participants gave the mindless math answer of $3.50, the sum of the three items (N = 200). Importantly, the difficulty of the actual problem (i.e., respond with a single number) is unchanged; and when people are incorrect, they are not simply being lazy: Performing the mindless math of adding the three numbers to reach $3.50 involves cognitive effort.

The effect of math difficulty is interesting because people give *more* math-based answers when the math is harder to compute. Though $3.50 is harder to compute than $6.00, it is a more common response. This is in contrast to literature on fluency which suggests that i) answers will be judged as truer if they are reached fluently (e.g., Alter & Oppenheimer, 2009), and ii) that disfluency may lead to deeper processing and greater accuracy. For example, Keysar et al. (2012) find that decision biases are reduced when people operate in a foreign language because they have to think more carefully.[3] Both of these findings suggest that the harder numbers — which are more disfluent — should activate analytical thinking and increase accuracy, rather than leading to more mindless math.

Is this effect limited to math? Let's consider a verbal formulation of the 'chips problem': *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?* Participants fluently form the word "cat", and this is reflected in their responses. 41% answered correctly with "c", but 34% of participants answered incorrectly with "cat" (N = 203).[4] When we make the tempting assembly of letters less fluent (by yielding a non-word) — *Joey went to the store and bought a pack of chips. A*

---

[3]Alter et al. (2007) report that people perform better in the Cognitive Reflection Test (Frederick, 2005) when it is presented in disfluent fonts, but strong evidence disputes such an effect (Meyer et al., 2015).

[4]These responses were collected using an open-ended text entry box. (See Table S5 for full results; and Table S3 for the results of an earlier iteration that supported the same conclusion.) Other popular responses included "chips" (4.9%), "floor" (2.9%), or "the floor" (2.7%). If we also count these responses as accurate then accuracy was 53.2% in the "cat" condition and 57.8% in the "cta" condition (the gap was slightly attenuated).

There were also other types of wrong answer (e.g., people responding "cat" in the "cta" condition, which is unambiguously wrong as only the chips with the "C" were purchased).

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? — people provide more correct answers (49%) and fewer mindless, "cta" answers (3%, N = 204). In this setting people perform better when the letter combination (cta) is less fluent than when it is more fluent (cat). This is consistent with past fluency work (e.g., Alter & Oppenheimer, 2009). In contrast, when answering our problems that contain math, harder calculations lead people to undertake the unnecessary math more. This shows that something specific about the act of doing the math calculations affects people's accuracy, rather than just the answer being disfluent in general.

Across three main pre-registered studies (total N = 3,193) and eight supplementary studies (N = 4,485), we investigated the relationship between numeric demands and individuals' ability to represent and solve problems correctly. Study 1 showed that participants facing higher numeric demands were more likely to respond with mindless math answers and less likely to answer the problems correctly. Study 2 replicated the results of Study 1 using a modified paradigm to test the difficulty effect's robustness, and collected self-reports of participants' decision processes to help elucidate the mechanism. Study 3 tested the effect's robustness across a wider range of numeric demands, and compared the effect of numeric demands on accuracy between the problems discussed so far where the calculation is conflicting with the right approach, and problems where math is the right approach.

2 Study 1

In Study 1, we manipulated the numeric demands of problems that contained tempting math that was unnecessary to reach the correct answer, and measured the rate of correct responses and the rate of incorrect responses that resulted from completing the tempting math (mindless math answers). In doing so, we hoped to investigate how changing the difficulty of tempting, but unnecessary math affects people's reasoning processes.

2.1 Methods

Participants. We initially recruited 1,920 participants from the survey platform Prolific Academic. 126 participants were excluded prior to random assignment for failing a comprehension check, leaving us with a final sample of 1,794 participants (M_{age} = 34.9y, 48.5% women, 2.1% non-binary), who completed the study for $0.18. Our sample size and analyses were pre-registered (https://osf.io/wuanj/.)

Procedure. Participants answered one of three possible questions (chips, pens, paper) with one of two levels of numeric demands (easier, harder) in a 3 x 2 full-factorial between-subjects design. The problems are listed in full in Table 1. Each problem contained tempting but unnecessary math that could lead to forming erroneous numeric problem

representations. The chips problem tempted participants to do addition, the pens tempted subtraction, and the paper tempted division. After answering their problem, participants provided demographic information (age, gender, education, race).

TABLE 1: Proportion of different types of responses to each problem.[a]

| Problem | N | Proportion of responses... | | |
		Correct	Mindless	Neither
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)	300	0.73	0.23	0.04
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)	298	0.58	0.37	0.05
Billy has 4 pens. He gives me 2 pencils. How many pens does he have left?	301	0.90	0.10	0.00
Billy has 28 pens. He gives me 9 pencils. How many pens does he have left?	300	0.72	0.27	0.01
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?[b]	301	0.92	0.07	0.01
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?	294	0.72	0.24	0.04

[a] The exact wording of the chips problem varies slightly across studies. The minor changes did not affect our conclusions in any instance.

[b] The third item differs from the first two in that the *correct* answer requires some math ($10 / 5 = 2$; $288 / 12 = 24$), but does not require the additional step that people perform, erroneously responding with the number of *crates* needed.

For each question, numeric responses were coded for two dependent variables: whether or not the answer was an instance of mindless math (MM) in which participants gave the incorrect mathematical answer and whether or not the answer was correct. Although these two dependent variables are related, they are not identical because it is possible to give an incorrect answer that does not match the MM response (e.g., saying that the volume of a 3' x 3' x 3' hole is 20 cubic feet). The joint use of both dependent variables allows us to distinguish between noise and MM responses. We pre-registered the MM responses for each problem.

2.2   Results

As shown in Table 1, making the numeric elements more complex increased the rate of MM responses. These effects were significant for each problem and overall (all ps < .001 using Fisher's exact test). In these problems, harder math difficulty led to a a higher rate of mindless math responses. Further, the incidence of mindless math answers was high overall (21%), and the MM responses made up the majority of errors (91%). There was a corresponding decrease to the rate of correct answers, both for each item and overall (all ps < .001).

2.3   Discussion

In Study 1, we presented three items where increasing the numeric demands of the numerical content led to participants responding with more mindless math answers, and so less accurately. We have replicated this result with these items many times. We also ran two studies — Study S3 and Study S4 — that tested whether other items exhibited this property. These studies tested a total of 9 items, of which 7 displayed the pattern whereby mindless math was more common when the numeric demands were higher, and thus accuracy was lower. Three of these items displayed statistically significant effects of the impact of numeric demands on both correct and mindless math responding. In short, we found the mindless math effect across some, but not all, of our different problems. In these supplementary studies, we also found that participants who scored higher on the Cognitive Reflection Test (Frederick, 2005) and Berlin Numeracy Test (Cokely et al., 2012) were more accurate and provided fewer MM responses, but neither of these scales interacted with our manipulation of the numeric demands of the problems.

Subjects performed unnecessary mathematical operations *more often* when those operations were more difficult. In the next study, we tried to understand whether participants who provided correct answers were able to avoid performing the mindless math completely or whether they computed but rejected it. Second we investigated whether greater scrutiny of question wording arose because problems in the easy condition were "too easy", leading respondents to suspect a trick.

3   Study 2

Study 1 suggests that problems whose solutions require no math are nevertheless solved at lower rates when the tempting, but unnecessary math is more difficult. This is because people do this math and return more MM answers. Yet it is less clear how people arrive at correct answers. It could be the case that when the math is easier, people do the math, and then have time left over to scrutinize their answers. This extra time could lead people to realize that the sum is incorrect, and revisit their representation to reach a correct answer. Alternatively, when the math is easier, people may feel less pressure to start quickly, and

spend more time representing the problem. We compare the evidence for each of these two explanations in Study 2.

We also wanted to test whether people perceived a trick from the experimenter at a higher rate when the math was easier. It could be the case that when the math is easy, people do not believe that someone would ask them to do it. This could lead people to perceive that the experimenter is trying to trick them, which could lead them to reread the question and reach the correct problem representation. To test this, we asked respondents to report whether they perceived a trick and whether they re-read the question.

Finally, to address the possibility that the observed mindless math effect was an artifact of the specific numbers we used — either because these numbers specifically inflated the temptation of the math, or because respondents inferred from their presence a specific intention from the experimenter — we designed a study where participants answered a single study item in which they selected all of the numbers themselves.

3.1   Methods

Participants.   Participants (N = 602) were recruited from Amazon's Mechanical Turk platform (registration: https://osf.io/y35w2). A further one-hundred fifty-seven participants were excluded prior to random assignment due to failing our comprehension check. We collected no data from these participants. The final sample of 602 respondents (Mean age = 38.2) contained 254 females and 348 males. Participants were paid $0.31 for participation.

Procedure.   All participants were given a version of the chips problem used in Study 1. As in Study 1, we manipulated the levels of the numeric demands. In this case, the manipulations were created by having subjects choose their own numbers from lists we manipulated: we attempted to induce participants to view the numbers as *their* choice, rather than the *experimenter's* choice.[5]

After providing consent, participants were asked to select three numbers from three identical dropdown menus. Half the subjects were randomly assigned to a dropdown list containing values ending in round numbers (0.5 to 5 in increments of 0.5). The other half saw a harder list containing values ending in less round numbers (e.g., 1.05, 3.80, etc.); see Table 2. After subjects chose three numbers, they read the chips problem with these values instantiated as the prices of the three items in a chips problem: *Imagine Joey is going to the store to buy a pack of chips. A bottle of water costs [N1], a pack of chips costs [N2] and a pack of gum costs [N3]. How much does he spend in total? (in dollars)*

After answering, participants were presented with a menu of possible strategies they might have used in answering the question (see Table 3). Participants were asked to report which process best described how they reached their answer and how many times they read

---

[5]Asking participants to choose their own numbers could *encourage* them to add them, as they may take ownership of the numbers and wish to do something with them. Yet this can't be the sole explanation of mindless math, as people still perform operations on the numbers in Study 1 when they do not choose them.

TABLE 2: Drop down lists of numbers to instantiate into chips problem. (Subjects did not see the word "easier" or "harder".)

Easier	Harder
0.50	0.30
1.00	0.75
1.50	1.05
2.00	1.70
2.50	2.15
3.00	2.60
3.50	3.35
4.00	3.80
4.50	4.40
5.00	4.85

the first sentence of the chips problem. If they answered that they read the first sentence more than once, they were then asked to select all options that applied to why they re-read the first sentence. Finally, participants were asked how difficult the problem was relative to their expectations, coded from -3 (a lot easier than expected) to 3 (a lot harder than expected).

3.2   Results

We again found that participants in the harder numeric demands condition gave the MM answer at a higher rate (68% versus 53%, Fisher's exact $p < .001$) and the correct answer at a lower rate (17% versus 37%, $p < .001$). This shows that the effect is robust when it is less likely that participants infer a specific intention from the experimenter about the presented numbers. Respondents are tempted to do the unnecessary math, even when it is of their own construction.

In comparing participants' self-reports of their processes, we found that participants in the easier condition were significantly less likely to pursue the wrong strategy of doing the math (Proportion$_{easier}$ = 0.53 vs. Proportion$_{harder}$ = 0.69, $p < .001$), and more likely to pursue the normative strategy of doing no math (Proportion$_{easier}$ = 0.34 vs. Proportion$_{harder}$ = 0.18, $p < .001$). There were no significant differences in the rate at which participants reported that they completed the sum but did not incorporate it into their answer (Proportion$_{easier}$ = 0.13 vs. Proportion$_{harder}$ = 0.13, $p = 1.00$). This evidence suggests that participants in the easier numeric demands condition were more likely to represent the problem correctly the first time around, and thus avoid the need to do any math.

Accurate participants overwhelmingly reported that they did not complete the math (89% in the easier numeric demands condition and 88% in the harder numeric demands

TABLE 3: Process questions from Study 2.

Problem solving strategy	Please select the option that best describes how you reached your answer. Please be honest, it will not affect your payment or approval in any way.[a] • I did not add any numbers together. • I added numbers together, but did not use the sum in my final answer. • I added numbers together, and I used the sum in my final answer.
Reading first sentence	How many times did you read the first sentence of the problem? ————
Reason for re-reading	Why did you reread it? Please check all that apply. • I could not remember the question. • The problem looked easier than I expected. • I thought the experimenter might be tricking me.
Experienced difficulty	This question was: • A lot easier than expected. • Somewhat easier than expected. • Slightly easier than expected. • About as difficult as I expected. • Slightly harder than expected. • Somewhat harder than expected. • A lot harder than expected.

[a] It is possible that people interpreted this question as only asking about operations they performed to directly reach the final answer, and not earlier steps that were conducted prior to the processes that produced that final answer.

condition). A smaller percentage of accurate participants indicated that they completed math but did not use it (10% in each condition).[6] On the other hand, most MM participants (93% in the easier condition and 92% in the harder condition) reported that they completed math and used it (as we would expect). Overall, this evidence suggests the different numeric demand levels affected how participants represented and processed the problem *prior* to completing math, rather than through allowing them additional time to go back and check

---

[6] Of the participants who gave neither correct nor MM responses, a high rate of respondents (50%) indicated that they did math but did not use it. We cannot identify any possible alternative interpretation of the question that led to these results. These participants appear to just be noisy.

their answers.

One factor that could encourage people to scrutinize their representations and avoid doing math would be suspicion that an especially easy sum is not the correct operation. Though slightly more participants indicated they re-read the stem in the easier condition relative to the harder condition (48.5% vs. 45.8%), there was no significant difference in participants' reports of the number of times they read the first sentence of the problem across numeric demands conditions (M_{easier} = 1.58 vs. M_{harder} = 1.62, $t(585)$ = –0.60 p = .500), suggesting this is not the explanation.

To further probe this, we used the numeric demands and the number of times the participant read the item as predictors of whether a participant reached a correct answer. Both numeric demands (b = –1.18, p < .001) and times read (b = 0.591, p < .001) achieved statistical significance: the hard-easy effect remained even after controlling for the number of times a respondent read the problem. In fact, when including harder numeric demands, times read, and their two-way interaction in the model, there was a significant interaction (b = –0.655, p = .007), suggesting that harder numeric demands reduced the benefit of re-reading the problem.

Among the subset of participants who did re-read the problem (n = 284), there were no differences in the reasons they (non-exclusively) reported for why they chose to re-read the problem across numeric demands condition. Specifically, subjects reported the following justifications for why they re-read the question stem at the same rates across conditions; *I could not remember the question* (Proportion$_{easier}$ = 0.12 vs. Proportion$_{harder}$ = 0.14, p = .726), *The problem looked easier than I expected* (Proportion$_{easier}$ = 0.46 vs. Proportion$_{harder}$ = 0.51, p = .406), and *I thought the experimenter might be tricking me* (Proportion$_{easier}$ = 0.73 vs. Proportion$_{harder}$ = 0.65, p = *.160*).[7] It does not appear that the hard-easy effect is explained by suspicion of a trick in the easier numbers condition.[8]

Participants did report that the problems in the easier numeric demands condition were significantly easier (relative to their expectations) than participants in the harder numeric demands condition (M_{easier} = –1.27 vs. M_{harder} = –0.82, p = .002). This result suggests that participants' perceptions of the difficulty of the problem may have contributed to whether they did the math or not, but not due to suspicion of the experimenter leading them to re-read the question.

---

[7]These categories are not wholly exclusive — someone could suspect a trick *because* the question was easier than expected — which is why participants were able to select more than one possible reason for re-reading.

[8]Note that when collapsing across numeric demands conditions, correct respondents reported re-reading the question stem more frequently than mindless respondents ($M_{correct}$ = 1.91, M_{MM} = 1.41, $t(388)$ = 8, p < .001), in part because they suspected a trick (Proportion$_{correct}$ = 0.62 vs. Proportion$_{MM}$ = 0.22, p < .001) and in part because the problem was easier than expected (Proportion$_{correct}$ = 0.34 vs. Proportion$_{MM}$ = 0.17, p < .001).

3.3   Discussion

In Study 2 we again found a higher rate of mindless math when numeric demands were harder, replicating Study 1's results in a new design that allowed participants to select their own numbers. We also ruled out several possible explanations for the hard-easy effect: Participants in the easier numeric demands condition did not indicate greater suspicion of experimenter tricks that led to them re-reading the question, though it did appear that the *returns* to re-reading the stem may be higher when the numeric demands were easier. In other words, though participants suspecting a trick from the experimenter does not appear to be a compelling explanation of the observed effect, harder numeric demands appear to impede people's ability to correctly solve the problem even when re-reading the stem.

Although self-reported decision processes are subject to potential response biases (Austin & Delaney, 1998; Ericsson & Simon, 1998), Study 2 suggests that participants who are able to reach *correct* answers do not complete any math, but do re-read and scrutinize the problem more thoroughly. People got the answer right or wrong for the same reasons in both the easier and harder conditions, and further work is required to understand why the reasoning processes that led to reaching correct answers were differentially utilized across the numeric demands conditions.

4   Study 3

In Study 1 we found evidence that, in our problems, increasing the numeric demands of tempting but unnecessary math caused participants to give more mindless math answers and fewer correct ones. In Study 2, we replicated this result with a new design, and found that the difference in participants' performance (in terms of the rate of correct and MM responses) did not seem to be explained by participants re-reading the question or suspecting experimenter tricks at a different rate.

In Study 3, we sought to test the relationship between numeric demands and accuracy over a wider range of values and with a behavioral measure of math difficulty. Participants either responded to a conflict version of the chips problem (i.e., the original) or a non-conflict version (where the correct answer did involve a sum). This design allowed us to analyze one set of participants' accuracy in the conflict version, and another set of participants' response times in the non-conflict version (i.e., how long did the tempting math take to complete). The participants' response times in the non-conflict version served as a measure of the difficulty of the math: accuracy was generally very high in these conditions and so this variable captured how long on average it took to do the math correctly. As a result, a negative correlation between the response times taken to solve the non-conflict versions and the rate of correct responding in the conflict versions would show that as the math got more difficult, people were less able to reach correct answers in the conflict versions where no math was required. We predicted that this would be the case, consistent with the difficulty effect observed between conditions in Studies 1 and 2.

4.1   Methods

Participants.   Participants (N = 1005) were recruited from Amazon's Mechanical Turk platform (registration: https://osf.io/mysut). 208 participants were prevented from completing the study for failing our comprehension check (prior to random assignment). This left us with 797 complete observations.[9] The sample (Mean age = 36.3) was 44.5% male. Participants were paid $0.20 for participation.

Procedure.   After providing consent, participants answered a version of the chips problem used in Studies 1 and 2. We manipulated two independent variables in an 8 x 2 factorial between-subjects design. The independent variables were the numeric demands of the math in the study item and whether the item was the conflict version of the problem in which the math is tempting but unnecessary (as in Studies 1 and 2) or a non-conflict version in which the tempting math is the correct approach. We use the term *conflict* to convey whether an intuitive response is cued that is different from the normative response (De Neys, 2012; Evans, 2010).[10]

Table 4a shows the eight levels of math difficulty: The first three levels used only whole dollars (e.g., $2.00), the next two levels used dollar amounts that varied in the first decimal place (e.g., $1.30), and the next three levels used dollar amounts that varied in first and second decimal places (e.g., $3.25). Table 4b shows two versions of a problem with the same level of numeric demands. On the left is the conflict version (as used in Studies 1 and 2) in which people may erroneously respond as if Joey is buying all three items. On the right is the non-conflict version in which the representation of the problem as a sum is the correct one.

As we continued to increase the difficulty of the numeric demands conditions, people's attempts to do math included more errors. Thus, here we focused on the rate of correct responses (which imply an absence of performing mindless math) rather than mindless math responses (which demonstrate its presence). Further, there was no MM response to the non-conflict version, as the correct response was computing the sum. As well as accuracy, we also measured participants' response times.

4.2   Results

In Table 4a, we see that non-conflict items that were designed to be harder did in fact take more time on average to answer (as shown by the geometric mean response time increasing

---

[9]We had initially pre-registered our intention to remove a further 29 participants for responding with the incorrect sum in the non-conflict version of the 'chips problem' to use the condition as a measure of how long it took to successfully complete the sum, but it was pointed out to us that this introduced a confound. Our statistical analyses were robust in both cases, as is noted later.

[10]The conflict is between the participant's intuitively cued response and the correct response; participants may not *feel* a sense of conflict.

TABLE 4: A: The 8 levels of numeric demands for Study 3.

Numeric Demands	Price of Water	Price of Chips	Price of Gum	Geometric Mean RT (non-conflict)	Proportion Correct (non-conflict)	Proportion Correct (conflict)	Difficulty
1	3.00	1.00	2.00	20.2	0.98	0.66	Easy
2	4.00	2.00	5.00	21.4	0.98	0.60	
3	5.00	1.00	7.00	21.3	0.94	0.58	
4	3.10	1.30	2.00	30.1	0.94	0.44	Medium
5	3.30	1.40	2.20	30.8	0.86	0.38	
6	3.25	1.50	2.65	39.1	0.88	0.47	Hard
7	3.15	1.45	2.65	38.5	0.96	0.37	
8	3.15	2.34	1.85	43.1	0.88	0.41	

B: A non-conflict and conflict version of the 'chips problem'.

Conflict	Non-conflict
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 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)**

across the range of numeric demands). The non-conflict versions were also answered less correctly across the range of numeric demands (98% to 88%).

On conflict items, accuracy decreased to a greater degree as the tempting math got harder – from 66% for the easiest to 41% for the hardest. Using logistic regression to predict correct answers with the level of numeric demands (1–8)[11], whether the problem was the conflict version, and their two way interaction revealed that the slope was significantly steeper for the conflict version ($b = -0.026$, $p = .030$). In other words, harder numeric demands had a larger negative effect on accuracy for the conflict items than the non-conflict items (25 percentage points vs. 10).

A negative correlation between geometric mean response time in the non-conflict condition and accuracy in the conflict condition ($r(6) = -0.86$; 95% CI [-1.00, -0.51]) supports the central claim that participants are more likely to engage in mindless math (and less likely to be accurate) when the math is harder.

---

[11]We used the pre-specified task difficulty as our measure of difficulty here because this analysis occurred at the level of the individual rather than the level of the problem. For the problem-level analyses, we considered the geometric mean response time for each difficulty level to be the best measure of *problem* difficulty.

In a pre-registered one-tailed test of the Pearson's product-moment correlation coefficient, this was statistically significant ($p = .003$). This relationship was robust if we included only correct answers when we calculated the average log time taken in the non-conflict condition ($r = -0.86$, $p = .003$), and when using our *a priori* ranking of the difficulty of numeric demands (1–8), rather than a response time based measure ($r = -0.87$, $p = .002$).

4.3   Discussion

Study 3 complemented earlier findings by showing that a behavioral measure of math difficulty (average response time to reach a correct answer in a non-conflict version) was negatively associated with accuracy in conflict versions of problems over a range of numeric demands. For the non-conflict items, the effect was smaller. In the discussion, we offer a possible interpretation of *why* harder numeric demands might reduce accuracy and increase the rate of mindless math.

5   General Discussion

Across three pre-registered studies (total N = 3,193), we studied the relationship between the difficulty of tempting math, problem representation, and problem solving. We found that higher numeric demands led to people performing *mindless math* at a higher rate, giving answers based on unnecessary and inappropriate calculations. In self-reports, participants who reached correct answers generally said that they did not perform any math. This result suggests that harder numeric demands lead to reduced accuracy because they induce people to do math, rather than because they leave less time to scrutinize the math they've done.

There are several possible reasons why harder numeric demands might induce someone to do math, or otherwise be less skeptical that math is the right thing to do. To identify some of these, in a supplementary study we collected item-level ratings of some properties of the problems from Study 3. Essentially we looked for variables that covaried with math difficulty that could mediate the relationship between difficulty and likelihood of performing mindless math. We found two variables that were strongly associated with the difficulty of the math: the anticipated cognitive effort associated with the math ($r = 0.93$), and the belief that doing the math would be more impressive ($r = 0.86$). Each of these factors was associated with a reduced likelihood that participants solved the problem correctly and a greater likelihood of performing mindless math. These are two possible pathways that might explain the relationship between math difficulty and people giving mindless math responses — the former by leading to reallocation of attention away from problem representation, and the latter by increasing the social returns to performing the math. For further details of these analyses and our consideration of alternative mechanisms, refer to the supplement.

At a high level, research in judgment and decision making argues that different processes guide intuition (System 1) and deliberation (System 2), where deliberation is often

needed both to detect and correct errors that arise from intuition (Kahneman, 2011; Sloman, 1996). The current research represents an interesting example where one System 2 process (performing math operations) distracts attention from a second System 2 process (monitoring for errors in problem representation). The errors come from System 2 being actively engaged in the wrong deliberative task, rather than not being engaged at all. People are not being lazy, but misguided. Kahneman (2000) notes that for System 2 to be able to correct a System 1 reasoning error, the cues that evoke applying a rule must be present, and the rule must be known. The act of doing math appears to suppress rather than elucidate the fact it is the wrong approach, which makes overcoming this error especially difficult. In this sense, working in an incorrect problem representation is like attempting to solve a problem without the necessary inferential rule — without this contextualizing information, trying harder may not be very helpful (Lawson et al., 2020).

Given that the benefits of engaging in deliberation are frequently contingent — either on possessing a rule (as in Lawson et al. (2020)) or forming a good problem representation — attempts to debias reasoning errors may benefit from guiding people to a *process* as well as a decision speed (Larrick & Lawson, 2021). For example, consider Heath & Heath's (2013) WRAP framework that compels decision makers to: Widen your options, Reality-test your assumptions, Attain distance before deciding, and Prepare to be wrong. Such a framework provides people with guidance for what to do with additional time spent deliberating, which may otherwise be wasted. Similarly, Ralph Keeney's (1996) "value-focused thinking" aims to provide guidance for how to create better alternatives for decision problems, which helps to guide how to productively use time spent deliberating. In the case of people performing mindless math, a debiasing intervention might go past telling people to deliberate — which may only lead to people more diligently performing unnecessary math — and instead ask decision makers to list the elements of the problem and draw a diagram that establishes how they relate to each other, before engaging in any computation.

People's propensity to jump into tasks has widespread implications in organizational contexts. In strategic problem-solving, researchers have highlighted the presence of a "plunging-in bias", where people start solving problems before understanding them (Bhardwaj, Crocker, Sims & Wang, 2018). The present research shows empirically how this tendency can be harmful, as failing to represent a problem correctly can lead to mindless math. In our problems, people would benefit from spending more time representing the problem up front, rather than jumping into the math. The general importance of this is summarized in a saying from the U.S. Navy Seals: "slow is smooth, smooth is fast".

A key question that follows from highlighting the pitfalls of getting underway too quickly is, how can we incentivize people taking their time represent problems? One reason why people might jump into doing math is that the output of such operations is tangible and demonstrates effort, despite being wrong. The act of representing a problem is less tangible and therefore harder to hold up as demonstrable proof of work. Yet good processes — which sometimes can seem intangible — are essential to organizational success. In

uncertain environments, organizations should want their members to fail quickly and often, and constantly learn from their mistakes. This requires setting out in directions that provide opportunities to learn, and consistently re-evaluating one's approach. The only way to achieve this is through a good process — rather than over-emphasizing specific outcomes. To avoid the errors of diving into problems too quickly and misrepresenting them, greater emphasis must be placed on these processes.

A notable implication of hard-easy effect observed with mindless math is that this propensity to jump into doing something will be particularly pronounced in situations when the task is difficult, which presumably warrant even *more* time spent representing the problem, rather than less. Organizations could seek to safeguard against this tendency by putting embargos on project tasks (times they cannot be completed before) rather than deadlines. Setting an external time to *start* implementing could ensure that representing a problem is given sufficient attention, thus minimizing the likelihood of acting *mindlessly*.

5.1 Limitations

One limitation of the present work is the specificity of the problems — in the discussion of Study 1 we referenced some additional cases of the hard-easy mindless math effect, but this effect will not be present for most problems. As a result, our conclusions are limited to cases where a task invites people to dive into doing familiar operations, but these operations are not helpful for solving the task.

Second, though Study 3 showed that over some ranges increasing the numeric demands of tempting calculations increases the rate of mindless math in our problems, at some point this effect inverts, as debilitatingly difficult computations may also serve as a cue to reinspect one's representation of a problem. Thus, increasing the numeric demands of problems may increase or decrease respondents' use of mathematical strategies. In Table 5 we see this – moving from the easiest problem in the first row to the harder one in the second, accuracy decreases, but for the most difficult computation (the last row), accuracy is higher relative to the *moderately* difficult version (row two).

Another limitation of the present research is that it contains data only from online samples facing low stakes. For example, though the MTurk population is highly attentive and engaged[12] (Hauser & Schwarz, 2016), it is not representative of the broader US population (Huff & Tingley, 2015), and hence our results may look different in a different sample. Yet it is unclear whether higher stakes would improve accuracy or not. Larger financial incentives, for example, might lead to broader allocation of attention to tasks such as representation, but could simply induce greater application of effort to the wrong strategy, which would not improve accuracy (Lawson et al., 2020; Enke et al., 2021). It would also be beneficial to expand the field of problems studied to those that involve more complex sequences of

---

[12]Perhaps unrepresentatively so.

TABLE 5: The rate of correct and mindless answers across an even wider range of numeric demands.

Joey went to the store and bought a pack of chips. A bottle of water costs N1, a pack of chips costs N2 and a pack of gum costs N3. How much did he spend in total? (in dollars)			Percentage of answers...		
Price of water (N1)	Price of chips (N2)	Price of gum (N3)	Correct	Mindless	Neither
$3.00	$1.00	$2.00	0.58	0.41	0.03
$1.05	$0.75	$1.70	0.46	0.51	0.03
$0.69	$0.93	$3.43	0.52	0.39	0.08
$3.13	$2.43	$7.26	0.63	0.29	0.07

computations. We speculate that with greater complexity, the temptation to engage in mindless math may be stronger, and continue to impede identifying the correct path.

5.2 Conclusion

When faced with difficult tasks, people are often eager to start doing *something*. But this eagerness may preclude them from correctly representing the problems that they are so eager to start solving. When we ask people to "take their time", we have to be more specific about *when* they should take their time. The execution of operations gives the illusion of progress, but if the problem is represented incorrectly, it remains just that, an illusion.

References

Adamson, R. E. (1952). Functional fixedness as related to problem solving: A repetition of three experiments. *Journal of Experimental Psychology*, *44*(4), 288–291. https://doi.org/10.1037/h0062487

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review, 13*(3), 219–235.

Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General, 136*(4), 569–576. https://doi.org/10.1037/0096-3445.136.4.569

Ash, I. K., & Wiley, J. (2006). The nature of restructuring in insight: An individual-differences approach. *Psychonomic Bulletin & Review*, *13*(1), 66–73. https://doi.org/10.3758/BF03193814

Austin, J., Delaney, P. F. (1998). Protocol analysis as a tool forbehavior analysis. *The Analysis of Verbal Behavior, 15*, 41–56. https://doi.org/10.1007/BF03392922.

Bar-Hillel, M., Noah, T., & Frederick, S. (2018). Learning psychology from riddles: The case of stumpers. *Judgment and Decision Making, 13*, 112–122.

Bar-Hillel, M., Noah, T., & Frederick, S. (2019). Solving stumpers, CRT and CRAT: Are the abilities related? *Judgment and Decision Making, 14*(5), 620–623.

Bassok, M. (1996). Using Content to Interpret Structure: Effects on Analogical Transfer. *Current Directions in Psychological Science*, *5*(2), 54–58. https://doi.org/10.1111/1467-8721.ep10772723

Bhardwaj, G., Crocker, A., Sims, J., & Wang, R. D. (2018). Alleviating the plunging-in bias: Elevating strategic problem-solving. *Academy of Management Learning & Education*, *17*(3), 279–301. https://doi.org/10.5465/amle.2017.0168

Chein, J. M., Weisberg, R. W., Streeter, N. L., & Kwok, S. (2010). Working memory and insight in the nine-dot problem. *Memory & Cognition*, *38*(7), 883–892. https://doi.org/10.3758/MC.38.7.883

Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. (2012). Measuring risk literacy: The Berlin Numeracy Test. *Judgment and Decision Making*, *7*(1), 25–47.

Cooper, G., & Sweller, J. (1987). Effects of schema acquisition and rule automation on mathematical problem-solving transfer. *Journal of Educational Psychology*, *79*(4), 347–362. https://doi.org/10.1037/0022-0663.79.4.347

De Neys, W. (2012). Bias and Conflict: A Case for Logical Intuitions. *Perspectives on Psychological Science*, *7*(1), 28–38. https://doi.org/10.1177/1745691611429354

Duncker, K. (1945). On problem-solving. *Psychological Monographs, 58* Whole No. 270. http://dx.doi.org/10.1037/h0093599.

Enke, B., Gneezy, U., Hall, B., Martin, D., Nelidov, V., Offerman, T., & van de Ven, J. (2021). Cognitive biases: Mistakes or missing stakes? *The Review of Economics and Statistics*, 1–45. https://doi.org/10.1162/rest_a_01093.

Ericsson, K. A., & Simon, H. A. (1998). How to study thinking in everyday life: Contrasting think-aloud protocols with descriptions and explanations of thinking. *Mind, Culture, and Activity*, *5*(3), 178–186. https://doi.org/10.1207/s15327884mca0503_3

Evans, J. S. B. T. (2010). Intuition and Reasoning: A Dual-Process Perspective. *Psychological Inquiry*, *21*(4), 313–326. https://doi.org/10.1080/1047840X.2010.521057

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*(4), 25–42. https://doi.org/10.1257/089533005775196732.

Greeno, J. G., & Simon, H. A. (1988). Problem solving and reasoning. In *Stevens' handbook of experimental psychology: Perception and motivation; Learning and cognition, Vols. 1-2, 2nd ed* (pp. 589–672). John Wiley & Sons.

Hauser, D. J., & Schwarz, N. (2016). Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods*, *48*(1), 400–407. https://doi.org/10.3758/s13428-015-0578-z

Heath, C., & Heath, D. (2013). *Decisive: How to make better choices in life and work*. New York: Random House.

Huff, C., & Tingley, D. (2015). "Who are these people?" Evaluating the demographic characteristics and political preferences of MTurk survey respondents. *Research & Politics*, *2*(3), 2053168015604648. https://doi.org/10.1177/2053168015604648

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Harvard University Press.

Kahneman, D. (2000). A psychological point of view: Violations of rational rules as a diagnostic of mental processes. Behavioral and Brain Sciences, 23(5), 681–683. https://doi.org/10.1017/S0140525X00403432 .

Kahneman, D. (2011). *Thinking Fast and Slow*. Farrar, Straus and Giroux.

Keeney, R.L. (1996), Value-focused thinking: Identifying decision opportunities and creating alternatives, *European Journal of Operational Research, 92*(3), 537–549.

Keysar, B., Hayakawa, S. L., & An, S. G. (2012). The Foreign-Language Effect: Thinking in a Foreign Tongue Reduces Decision Biases. *Psychological Science, 23*(6), 661–668. https://doi.org/10.1177/0956797611432178.

Larrick, R. P., & Lawson, M. A. (2021, August 31). Judgment and Decision-Making Processes. Oxford Research Encyclopedia of Psychology. https://doi.org/10.1093/acrefore/9780190236557.013.867

Lawson, M. A., Larrick, R. P., & Soll, J. B. (2020). Comparing fast thinking and slow thinking: The relative benefits of interventions, individual differences, and inferential rules. *Judgment and Decision Making*, *15*(5), 660–684.

Luchins, A. S., & Luchins, E. H. (1950). New experimental attempts at preventing mechanization in problem solving. *Journal of General Psychology*, *43*, 279–297.

McNamara, T. P. (1994). Chapter 3 — Knowledge representation. In R. J. Sternberg (Ed.), *Handbook of Perception and Cognition: Thinking and problem solving* (Vol. 2, pp. 81–117). Academic Press. https://doi.org/10.1016/B978-0-08-057299-4.50009-8.

Meyer, A., & Frederick, S. (2022). *Forming and revising intuitions* (SSRN Scholarly Paper No. 4039414). https://doi.org/10.2139/ssrn.4039414.

Meyer, A., Frederick, S., Burnham, T. C., Guevara Pinto, J. D., Boyer, T. W., Ball, L. J., Pennycook, G., Ackerman, R., Thompson, V. A., & Schuldt, J. P. (2015). Disfluent fonts don't help people solve math problems. *Journal of Experimental Psychology: General, 144*(2), e16–e30. https://doi.org/10.1037/xge0000049

Newell, A., & Simon, H. *Human problem solving*. Englewood Cliffs, N.J.: Prentice-Hall, 1972.

Reimann, P., & Chi, M. T. H. (1989). Human Expertise. In K. J. Gilhooly (Ed.), *Human and Machine Problem Solving* (pp. 161–191). Springer US. https://doi.org/10.1007/978-1-4684-8015-3_7

Sheridan, H., & Reingold, E. M. (2013). The mechanisms and boundary conditions of the Einstellung effect in chess: Evidence from eye movements. *PLOS ONE*, *8*(10), e75796. https://doi.org/10.1371/journal.pone.0075796

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*(1), 3–22.

Thomson, K. S., & Oppenheimer, D. M. (2016). Investigating an alternate form of the cognitive reflection test. *Judgment and Decision Making*, *11*(1), 99–113.

Wiley, J., & Jarosz, A. F. (2012). Working memory capacity, attentional focus, and problem solving. *Current Directions in Psychological Science*, *21*(4), 258–262. https://doi.org/10.1177/0963721412447622.