Algebraic reasoning and bat-and-ball problem variants: Solving isomorphic algebra first facilitates problem solving later

Jerome D. Hoover¹ · Alice F. Healy¹

Published online: 14 February 2017
© Psychonomic Society, Inc. 2017

Abstract The classic bat-and-ball problem is used widely to measure biased and correct reasoning in decision-making. University students overwhelmingly tend to provide the biased answer to this problem. To what extent might reasoners be led to modify their judgement, and, more specifically, is it possible to facilitate problem solution by prompting participants to consider the problem from an algebraic perspective? One hundred ninety-seven participants were recruited to investigate the effect of algebraic cueing as a debiasing strategy on variants of the bat-and-ball problem. Participants who were cued to consider the problem algebraically were significantly more likely to answer correctly relative to control participants. Most of this cueing effect was confined to a condition that required participants to solve isomorphic algebra equations corresponding to the structure of bat-and-ball question types. On a subsequent critical question with differing item and dollar amounts presented without a cue, participants were able to generalize the learned information to significantly reduce overall bias. Math anxiety was also found to be significantly related to bat-and-ball problem accuracy. These results suggest that, under specific conditions, algebraic reasoning is an effective debiasing strategy on bat-and-ball problem variants,

and provide the first documented evidence for the influence of math anxiety on Cognitive Reflection Test performance.

Keywords Decision-making · Judgment · Reasoning · Math anxiety · Algebraic reasoning · Debiasing · Cognitive Reflection Test

The classic bat-and-ball problem, which is part of the Cognitive Reflection Test (CRT) (Frederick, 2005), has been used widely to demonstrate biased and correct reasoning in decision-making. It is presented as follows:

A bat and ball together cost \$1.10. The bat costs \$1.00 more than the ball.
How much does the ball cost?

Five cents for the ball is the correct solution. However, participants tend to respond quickly, “10 cents,” without considering all the relevant information provided in the problem description. The relevant wording that necessitates further consideration is the “more than” phrase (e.g., De Neys, Rossi, & Houdé, 2013; Hathorn & Healy, 2016). Individuals who give the “10 cents” answer are incorrectly equating, “The bat costs \$1.00 more than the ball,” to “The bat costs \$1.00,” for which 10 cents would indeed be the correct solution.

Some have proposed that biased reasoners respond in such a manner because they are cognitive misers (Kahneman, 2011; Kahneman & Frederick, 2002; Tversky & Kahneman, 1974). Motivated by a predisposition not to expend more internal resources than necessary, reasoners solve a simpler, more intuitive problem in lieu of the actual, more challenging problem—a phenomenon referred to as *attribute substitution* (Kahneman, 2011; Kahneman & Frederick, 2002). Attribute substitution is thought to be the process by which a number of

Electronic supplementary material The online version of this article (doi:10.3758/s13423-017-1241-8) contains supplementary material, which is available to authorized users.

✉ Alice F. Healy
Alice.Healy@Colorado.EDU

Jerome D. Hoover
Jerome.Hoover@Colorado.EDU

¹ Department of Psychology and Neuroscience, University of Colorado Boulder, Boulder, CO 80309-0345, USA

logical fallacies and cognitive heuristics take place (Kahneman & Tversky, 1973). As such, when reasoners engage in attribute substitution in the real world, it may not be as trivial as a miscalculation on a logic problem. Instead, it may come in the form of logical fallacies and cognitive heuristics that result in stereotyping, base rate neglect, etc.

This description is often discussed in terms of a dual-process account of human judgment and decision-making (Evans, 2008; Evans & Stanovich, 2013; Kahneman 2011; Kahneman & Frederick, 2002; Stanovich & West, 2002). System 1 processing is quicker, intuitive, and less demanding on the reasoner's limited pool of cognitive resources, whereas System 2 processing is slower, responsible for the critical evaluation of information, and resource intensive. As a result, heuristics provide utility because they save time and cognitive resources, and are sufficiently accurate in many situations, particularly concerning decisions in which reasoners have significant expertise (e.g., Gigerenzer & Selten, 2002). By this account, in the bat-and-ball problem, biased answers derive from System 1, whereas correct responses derive from System 2.

However, not all theorists agree that the dual-process account is the most appropriate framework to discuss human reasoning. In particular, unconscious (System 1) processes have not been demonstrated to be useful explanatory constructs (e.g., Newell & Shanks, 2014). Dual-process theories also tend to assume that the cognitive processes that regulate either a biased or a correct response are distinct to these respective systems; however, the overlap is substantial and heavily dependent on the individual reasoners' goals in addition to their higher order cognitive abilities (Cokely & Kelley, 2009; Keren & Schul, 2009). For this reason, a single-system model may be theoretically preferable in an effort to expand predictive power (Keren & Schul, 2009).

More specific to the CRT itself, the bat-and-ball problem is not just a measure of cognitive reflection (i.e., the ability to inhibit a prepotent intuitive response and engage in further consideration), but numeracy and cognitive tendencies toward *Actively Open-minded Thinking* (AOT) are likewise determinants of CRT accuracy, at least for men (Baron, Scott, Fincher, & Metz, 2015; Campitelli & Gerrans, 2014). AOT, broadly defined, can be thought of as the disposition toward the search for alternative answers to the intuitive response, and a strategy to counteract myside bias—the tendency for participants' confidence in their initial response to increase (Baron et al., 2015).

Approximately 80% of undergraduate students exhibit biased reasoning on the bat-and-ball problem (e.g., Bourgeois-Gironde & Vanderhenst, 2009; Frederick, 2005). However, this tendency is lessened depending on education level: undergraduates tested at more prestigious academic institutions have been found to demonstrate lower proportions of biased reasoning (e.g., Frederick, 2005; Thompson et al., 2013). This difference can be accounted for in part by variability in IQ;

however, educational attainment is also relevant, implying that training might decrease bias in the bat-and-ball problem.

Previous research findings have demonstrated a number of successful debiasing strategies for the bat-and-ball problem (e.g., Alter, Oppenheimer, Epley, & Eyre, 2007; Bourgeois-Gironde & Vanderhenst, 2009; Mastrogiorgio & Petracca, 2014; Thompson et al., 2013). For example, degrading or stylizing the lettering of the font such that it was more difficult to read resulted in higher accuracy (Alter et al., 2007). However, cueing the algebraic structure of the bat-and-ball problem as a debiasing strategy has not yet been investigated. The extracted algebraic structure of the bat-and-ball problem is:

$$\begin{aligned}\text{Bat} + \text{Ball} &= \$1.10 \\ \text{Bat} &= \$1.00 + \text{Ball} \\ \text{Solve for Ball}\end{aligned}$$

By reducing the verbal information to a pair of algebraic equations, the “more than” phrase and the substitution inducing potential are absent. The only skill required at this point is basic algebraic calculation. Is it possible to facilitate problem solution by cueing participants to reason through the bat-and-ball problem from its algebraic perspective?

We explored the possible benefits of algebraic cueing across the following between-subjects experimental cueing conditions: (1) by initially presenting the relevant algebra in the context of the question (*algebra plus words*); (2) by initially cueing the relevant algebra without question context (*algebra only*); and (3) by cueing unrelated algebra (*algebra irrelevant*). We predicted that, overall, algebraically cued participants would outperform non-cued participants on a subsequent critical question presented in its standard form. However, we also expected performance to be highest among participants in the algebra plus words condition, because they were cued with the relevant mathematical equations in the context of the question itself. Accuracy on the cueing questions might also contribute to cueing effectiveness on the critical question.

Thus, we are primarily interested in determining if algebra cueing without feedback increases accuracy on a subsequent question presented in its standard form. If accuracy on the second (“critical”) question is improved by an algebraic cue, we will infer that participants were able to generalize the learned information from their cueing question to their subsequent critical question (of identical structure, but containing differing item content and dollar amounts). Therefore, accuracy on the second question was the primary measure of interest. We expected to see a decrease in biased responding on this question among algebraically cued participants.

Following De Neys et al. (2013), confidence ratings were solicited after the critical question. This measure might provide additional insight into the extent to which reasoners were

led to modify their judgment across the cueing conditions' critical question.

As additional measures, we also included math anxiety indices. Accuracy on the bat-and-ball problem depends on numeric ability in addition to analytical predisposition (see Pennycook & Ross, 2016; Sinayev & Peters, 2015). There is also evidence demonstrating the deleterious effects of anxiety in certain decision-making contexts (e.g., in gambling tasks; Miu, Heilman, & Houser, 2008; Raghunathan & Pham, 1999). However, the relationship between reasoning on bat-and-ball problem types and math anxiety specifically has yet to be considered from an empirical perspective. Therefore, two math anxiety scales were included to explore the hypothesized negative impact of math anxiety on critical question accuracy.

Methods

Participants

One hundred participants were recruited online via Amazon Mechanical Turk ($M_{Age} = 26.04$, $SD = 6.31$, range = 18–49, men = 69%), and 97 University of Colorado undergraduate psychology students ($M_{Age} = 19.38$, $SD = 1.09$, range = 18–22, men = 51%) were recruited as laboratory participants. Online participants were restricted to current college students (indicated via self-report) at least 18 years of age, so as to be comparable to the laboratory sample.

Materials and procedure

The surveys were developed in Qualtrics survey software and distributed via Amazon Mechanical Turk for the online sample. Online participants were asked to provide their own paper and writing utensils to work out the math problems. Laboratory participants were tested on the same set of surveys on iMac desktop computers. Pencils and paper were provided for laboratory participants who did not have these items.

The items and item amounts used across cueing conditions followed De Neys et al. (2013): Instead of bat/ball/\$1.10, the counterbalanced variants of the problems were pencil/eraser/\$1.10 and magazine/banana/\$2.90 (with \$2.90 replacing \$1.10 and \$2.00 replacing \$1.00). For the magazine/banana/\$2.90 question, biased participants were expected to tend toward subtracting \$2.00 from \$2.90 with just as much intuitive ease as \$1.00 from \$1.10. Moreover, the bat-and-ball problem has become fairly well known, so inclusion of these alternative questions was to limit any possible improved performance due to previous exposure. In addition, participants' confidence ratings were obtained after each critical question with the following prompt: "How confident are you in your

response? Please write down a number between 0% (totally not sure) to 100% (totally sure)."

Cueing condition descriptions

Overview No feedback was provided after any of the questions. All participants in each cueing condition solved one problem variant in its standard form as the critical question, counterbalanced such that half of the participants from each source were given one variant and the other half given the other variant. In one cueing condition (control with no initial problem), the participants were shown only the pencil/eraser/\$1.10 critical question or the magazine/banana/\$2.90 critical question, both in their standard form. In all other cueing conditions, the critical question was preceded by a cue. In one of these cueing conditions (algebra irrelevant), the cue was an unrelated algebra problem. In the remaining cueing conditions, the cue was some form of the alternate problem. Thus, participants never saw a given problem variant, nor equations corresponding to a problem variant, more than once.

Algebra plus words Half of the participants in this cueing condition were initially cued with the following question and corresponding algebraic equations:

A pencil and an eraser together cost \$1.10. The pencil costs \$1.00 more than the eraser. How much does the eraser cost?

$$\begin{aligned} P + E &= \$1.10 \\ P &= \$1.00 + E \\ \text{Solve for } E \end{aligned}$$

Because these were complementary forms of the same question, on the Qualtrics survey screen only one text box was available for an answer. Following this question, participants were asked to solve the critical question magazine/banana/\$2.90 in its standard form (i.e., without a cue):

A magazine and a banana together cost \$2.90. The magazine costs \$2.00 more than the banana. How much does the banana cost?

The remaining half of the participants in this cueing condition were counterbalanced such that their initial cue was the magazine/banana/\$2.90 problem and corresponding equations, with the pencil/eraser/\$1.10 problem in its standard form as the critical question. The goal was to limit ambiguity about what the question was asking without being explicit about the solution.

Algebra only Half of the participants in this cueing condition were initially cued with just the pencil/eraser/\$1.10 algebraic equations presented as follows:

Consider the following equations. Use the substitution method to solve for the value of E

$$\begin{aligned} P + E &= \$1.10 \\ P &= \$1.00 + E \\ \text{Solve for E} \end{aligned}$$

Following this question, participants were asked to solve the critical question magazine/banana/\$2.90 in its standard form (i.e., without the relevant algebraic equations). To counterbalance problem variants, the remaining half of the participants in this cueing condition were cued with the magazine/banana/\$2.90 algebraic equations, and their critical question was the pencil/eraser/\$1.10 in its standard form. This cueing condition allowed us to explore the effect of algebraic cueing without the written expression of the question included in the cue.

Algebra irrelevant All participants in this cueing condition were initially cued with the following algebra question:

Consider the following equations. Use the substitution method to solve for the value of X

$$\begin{aligned} X + 3Y &= 13 \\ Y &= 6 - 2X \\ \text{Solve for X} \end{aligned}$$

After this question, half of the participants were asked to solve the critical question magazine/banana/\$2.90 in its standard form (i.e., without the algebraic equations). The other half of the participants were asked to solve the pencil/eraser/\$1.10 critical question in its standard form. This cueing condition allowed us to determine the effect of algebraic cueing in the context of an algebra problem unrelated to the specific structure of bat-and-ball problem types.

Control with initial problem Half of the participants in this cueing condition were first asked the pencil/eraser/\$1.10 question followed by the magazine/banana/\$2.90 critical question, both in their standard form. The order of question presentation was reversed for the remaining participants. This control allowed us to determine the unique effect of algebraic cueing beyond the effect of simply being asked an initial relevant question.

Control with no initial problem Half of the participants in this cueing condition were asked the pencil/eraser/\$1.10 critical question, whereas the other half were asked the magazine/banana/\$2.90 critical question, both in their standard form with no prior cue. Participants were expected to demonstrate error rates consistent with prior related research.

Participants were then asked to complete the five-question Mathematics Anxiety/Confidence (MANX) questionnaire to determine the extent to which overall math anxiety and math

confidence were related to problem solution accuracy (see Chipman, Krantz, & Silver, 1992). Participants selected a rating on a scale with five levels, ranging from strongly agree to strongly disagree (e.g., “Working on mathematics problems makes me tense”). Low scores indicated math anxiety; high scores indicated math confidence.

Finally, participants were asked to provide a rating to the following question: “On a scale of 1 to 10, how math anxious are you?” Ashcraft (2002) found that, for quick determination of math anxiety, simply asking this question was sufficient and correlated with scores on the shortened Mathematics Anxiety Rating Scale.

Results

Survey completion was 100% with no attrition. See Table 1 for cueing question descriptive statistics.

A two (subject group: Mechanical Turk and laboratory) $\times$ five (cueing condition) analysis of variance (ANOVA) was conducted for critical question accuracy. Participants’ accuracy on the critical question depended on cueing condition (algebra plus words $M = .475$, $SD = .506$; algebra only $M = .744$, $SD = .442$; algebra irrelevant $M = .526$, $SD = .506$; control with initial problem $M = .425$, $SD = .501$; control with no initial problem $M = .225$, $SD = .423$); the main effect of cueing condition was significant, $F(4, 187) = 5.96$, $p < .001$, $\eta^2 = .11$ (see Fig. 1). There was no main effect for subject group, $F < 1$, and no Cueing Condition $\times$ Subject Group interaction, $F < 1$. As predicted, there was a significant main effect of algebraic cueing on accuracy (collapsing across experimental cueing conditions: algebra plus words, algebra only, and algebra irrelevant), compared to not being cued to reason algebraically (collapsing across control cueing conditions: control with initial problem and control with no initial problem), $F(1, 192) = 13.76$, $p < .001$, $\eta^2 = .07$.

Post hoc comparisons using Tukey’s HSD indicated significant differences ($p < .05$) between algebra only and control with initial problem, algebra only and control with no initial problem, and algebra irrelevant and control with no initial problem.

Table 1 Mean, number, and standard deviation for cueing question accuracy by subject group and cueing condition

Cueing condition	Mechanical Turk			Laboratory		
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
Algebra plus words	20	.400	.503	20	.650	.489
Algebra only	20	.950	.224	19	.947	.229
Algebra irrelevant	20	.700	.470	18	.778	.428
Control with initial problem	20	.300	.470	20	.250	.444

Fig. 1 Proportion of correct responses on the critical question as a function of cueing condition and subject group. The error bars represent standard errors of the mean. AW algebra plus words, AO algebra only, AI algebra irrelevant, CP control with initial problem, CN control with no initial problem

A logistic regression analysis was also conducted on the critical question because of the dichotomous dependent variable (correct or incorrect). This analysis revealed identical patterns of significance to the ANOVAs (see [Supplemental Materials](#)).

There was a significant positive correlation between cueing question accuracy and critical question accuracy, $r(155) = .486, p < .001$, such that those who correctly solved the initial question were more likely to solve the critical question accurately. There was a significant negative correlation between the Ashcraft (2002) question and accuracy on the critical question, $r(195) = -.303, p < .001$. That is, those who indicated higher levels of math anxiety were less likely to answer correctly. There was also a positive correlation between the MANX scale and accuracy on the critical question, $r(195) = .345, p < .001$. On this scale, high scores indicated more math confidence. Therefore, this scale indicates that higher degrees of math confidence are related to an overall tendency to answer correctly on bat-and-ball problem variants. Unsurprisingly, there was a strong negative association between these anxiety scales, $r(195) = -.681, p < .001$.

Confidence ratings were significantly related to critical question accuracy, $r(195) = .155, p = .030$, to responses from the MANX scale, $r(195) = .335, p < .001$, and to the single question from Ashcraft (2002), $r(195) = -.278, p < .001$.

An ANOVA on confidence ratings revealed no significant effects (all $F < 1$), presumably because of ceiling effects such that the modal response was 100% (Mechanical Turk $M = 93.9\%$, $SD = 14.4\%$; Laboratory $M = 92.4\%$, $SD = 19.1\%$).

Two (subject group) $\times$ five (cueing condition) $\times$ two (problem variant) factorial ANOVAs were conducted on both accuracy and confidence to compare the two problem variants as critical questions (pencil/eraser/\$1.10, magazine/banana/\$2.90). For both accuracy and confidence, the factor of problem variant did not yield a significant main effect nor enter into any significant interactions ($p > .05$ in each case). For accuracy, pencil/eraser/\$1.10 $M = .500$, $SD = .503$; magazine/banana/\$2.90 $M = .455$, $SD = .500$.

Discussion

Support for the main hypothesis was found. Across both the Mechanical Turk and laboratory samples, algebraically-cued participants significantly outperformed non-cued participants. Post hoc comparisons revealed that the algebra only condition was more than three times as accurate on the critical question as was the control with no initial problem, and the algebra only condition was the only experimental cueing condition that significantly outperformed the control with initial problem. These findings suggest that most of the unique effect of algebraic cueing occurred in the algebra only condition. Despite evidence for the predicted overall effect of algebra-cueing (comparing experimental to control cueing conditions), this result indicates that general algebraic cueing is not the most important determinant. We had predicted best performance in the algebra plus words condition, because these participants were presented with not only the related mathematical equations but also the context of the question itself. However, the data suggest that this was not the case.

Only 52.5% of the participants in the algebra plus words condition solved their cueing question correctly. On the other hand, participants in the algebra only condition (who were asked to solve just the relevant algebraic equations) overwhelmingly answered their cueing question correctly (94.9% correct overall). The prepotent intuitive response elicited (as a result of the way the bat-and-ball problem types are phrased) is so strong that participants in the algebra plus words condition appear to have arrived at the intuitive answer during algebraic cueing and did not feel the need for further calculation, despite having all the relevant algebraic information presented to them. Yet, for participants in the algebra only condition, who were forced to solve the relevant algebra, a benefit to overall accuracy on the critical question was observed.

A possible explanation for this finding is that being forced to work through the math (as opposed to being given the option to answer intuitively) explicitly highlighted the mathematical structure of the “more than” phrase during cueing (e.g., $P = \$1.00 + E$), which reasoners were then able to generalize to the critical question, despite the differing item content and dollar amounts from cue to critical question. These findings also emphasize the extent of reasoners’ cognitive

miserliness on the bat-and-ball problem, and are consistent with common descriptions of System 2 as all too quick to accept the intuitions provided by System 1 (e.g., Evans & Stanovich, 2013). However, it is worth bearing in mind the limitations of dual-process explanations: None of the reported effects necessitate the use of unconscious processing, which had been attributed to automatic, System 1 thinking. Instead, by having participants initially solve a related question in the form of algebra equations, the likelihood was increased that reasoners would explicitly detect and judge the “more than” phrase to be a relevant and necessary algebraic cue to inform their decision on the critical question. Given the importance of numeracy in solving CRT questions (e.g., Campitelli & Gerrans, 2014), it is likely that, when cued with algebra, many correct responders utilized algebraic reasoning as their primary strategy to solve the problem, perhaps without, or with less, consideration of the intuitive answers.

The high performance on the critical question in the algebra only condition was not observed in the algebra irrelevant condition, although performance in that cueing condition was significantly higher than in the control with no initial problem. Several explanations seem possible for this finding: (1) because fewer participants were able to solve this algebraic cueing question correctly (73.7% in algebra irrelevant and 94.9% in algebra only), it is possible that this problem was simply too challenging to elicit a change in problem solution; (2) cueing algebraic reasoning with equations unrelated to the structure of bat-and-ball problem types is not as effective as more closely related mathematical structure in eliciting a change in problem solution; or (3) a combination of these two possibilities.

Significant correlations between accuracy on the critical questions and responses on the MANX scale and the single question from Ashcraft (2002) were also observed. This finding indicates that, as hypothesized, math anxiety is associated with the overall tendency to engage in biased reasoning on the critical question. Furthermore, confidence ratings were significantly related both to critical question accuracy and to responses from the math anxiety measures. Because we did not assess participants’ numeric abilities, we are limited in terms of the implications our math anxiety findings have on the analytical predisposition versus numeracy debate (Pennycook & Ross, 2016; Sinayev & Peters, 2015).

A limitation of the present investigation was that only one of the three problems traditionally asked in the full CRT was addressed (Frederick, 2005). However, cueing algebraic reasoning is unlikely to be the path toward eliciting correct reasoning for the other two problems. The bat-and-ball problem has a very clear algebraic structure, whereas the other two problems lack such structure. Despite these limitations, cueing the numeric components found in the other two CRT problems, and across other decision-making tasks that necessitate numeric ability, provides fruitful grounds for subsequent debiasing investigations.

In conclusion, the findings of the present investigation, demonstrating more than a threefold increase in critical question accuracy from the control with no initial problem to the algebra only condition, indicate that algebraic reasoning is an effective debiasing strategy for the bat-and-ball problem, and shed new light on the role that math anxiety plays in CRT performance.

Acknowledgements This research was supported in part by NSF Grant DRL1246588 to the University of Colorado Boulder. We thank Lesley Hathorn and Susan Chipman for their helpful discussions about this research and direction to the math anxiety measures we employed. We also thank Sabine Doebel, Shaw Ketels, Mason Eastwood, and other members of the Healy/Jones laboratory for their helpful comments and suggestions. We are particularly grateful to Shaw Ketels for his help in our discussion of theoretical issues and guidance to the relevant literature. A special thanks to the reviewers Guillermo Campitelli and Mark H. Ashcraft, and to the action editor Ben Newell, whose comments and suggestions greatly improved the quality of our report.

References

- 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, 569–576. doi:10.1037/0096-3445.136.4.569
- Ashcraft, M. H. (2002). Math anxiety: Personal, educational, and cognitive consequences. *Current Directions in Psychological Science*, 11, 181–185. doi:10.1111/1467-8721.00196
- Baron, J., Scott, S., Fincher, K., & Metz, S. E. (2015). Why does the Cognitive Reflection Test (sometimes) predict utilitarian moral judgment (and other things)? *Journal of Applied Research in Memory and Cognition*, 4, 265–284. doi:10.1016/j.jarmac.2014.09.003
- Bourgeois-Gironde, S., & Vanderhenst, J. B. (2009). How to open the door to System 2: Debiasing the Bat and Ball problem. In S. Watanabe, A. P. Bloisdell, L. Huber, & A. Young (Eds.), *Rational animals, irrational humans* (pp. 235–252). Tokyo: Keio University Press. doi: ijn_00432320
- Campitelli, G., & Gerrans, P. (2014). Does the cognitive reflection test measure cognitive reflection? A mathematical modeling approach. *Memory & Cognition*, 42, 434–447. doi:10.3758/s13421-013-0367-9
- Chipman, S. F., Krantz, D. H., & Silver, R. (1992). Mathematics anxiety and science careers among able college women. *Psychological Science*, 3, 292–295.
- Cokely, E. T., & Kelley, C. M. (2009). Cognitive abilities and superior decision making under risk: A protocol analysis and process model evaluation. *Judgment and Decision Making*, 4, 20–33.
- De Neys, W., Rossi, S., & Houdé, O. (2013). Bats, balls, and substitution sensitivity: Cognitive misers are no happy fools. *Psychonomic Bulletin & Review*, 20, 269–273. doi:10.3758/s13423-013-0384-5
- Evans, J. S. B. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, 59, 255–278. doi:10.1146/annurev.psych.59.103006.093629
- Evans, J. S. B., & Stanovich, K. E. (2013). Dual-process theories of higher cognition advancing the debate. *Perspectives on Psychological Science*, 8, 223–241. doi:10.1177/1745691612460685
- Frederick, S. (2005). Cognitive reflection and decision making. *The Journal of Economic Perspectives*, 19, 25–42. doi:10.1257/089533005775196732

- Gigerenzer, G., & Selten, R. (2002). *Bounded rationality: The adaptive toolbox*. Cambridge: MIT Press.
- Hathorn, L., & Healy, A. F. (2016). *Attribute substitution in the bat-and-ball problem*. Poster presented at the 28th APS Annual Convention, Chicago, IL.
- Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.
- Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgement. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49–81). Cambridge: Cambridge University Press.
- Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80, 237–251. doi:[10.1037/h0034747](https://doi.org/10.1037/h0034747)
- Keren, G., & Schul, Y. (2009). Two is not always better than one: A critical evaluation of two-system theories. *Perspectives on Psychological Science*, 4, 533–550. doi:[10.1111/j.1745-6924.2009.01164.x](https://doi.org/10.1111/j.1745-6924.2009.01164.x)
- Mastrogiorgio, A., & Petracca, E. (2014). Numerals as triggers of System 1 and System 2 in the ‘bat and ball’ problem. *Mind & Society*, 13, 135–148. doi:[10.1007/s11299-014-0138-8](https://doi.org/10.1007/s11299-014-0138-8)
- Miu, A. C., Heilman, R. M., & Houser, D. (2008). Anxiety impairs decision-making: Psychophysiological evidence from an Iowa Gambling Task. *Biological Psychology*, 77, 353–358. doi:[10.1016/j.biopsycho.2007.11.010](https://doi.org/10.1016/j.biopsycho.2007.11.010)
- Newell, B. R., & Shanks, D. R. (2014). Unconscious influences on decision making: A critical review. *Behavioral and Brain Sciences*, 37, 1–19. doi:[10.1017/S0140525X12003214](https://doi.org/10.1017/S0140525X12003214)
- Pennycook, G., & Ross, R. M. (2016). Commentary: Cognitive reflection vs. calculation in decision making. *Frontiers in Psychology*, 7, 9. doi:[10.3389/fpsyg.2016.00009](https://doi.org/10.3389/fpsyg.2016.00009)
- Raghunathan, R., & Pham, M. T. (1999). All negative moods are not equal: Motivational influences of anxiety and sadness on decision making. *Organizational Behavior and Human Decision Processes*, 79, 56–77. doi:[10.1006/obhd.1999.2838](https://doi.org/10.1006/obhd.1999.2838)
- Sinayev, A., & Peters, E. (2015). Cognitive reflection vs. calculation in decision making. *Frontiers in Psychology*, 6, 532. doi:[10.3389/fpsyg.2015.00532](https://doi.org/10.3389/fpsyg.2015.00532)
- Stanovich, K. E., & West, R. F. (2002). Individual differences in reasoning: Implications for the rationality debate? In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 421–440). Cambridge: Cambridge University Press. doi:[10.1017/CBO9780511808098.026](https://doi.org/10.1017/CBO9780511808098.026)
- Thompson, V. A., Prowse Turner, J. A., Pennycook, G., Ball, L. J., Brack, H., Ophir, Y., & Ackerman, R. (2013). The role of answer fluency and perceptual fluency as metacognitive cues for initiating analytic thinking. *Cognition*, 128, 237–251. doi:[10.1016/j.cognition.2012.09.012](https://doi.org/10.1016/j.cognition.2012.09.012)
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124–1131. doi:[10.1126/science.185.4157.1124](https://doi.org/10.1126/science.185.4157.1124)