

BATS, BALLS, AND SUSTITUTION SENSITIVITY: COGNITIVE MISERS ARE NO HAPPY FOOLS

Wim De Neys^{1,2,3}, Sandrine Rossi^{1,2,3}, Olivier Houdé^{1,2,3}

1 - CNRS, Unité 3521 LaPsyDÉ, France

2 - Université Paris Descartes, Sorbonne Paris Cité, Unité 3521 LaPsyDÉ, France

3 - Université de Caen Basse-Normandie, Unité 3521 LaPsyDÉ, France

Word Count : 3408 words

Mailing address: Wim De Neys
LaPsyDÉ (Unité CNRS 3521, Université Paris Descartes)
Sorbonne - Labo A. Binet
46, rue Saint Jacques
75005 Paris
France

wim.de-neys@parisdescartes.fr

Abstract

Influential work on human thinking suggests that our judgment is often biased because we minimize cognitive effort and intuitively substitute hard questions by easier ones. A key question is whether or not people realize they are doing this and notice their mistake. Here we test this claim with one of the most publicized examples of the substitution bias, the bat-and-ball problem. We designed an isomorphic control version in which reasoners experience no intuitive pull to substitute. Results show that people are less confident in their substituted, erroneous bat-and-ball answer than in their answer on the control version that does not give rise to the substitution. Contrary to popular belief, this basic finding indicates that biased reasoners are not completely oblivious to the substitution and sense that their answer is questionable. This questions the characterization of the human reasoner as a happy fool who blindly answers erroneous questions without realizing it.

Bats, Balls, and Substitution Sensitivity: Cognitive Misers Are No Happy Fools

Human reasoners have been characterized as cognitive misers who show a strong tendency to rely on fast, intuitive processing rather than on more demanding, deliberate thinking (Evans, 2008; Kahneman, 2011). Although the fast and effortless nature of intuitive processing can sometimes be useful, it can also bias our reasoning. It has been argued that the key to this bias is a process of so-called attribute substitution – when people are confronted with a difficult question they often intuitively answer an easier one instead (e.g., Kahneman, 2011; Kahneman & Frederick, 2002). Consider the following example:

A bat and a ball together cost \$1.10. The bat costs \$1 more than the ball. How much does the ball cost?

When you try to answer this problem, the intuitive answer that immediately springs to mind is “10 cents”. Indeed, about 80% of university students who are asked to solve the “bat-and-ball” problem give the “10 cents” answer (e.g., Bourgeois-Gironde & Vanderhenst, 2009). But it is wrong. Obviously, if the ball were to cost 10 cents, the bat would cost \$1.10 (i.e., \$1 more) and then the total cost would be \$1.20, rather than the required \$1.10. The correct response is “5 cents”, of course (i.e., the bat costs \$1.05). The explanation for the widespread “10 cents” bias in terms of attribute substitution is that people substitute the critical relational “more than” statement by a simpler absolute statement. That is, “the bat costs \$1 more than the ball” is read as “the bat costs \$1”. Hence, rather than working out the sum, people naturally parse \$1.10, into \$1 and 10 cents which is easier to do. In other words, because of the substitution people give the correct answer to the wrong question.

The bat-and-ball problem is considered a paradigmatic example of people’s cognitive miserliness (e.g., Bourgeois-Gironde & Vanderhenst, 2009; Kahneman, 2011; Kahneman & Frederick, 2002; Toplak, West, & Stanovich, 2011). After all, the problem is really not that hard. Clearly, if people would reflect upon it for even a moment they would surely realize their error and notice that a 10 cents ball and a bat that costs a dollar *more* cannot total to \$1.10. Hence, the problem with attribute substitution seems to be that people do typically not notice that they are substituting and do not realize their error (Kahneman & Frederick, 2005; Thompson, 2009; Toplak et al., 2011). This can sketch a somewhat bleak picture of human rationality: Not only do we often fail to reason correctly, much like happy fools, we do not even seem to realize that we are making a mistake.

However, the fact that decision-makers do not deliberately reflect upon their response does not *necessarily* imply that they are not detecting the substitution process. That is, although people might not engage in deliberate processing and might not know what the correct answer is, it is still possible that they have some minimal substitution sensitivity and at least notice that their substituted “10 cents” response is not completely warranted. To test this hypothesis we designed a control version of the bat-and-ball problem that does not give rise to attribute substitution. Consider the following example:

A magazine and a banana together cost \$2.90. The magazine costs \$2. How much does the banana cost?

People will tend to parse the \$2.90 into \$2 and 90 cents just as naturally as they parse \$1.10 in the standard version. However, the control version no longer contains the relative statement (“\$2 more than the banana”) that triggers the substitution. That is, in the control version we explicitly present the easier statement that participants are supposed to be unconsciously substituting. After solving each version participants were asked to indicate their response confidence. If participants are completely unaware that they are substituting when solving the standard version, the standard and control version should be isomorphic and response confidence should not differ. However, if we are right that people might not be completely oblivious to the substitution and have some minimal awareness of the questionable nature of their answer, response confidence should be lower after solving the standard version.

Method

Participants

A total of 248 University of Caen undergraduates who took an introductory psychology course participated voluntarily.

Material and procedure

Participants were presented with a standard and control version of the bat-and-ball problem. The problems were translated in French and adjusted to the European test context (see Supplementary Material). To minimize surface similarity, we also modified the superficial item content of the two problems (i.e., one problem stated that a pencil and eraser together cost \$1.10, the other that a magazine and banana together cost \$2.90). Both problems were printed on separate pages of a booklet.

To make sure that the differential item content did not affect the findings, the item content and control status of the problem were completely crossed. For half of the sample we used the pencil/eraser/\$1.10 content in the standard version and magazine/banana/\$2.90 content in the control version. For the other half of the sample the content of the two presented problems was switched. Presentation order of the control and standard version was also counterbalanced: Approximately half of the participants solved the control version first, whereas the other half started with the standard version. An overview of the material is presented in the Supplementary Material section.

Immediately after participants wrote down their answer they were asked to indicate how confident they were that their response was correct by writing down a number between 0 (totally not sure) and 100% (totally sure). Note that we only intend to use this measure to contrast people's relative confidence difference in the standard and control versions. Obviously, the confidence ratings will be but a proxy of people's phenomenal confidence state. The response scale is not immune to measurement biases such as end preferences or social desirability effects (e.g., Berk, 2006). For example, since it might be hard to openly admit that one has given a response that one is not confident about, mere social desirability can drive people's estimates upwards. This implies that one needs to be cautious when interpreting absolute confidence levels. However, such interpretative complications can be sidestepped when contrasting the relative rating difference in two conditions. Any general response scale bias should affect the ratings in both conditions. Consequently, our analyses focus on the relative confidence contrast and we refrain from making claims based on the absolute confidence levels.

Results

Accuracy. In line with previous studies, only 21% (SE = 2.3%) of participants managed to solve the standard bat-and-ball problem correctly. Incorrect responses were almost exclusively (i.e., 194 out of 195 responses) of the "10 cents" type suggesting that biased participants were not simply making a random guess but indeed engaged in the postulated substitution process. As expected, the control version that did not give rise to substitution was solved correctly by 98% (SE = 1%) of the participants, $F(1, 247) = 714.94, p < .0001, \eta^2 p = .74^1$.

¹ The few incorrectly solved control trials and the "non-10 cents" incorrectly solved standard trial were discarded for the subsequent confidence analyses.

Confidence ratings. As Figure 1 shows, the response confidence of participants who engaged in a substitution process and gave the erroneous “10 cents” response on the standard version was significantly lower than their confidence in their control version answer that did not give rise to the substitution, $F(1, 194) = 58.54, p < .0001, \eta^2p = .23$. Consistent with our hypothesis, this establishes that biased reasoners are sensitive to the substitution and are not completely oblivious to the erroneous nature of their substituted judgment. Figure 1 also shows that this confidence decrease was far less clear for reasoners who solved the standard problem correctly, $F(1, 47) = 4.89, p < .05, \eta^2p = .09$. Indeed, it makes sense that the few non-misers who reflected upon their judgment and resisted the substitution also knew that their response was likely to be correct.

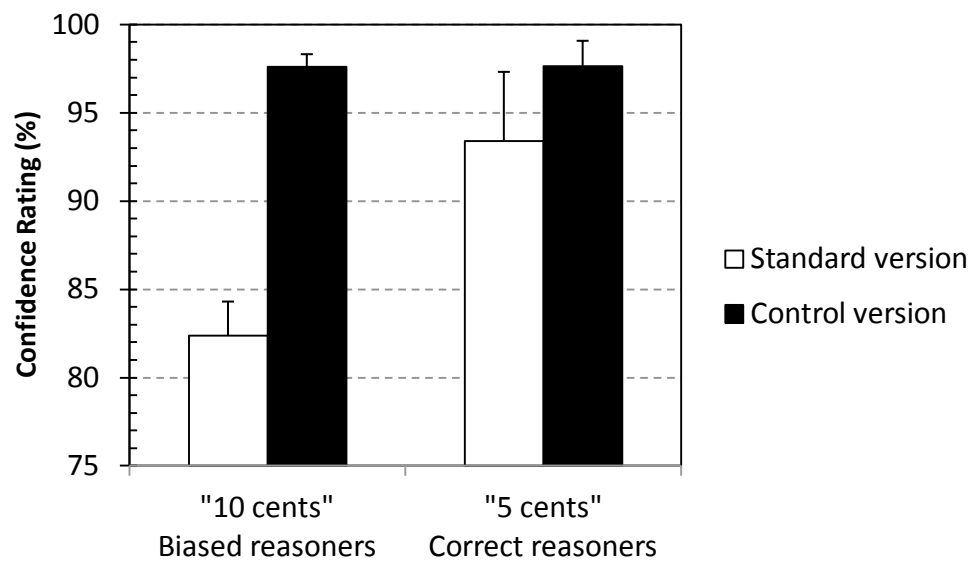


Figure 1. Response confidence on standard and control versions of the bat-and-ball problem for participants who answered the standard problem incorrectly (“10 cents” biased reasoners) and correctly (“5 cents” correct reasoners). Error bars are standard errors.

A critic might note that by presenting both the control and standard version to the same participants we might have artificially directed their attention to the substitution. We already tried to limit this problem by restricting the surface resemblance of the two problems but to eliminate this possible confound completely we also ran a control analysis that was restricted to the first problem that participants solved (i.e., half of the participants solved the control version first, whereas the other half started with the standard version). Results of this between-subject analysis confirmed our findings. Biased reasoners who failed to solve the standard version were less confident in their answer ($M = 85\%$,

SE = 1.8%) than reasoners who solved the control version (M = 98%, SE = 1.6%), $F(1, 222) = 26.89$, $p < .0001$, $\eta^2p = .11$.

Discussion

The present data establish that reasoners are not completely oblivious to their substitution bias. When people substitute a harder question for an easier one, their response confidence indicates that they show some minimal awareness of the questionable nature of their answer. Hence, although reasoners might typically fail to reflect on the problem and might not know the correct answer (Frederick, 2005), they at least seem to sense that their substituted response is not fully warranted. Bluntly put, what these data suggest is that although we might be cognitive misers, we are no happy fools who blindly answer erroneous questions without realizing it.

To be clear, our findings do not argue against the popular characterization of the human decision-maker as a cognitive miser *per se*. Note that we replicated the massive preference for the substituted “10 cents” answer, for example. In line with most authors (e.g., Evans, 2010; Frederick, 2005; Kahneman, 2011; Stanovich, 2010), we also believe that the key reason for the substitution bias is that reasoners tend to minimize cognitive effort and stick to mere intuitive processing. However, the point we want to stress is that despite this lack of deliberate reflection, reasoners are not at the blind mercy of a substitution process. More generally, our findings suggest that cognitive misers might have more accurate intuitions about the substitution process than hitherto believed². Although people experience a strong intuitive pull to engage in substitution and fail to deliberately reflect on their answer, our data suggest that at the same time they also sense that the substituted response is questionable.

At a more general level, a number of authors have recently suggested that such intuitive conflict sensations might act as a cue that allows our reasoning engine to determine whether it is needed to engage in deliberate thinking (e.g., Alter, Oppenheimer, Epley, & Eyre, 2007; De Neys, 2012; Thompson & Morsanyi, 2012; Thompson, Turner, & Pennycook, 2011). Thompson and colleagues (Thompson & Morsanyi, 2012; Thompson et al., 2011; see also Oppenheimer, 2008) have linked this process to the metacognitive memory literature (e.g., Koriat, 1993) and labeled it the “Feeling of Rightness”. In terms of

² See De Neys (2012) for a related suggestion with respect to the intuitive detection of violations of logical and probabilistic rules in deductive and statistical reasoning tasks.

this model, one could argue that the present data indicate that people's "Feeling of Rightness" is lowered when they substitute. Bluntly put, people feel that their biased response is not "right". In line with our claims, this suggests that the problem with substitution and judgment bias in general is not that people do not realize that they need to think harder, but rather that this deliberate processing is not successfully engaged.

It is important to clarify some potential misconceptions and critiques about our work. For example, some critics might spontaneously argue that since our control bat-and-ball version is easier than the standard version our findings are trivial since they simply show that people are more confident when answering an easy question than when answering a hard question. It is important to stress that this critique is begging the question. The crucial question is of course whether or not people realize that the classic version is hard. That is, the control version presents the easier statement that participants are supposed to be unconsciously substituting. What we want to know is whether or not people note this substitution. If people do not notice it, then the two problems should be isomorphic and they should be considered equally hard. In other words, arguing that people notice that the classic problem is harder than the control problem underscores the point that they are not oblivious to the substitution.

A related spontaneous critique is that our confidence findings might result from mere guessing rather than from substitution sensitivity. In general, if people do not know an answer to a problem and guess, they presumably realize this and will also give a low confidence rating. Hence, a critic might argue that the lower confidence does not necessarily point to substitution sensitivity but merely to a rather trivial "guessing awareness". However, this critique is readily discarded. In the present study more than 99% of the erroneous bat-and-ball responses were of the "10 cents" type. This is the response that people should pick if they engage in the postulated substitution process. Clearly, if people were biased and less confident because they were merely guessing, we should have observed much more random erroneous answers.

In the present study we focused on the bat-and-ball problem because it is one of the most vetted and paradigmatic examples of people's substitution bias (e.g., Bourgeois-Gironde & Vanderhenst, 2009; Kahneman, 2011; Kahneman & Frederick, 2002; Toplak, West, & Stanovich, 2011). However, attribute substitution has also been proposed as an explanation for people's judgment errors in other classic reasoning tasks such as the base-rate neglect or conjunction fallacy task (Kahneman & Frederick, 2002). Although it has been argued that these tasks might be less suited to test substitution claims (e.g., Bourgeois-Gironde & Vanderhenst, 2009), one might nevertheless wonder whether the present findings can be generalized across these tasks. Some emerging evidence suggests they might. For example, a

recent study showed that when reasoners give a biased response to standard conjunction or base-rate neglect problems, they also indicate to be less confident about their response compared to control problems (e.g., De Neys, Cromheeke, & Osman, 2011; see also De Neys & Feremans, 2012). This gives us some initial indication of the generality of the present findings.

With the present paper we hope to have presented a critical building block to stimulate further research on people's substitution sensitivity. Obviously, we acknowledge that our findings will need to be extended but we nevertheless feel it is important to bring them under the attention of the wide range of researchers interested in human thinking. At the very least the present data should alert scholars that the popular idea that substitution typically goes unnoticed is disputable and needs closer empirical testing.

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.
- Berk, R. A. (2006). Thirteen strategies to measure college teaching: a consumer's guide to rating scale Construction, assessment, and decision making for faculty, administrators, and clinicians. Sterling, VA: Stylus.
- 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, Japan: Keio University Press.
- De Neys, W. (2012). Bias and conflict: A case for logical intuitions. *Perspectives on Psychological Science*, *7*, 28-38.
- De Neys, W., Cromheeke, S., & Osman, M. (2011). Biased but in doubt: Conflict and decision confidence. *PLoS ONE*, *e15954*. DOI:10.1371/journal.pone.0015954
- De Neys, W., & Feremans, V. (2012). Development of heuristic bias detection in elementary school. *Developmental Psychology*. DOI: 10.1037/a0028320. ISSN: 0012-1649
- Evans, J. St. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, *59*, 255-278.
- Evans, J. B. S. T. (2010). Intuition and reasoning: A dual process perspective. *Psychological Inquiry*, *21*, 313-326.

- Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*, 25-42.
- Kahneman, D. (2011). *Thinking, Fast and Slow*. New York: Farrar, Strauss, Giroux.
- Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics & biases: The psychology of intuitive judgment* (pp. 49-81). New York: Cambridge University Press.
- Kahneman, D., & Frederick, S. (2005). A model of heuristic judgment. In K. J. Holyoak, & R. G. Morrison (Eds.), *The Cambridge Handbook of Thinking and Reasoning* (pp. 267-293). New York: Cambridge University Press.
- Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review*, *100*, 609-639.
- Oppenheimer, D. M. (2008). The secret life of fluency. *Trends in Cognitive Science*, *12*, 237-241.
- Stanovich, K. E. (2010). *Rationality and the reflective mind*. New York: Oxford University Press.
- Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The cognitive reflection test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition*, *39*, 1275-1289.
- Thompson, V. A. (2009). Dual process theories: A metacognitive perspective. In J. Evans and K. Frankish (Eds.), *In Two Minds: Dual Processes and Beyond*. Oxford, UK: Oxford University Press.
- Thompson, V. A. & Morsanyi, K. (2012). Analytic thinking: Do you feel like it? *Mind & Society*, *11*, 93-105.
- Thompson, V. A., Turner, J. P., & Pennycook, G. (2011). Choosing between intuition and reason: The role of metacognition in initiating analytic thinking. *Cognitive Psychology*, *63*, 107-140.

Supplementary Material

Standard versions

French

Un crayon et une gomme coûtent 1.10 euro au total. Le crayon coûte 1 euro de plus que la gomme. Combien coûte la gomme?

_____ centimes

Un magazine et une banane coûtent 2.90 euros au total. Le magazine coûte 2 euros de plus que la banane. Combien coûte la banane?

_____ centimes

English translation

A pencil and an eraser cost 1.10 euro in total. The pencil costs 1 euro more than the eraser. How much does the eraser cost?

_____ cents

A magazine and a banana cost 2.90 euro in total. The magazine costs 2 euro more than the banana. How much does the banana cost?

_____ cents

Control versions

French

Un crayon et une gomme coûtent 1.10 euro au total. Le crayon coûte 1 euro. Combien coûte la gomme?

_____ centimes

Un magazine et une banane coûtent 2.90 euros au total. Le magazine coûte 2 euros. Combien coûte la banane?

_____ centimes

English translation

A pencil and an eraser cost 1.10 euro in total. The pencil costs 1 euro. How much does the eraser cost?

_____ cents

A magazine and a banana cost 2.90 euro in total. The magazine costs 2 euro. How much does the banana cost?

_____ cents