The unbearable burden of executive load on cognitive reflection: A validation of dual process theory

Eric D. Johnson (eric.johnson@ub.edu)
Department of Basic Psychology and IR3C, University of Barcelona, Barcelona, Spain

Elisabet Tubau (etubau@ub.edu)
Department of Basic Psychology and IR3C, University of Barcelona, Barcelona, Spain

Wim De Neys (wim.de-neys@parisdescartes.fr)
CNRS, LaPsyDE (CNRS Unit 8042), Paris Descartes University, Paris, France

Abstract

The Cognitive Reflection Test (CRT) is increasingly employed to measure the tendency of an individual to override a prepotent but incorrect response and to subsequently engage in further reflection. This interplay between fast intuitive responding and resource demanding reflection has been offered as a paradigmatic example of dual process theories of thinking. Despite its growing popularity both for dual process theories and as an easily deployed measure of intelligence, the basic assumption that the CRT relies on executive resources remains generally circumstantial. The present study directly tested these dual process assumptions by presenting the standard bat-and-ball problem and a no-conflict control version while manipulating executive resources with a secondary load task. With the no-conflict control problems, accuracy was uniformly at ceiling in no load, low load, and high load conditions. In sharp contrast, in the standard conflict problems accuracy clearly declined with increasingly load. These findings validate dual process assumptions by providing direct causal evidence that correctly resolving the bat-and-ball problem draws on executive resources.

Keywords: Cognitive reflection; bat-and-ball problem; Dual process; Executive resources; Reasoning; Decision making; Bias

Introduction

In the face of difficulty, human reasoners often appear to forego the effortful processing that may be required and opt instead for less demanding intuitive responses (Kahneman, 2011). While many fast and frugal heuristics are no doubt adaptive in complex and reoccurring environments (Gigerenzer, 2007), thinking fast can also lead to quite embarrassingly erroneous responses in less routine settings. Quickly 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?

Intuitively, the answer "10 cents" quickly springs to mind. In fact, typically around 80% of university students, including those from elite schools such as MIT and Harvard, respond with this *intuitive*—but *incorrect*—answer (e.g. Bourgeois-Gironde & Vanderhenst, 2009; Frederick, 2005). If the bat costs $1 more than a 10-cent ball, the bat itself must cost $1.10. Summing up, a $1.10 bat + a $0.10 ball would equal $1.20, not $1.10 as stated in the problem. Does this imply that highly educated young adults think that '110 + 10' = '110'? Of course not. Rather, it suggests that even educated reasoners often do not invest the necessary effort to correct their initial intuition, and instead settle for a quickly derived response.

The above bat-and-ball problem[1] is extremely popular amongst researchers, and extremely difficult for participants. This popularity and difficulty are highly intertwined. First, high correlations are often observed between performance on this problem and other measures of intelligence or cognitive capacity, making it a tempting short alternative to more cumbersome measures such as standard IQ or working memory tests (Toplak, West, Stanovich, 2011). Second, like many traditional judgment and reasoning tasks arising out the heuristics-and-biases era, the above bat-and-ball problem is a paradigmatic example of a situation whose initially primed response is incorrect and which therefore must be overridden in order to correctly solve the task.

This latter point has made the bat-and-ball problem particularly attractive for dual process theories of thinking and reasoning (Kahneman, 2011; Kahneman & Frederick, 2002; Thompson, 2009). According to dual processes theories, human reasoning is characterized by two systems, or processing types, which respond to encountered information in different ways. A first intuitive type thinking (System 1) automatically triggers responses based on prior knowledge/beliefs or via a number of simplifying heuristics, whereas a second more deliberate type (System 2) demands the more controlled utilization of limited executive resources.

[1] The bat-and-ball problem, together with two other items, forms part of the Cognitive Reflection Test (CRT; Frederick, 2005). For a recently extended test, see Toplak, West & Stanovich (2013).

The bat-and-ball problem has captured the attention of researchers working within dual process frameworks precisely because such a strongly compelling but wrong answer jumps to mind so quickly. In order to correctly solve the problem, one must first detect that something is amiss with the rapid, compelling response provided by System 1. In principle, this takes little more than a simple check: "Wait a minute, if the ball costs 10 cents, then the bat costs $1.10… that can't be right!". Once the conflict is detected, this erroneous intuition can then be overridden with the more effortful engagement of System 2.

Although it is widely assumed that correctly solving the bat-and-ball problem draws on the limited executive resources of System 2, no direct evidence has been presented to validate this claim. So far, this idea is based primarily on correlational data demonstrating the above mentioned relationships between CRT performance and other measures of cognitive capacity (e.g. Aguilar, Johnson & Tubau, 2014; Cokely & Kelley, 2009; Lesage, Navarrete, & De Neys, 2012; Liberali et al, 2011). This is particularly troublesome, especially given the variety of criticism aimed at dual process frameworks (see e.g. De Neys, 2006a; 2006b; Evans & Stanovich, 2013; Gigerenzer & Regier, 1996; Keren & Schul, 2009).

In the present study we tested this central dual process claim by experimentally limiting executive resources with a dual-task paradigm. Dual process theorists claim that cognitive load should place differential burdens on effortful and intuitive processing (Evans, 2008; Evans & Stanovich, 2013; Sloman, 1996; Stanovich & West, 2000). If the bat-and-ball problem does indeed inextricably depend on the availability of executive resources, then performance on this task should decrease under cognitive load (for further discussion see De Neys, 2006a, 2006b).

As a crucial counterpoint, we also presented the following no-conflict control version of the bat-and-ball problem:

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

On this control version the rapidly cued intuitive response ("90 cents") does not conflict with other normative considerations, and therefore no inhibition of an erroneous prepotent response is required. De Neys, Rossi, and Houdé (2013) demonstrated very high accuracy on this control problem when presented as a simple reasoning task without additional load manipulations. Crucial for the automaticity assumption in dual process theories, accuracy should not decline on these no-conflict problems in the presence of the additional cognitive burden. Accordingly, this problem helps to restrain interpretations of any observed decline in accuracy under load in the standard bat-and-ball problem. For example, if burdening cognitive capacity simply reduces general resources which may be required for reading, comprehending, and responding in an experimental setting, then a corresponding decline in performance should also be observed on the control problem. If, however, it is the need to overcome the rapidly available intuitive response on the standard conflict problems which draws on cognitive resources, then no performance decline should be observed under cognitive load on the control problems.

Experiment

Method & Material

Participants. A total of 190 undergraduate students from the University of Barcelona were recruited for this task in exchange for course credit. Seven of these students reported being familiar with the bat-and-ball problem, and therefore only data from the remaining 183 participants (mean age = 20.46, SE=.23) was analyzed and reported here.

Reasoning Task. The reasoning tasks included both a standard conflict and a no-conflict control version of the bat-and-ball problem introduced above. As in previous work (De Neys et al, 2013), different contextual and numerical contents were used (see Appendix). One problem presented a bat and ball, the other presented a magazine and banana. In one problem the total cost was $1.10 with one item costing $1 more than the other; in the other problem the total cost was $2.90 with one item costing $2 more than the other. Item contents and values for the conflict and control versions were fully counterbalanced across participants, which helps to ensure that any observed effects are general and not driven by the specific material used (e.g. the ease of partitioning 1.10 and 10, or background beliefs about the price of specific items)[2]. A blank box with the label "cents" appeared on screen following the problem. Participants therefore typed only their numerical response into the box.

Load Task: Dot Memory. In the load conditions, participants were presented a secondary visuospatial storage task (De Neys, 2006a; Franssens & De Neys, 2009). Prior to the reasoning task, a pattern of dots was briefly presented (900ms) in a 3 x 3 grid for participants to memorize and keep in mind while reasoning. After the reasoning task, participants were subsequently presented a blank grid into which they clicked with the mouse to reproduce the remembered pattern (an indicated dot could be removed by clicking again). Two load conditions were used in the present study. In the high load (HL) condition, four dots were presented in a complex interspersed pattern, which has been established to interfere specifically with effort-demanding executive resources (Miyake et al, 2001). In the low load (LL) condition, three dots were presented in a single column, which should place only a minimal burden on executive resources (De Neys, 2006a; De Neys & Verschueren, 2006; see examples in Figure 1).

[2] None of these factors had any impact on performance.

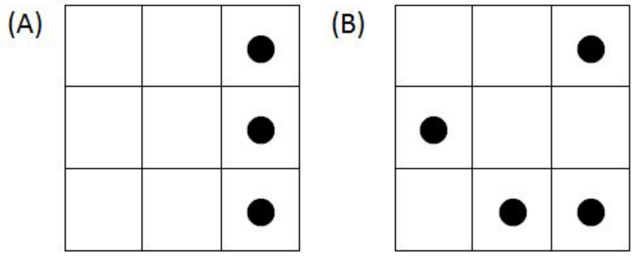

Figure 1. Example dots patterns presented as a dual-load task in the (A) Low Load condition, and (B) High Load condition.

Procedure. All tasks were adapted for computer-based testing. Participants were tested in small groups (up to four at a time) at individual computer terminals. All participants were randomized to receive the standard conflict (n=93) or the no-conflict control (n=90) problem in one of the three load conditions (HL, LL, NL). Appropriate task instructions were provided, along with a brief practice series to familiarize them with the testing environment, explained as follows.

In the (high and low) load conditions, participants first saw a simple and unrelated math story problem where they were to provide a single numerical response and press the Enter key. This was followed by a new screen with instructions explaining that they would also have to memorize a dot pattern to subsequently reproduce after the reasoning task. Participants then practiced the entire series: A pattern was briefly flashed for 900ms, followed by the same simple practice problem, and finally a blank grid appeared for participants to reproduce the previously seen dot pattern. As part of another study, prior to the appearance of the blank grid participants were also asked to provide a confidence judgment regarding the correctness of their response. The procedure was the same in the no load group, but without any mention or practice of the dot pattern task. Following the practice series the actual experiment began.

Results & Discussion

Load Task. On average, in the 3-dot low load condition participants correctly indicated 98% (2.95, *SD*=0.38) of the dot locations on the conflict problems and 100% (3.00, *SD*=0.00) of the dots on the control versions. In the 4-dot high load condition, 86% (3.42, *SD*=0.94) of the dot locations on the conflict problems and 85% (3.38, *SD*=0.79) of the dots on the control versions were correctly indicated. This shows that the secondary task was performed properly. There was no correlation between participants scores on the dot recall and reasoning task in the low load (r=.053, p=.68) or high load (r=.058, p=.65) conditions, which indicates that there was no performance trade-off between these tasks.

Accuracy. As expected, response accuracy was very high on the control problems in the no load (97%, *SE*=3.2%), low load (100%, *SE*=0.0%), and high load (100%, *SE*=0.0%) conditions. A logistic regression confirmed that performance was clearly not affected by executive load on these no-conflict problems ($\chi^2(1)<1$, $p>.95$). This establishes that the intuitive, and correct, response in the control version was automatically triggered with minimal involvement of executive resources.

In sharp contrast, on the standard conflict versions a clear decline in correct responses was observed with increasing load, from 23% (*SE*=8.4%) with no load, to 16% (*SE*=6.1%) with low load, to 3% (*SE*=3.3%) under high load (Figure 2). Logistic regression revealed that this effect of load on accuracy was significant (χ^2 (1)=4.25, p=.039, e^{β}=.41, 95% C.I.=.18 – .96). This finding provides direct causal evidence for the typically assumed claim the CRT bat-and-ball problem draws on executive resources.

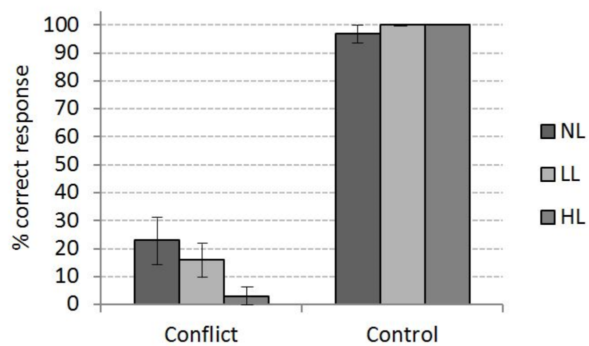

Figure 2. Response accuracy on the Conflict and Control problems under no load (NL), low load (LL), and high load (HL). Error bars are standard errors.

The response accuracy on the standard conflict problems in the no load condition (23%) is in line with previous studies (e.g. De Neys et al, 2013), demonstrating that performance was not generally influenced by the computer-based testing environment. Further analysis of erroneous responses showed that 95% (19/20) of biased reasoners provided the intuitive "10 cents" response in the absence of load. Importantly, this same high proportion of intuitively cued incorrect responses was observed under low (97%; 30/31) and high (93%; 27/29) loads, confirming that the presence of a demanding executive task does not simply lead to increased guessing or otherwise random responding[3].

It might be argued that the decrease in accuracy under load results not from increased demands on reasoning processes per se, but rather from the effect of load on reading or comprehension process, which are also known to draw on executive resources (Just, Carpenter, & Keller, 1996). This interpretation is ruled out, however, by the uniformly high accuracy on the no-conflict problems

[3] To be clear, the overall number of erroneous responses increased under load. The point here is simply that in all conditions the erroneous responses are not random but specifically entail the postulated intuitively cued "10 cents" response.

regardless of executive burden. This establishes that the observed effect of load on the standard bat-and-ball problem resulted specifically from the deployment of executive resources required to overcome the strongly compelling intuition.

General Discussion

The bat-and-ball problem has become a centerpiece for dual process theories of thinking (Bourgeois-Gironde & Vanderhenst, 2009; Kahneman, 2011; Thompson, 2009). The principle assumption is that an intuitive but incorrect answer (10 cents) is delivered via System 1 processing, with the intervention of System 2 required to override the erroneous intuition with correct reasoning. According to dual process theory, the quintessential difference between these two types of processing is the differential involvement of executive working memory (Evans, 2008; Evans & Stanovich, 2013; Stanovich, 2000). It is therefore essential for this theory to demonstrate that correctly solving the bat-and-ball problem does indeed draw on this limited executive resource pool.

Results of the reported experiment validate this dual-process assumption by experimentally manipulating the availability of executive resources while participants solved the bat-and-ball problem. When the intuitive response corresponded to the correct response in the control problem, accuracy was nearly perfect. Importantly, performance remained at ceiling under both low and high load conditions, confirming the basic assumption that the fast-acting System 1 intuition is triggered independently of executive involvement. In sharp contrast, a clear decline in accuracy was observed with increasing load on the standard problem which required the override of the erroneously cued response. This is all the more remarkable given that performance in the absence of load is already so low. When executive resources were burdened with the secondary task of maintaining and recalling a complex dot pattern, only a single person out of 30 was able to correctly solve the bat-and-ball problem. This pattern of results empirically confirms the direct involvement of executive working memory in the bat-and-ball problem, a suggestion supported previously by correlational evidence only.

Although dual process theories continue to gain in popularity across the social and psychological sciences, they have also received their fair share of criticism (see Evans & Stanovich, 2013). Many of these views offer unimodal or alternative approaches that may be more appealing for a variety of reasons. To be clear, the demonstrated load effect on the bat-and-ball problem does not in and of itself imply that dual process theories are indeed correct. Unimodal or other pluralistic models might also entail that the CRT draws on executive resources, however, the specific involvement of executive resources on conflict problems is a clear, a priori assumption of the general dual process model. Our results directly show that this assumption is warranted. In brief, the present study validates a key assumption of dual process theories, however, it clearly does not by itself argue against alternative frameworks.

Our findings validate the involvement of executive resources on the bat-and-ball problem, but they also raise additional questions regarding the nature of this involvement on these conflict problems. That is, it is not clear which component(s) of the bat-and-ball problem-solving process demand executive resources (see De Neys & Bonnefon, 2013). Recent work by De Neys, Rossi, and Houdé (2013) established that, in the absence of a secondary load, even incorrect reasoners were not completely oblivious to their erroneous response. In that study, reasoners who went with their intuition on the bat-and-ball problem expressed less confidence in their invalid response relative to confidence expressed on the no-conflict control version. The fact that even incorrect responders demonstrated some awareness that their response was not fully warranted suggests that conflict *detection* is not the primary difficulty of the problem. Other dual-task studies have indeed shown that conflict detection can operate independently of executive resources on a variety of judgment and reasoning tasks (e.g. De Neys, 2006; Franssens & De Neys, 2009). Nevertheless, this does not in and of itself establish that executive resources are not implicated in the detection of conflict on the bat-and-ball problem.

Of course even if reasoners detect the conflict in the problem and are in principle aware that the intuitive response must be inhibited/overridden, this does not imply that they still have the relevant knowledge or capacity to derive the correct response. While it is probably safe to assume that all university students can easily add 110 + 10, that does not entail that they can all set up the appropriate algebraic representation ($x + (x + 1) = 1.10$) from a reading of the bat-and-ball problem. Even if this knowledge is generally present in a student population, the additional cognitive load may impinge on the accessibility of this knowledge. An alternative strategy for solving this problem may be trial-and-error, adjusting the intuitive response (10 cents) little by little until an acceptable answer is found. This method of adjusting and checking also likely draws on executive resources and could account for the reduced performance observed under high load.

In short, the load effect observed in the present study could result from an additional burden on the retrieval of relevant knowledge, the conflict detection process, or the ability to inhibit/override the intuitive response and supplant it with a correct one. While this refinement remains for future research, the present study establishes that at whatever stage the trouble lies, successfully overcoming this difficulty depends on executive working memory resources. This executive involvement in the face of conflict, along with the automaticity of the intuitive control response, validates a key assumption of the dual process framework.

Acknowledgements

This research was supported by grants from the Generalitat de Catalunya (FI-DGR 2011), the Spanish Ministry of Economics and Competitiveness (PSI2012-35703), and the Agence Nationale de la Recherche (ANR-JSH2-0007-01).

References

Aguilar, D., Johnson, E.D. & Tubau, E. (2014). Reflection on explanation but not on practice facilitates overcoming erroneous probabilistic intuitions. Manuscript submitted for publication.

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.

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(1), 20–33.

De Neys, W. (2006a). Dual processing in reasoning: Two systems but one reasoner. *Psychological Science*, 17, 428–433.

De Neys, W. (2006b). Automatic-heuristic and executive-analytic processing in reasoning: Chronometric and dual task considerations. *Quarterly Journal of Experimental Psychology*, 59, 1070–1100.

De Neys, W., & Bonnefon, J. F. (2013). The whys and whens of individual differences in thinking biases. *Trends in Cognitive Sciences*, 17, 172-178.

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.

De Neys, W. & Verschueren, N. (2006). Working memory capacity and a notorious brain teaser: The case of the Monty Hall Dilemma. *Experimental Psychology*, 53, 123-131

Evans, J. St. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, 59, 255–278.

Evans, J. St. B. T., & Stanovich, K. (2013). Dual process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, 8, 223–241.

Franssens, S., & De Neys, W. (2009). The effortless nature of conflict detection during thinking. *Thinking & Reasoning*, 15, 105–128.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19, 25-42.

Gigerenzer, G. (2007). *Gut Feelings: The Intelligence of the Unconscious*. New York: Viking.

Gigerenzer, G., & Regier, T. (1996). How do we tell an association from a rule? Comment on Sloman (1996). *Psychological Bulletin*, 119, 23-26.

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.

Keren, G. & Schul, Y. (2009). Two Is Not Always Better Than One. *Perspectives on Psychological Science*, 4(6), 533-550.

Lesage, E., Navarrete, G., & De Neys, W. (2013). Evolutionary modules and Bayesian facilitation: The role of general cognitive resources. *Thinking and Reasoning*, 19(1), 27–53.

Liberali, J.M., Reyna, V.F., Furlan, S., Stein, L.M. & Pardo, S.T. (2011). Individual Differences in Numeracy and Cognitive Reflection, with Implications for Biases and Fallacies in Probability Judgment. *Journal of Behavioral Decision Making*. DOI: 10.1002/bdm.752

Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General*, 130, 621–640.

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

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, 23, 645–726.

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 and Cognition*, 39, 1275-1289.

Toplak, M.E., West, R. F. & Stanovich, K. E. (2013). Assessing miserly information processing: An expansion of the Cognitive Reflection Test. *Thinking & Reasoning*, DOI: 10.1080/13546783.2013.844729.

Thompson, V. A. (2009). Dual process theories: A metacognitive perspective. In J. Evans & K. Frankish (Eds.), *In Two Minds: Dual Processes and Beyond*. Oxford, UK: Oxford University Press.

Appendix

Standard versions

A bat and a ball together cost $1.10 . The bat costs $1 more than the ball. How much does the ball cost? ___ cents
[correct = .05 cents; intuitive = .10 cents]

A bat and a ball together cost $2.90. The bat costs $2 more than the ball. How much does the ball cost? ___ cents
[correct = .45 cents; intuitive = .90 cents]

A magazine and a banana together cost $1.10. The magazine costs $1 more than the ball. How much does the banana cost? ___ cents
[correct = .05 cents; intuitive = .10 cents]

A magazine and a banana together cost $2.90. The magazine costs $2 more than the ball. How much does the banana cost? ___ cents
[correct = .45 cents; intuitive = .90 cents]

Control versions

A magazine and a banana together cost $2.90. The magazine costs $2. How much does the banana cost?
[correct & intuitive = .90 cents]

A magazine and a banana together cost $1.10. The magazine costs $1. How much does the banana cost?
[correct & intuitive = .10 cents]

A bat and a ball together cost $2.90. The bat costs $2. How much does the ball cost?
[correct & intuitive = .90 cents]

A bat and a ball together cost $1.10. The bat costs $1. How much does the ball cost?
[correct & intuitive = .10 cents]