Smarter and Richer?: Executive Processing and the Monty Hall Dilemma

Wim De Neys (Wim.Deneys@psy.kuleuven.ac.be)
Department of Psychology, K.U.Leuven, Tiensestraat 102
B-3000 Leuven, Belgium

Abstract

The Monty Hall Dilemma (MHD) is a striking example of the human tendency to base probability judgment on intuitive, erroneous heuristics instead of an analytic, normative reasoning process. Two experiments tested the claim (e.g., Stanovich & West, 2000) that correct, normative reasoning draws on executive, working memory resources (WM) whereas heuristic reasoning would be purely automatic. Experiment 1A examined the link between MHD-reasoning and WM-capacity. Participants that solved the MHD correctly had a significantly higher WM-capacity. Experiment 1B presents a new approach to test the role of the WM-resources experimentally. Participants solved the MHD while WM-resources were burdened by a secondary task. Correct responses decreased under load. The results provide new evidence for the differential role of executive resources in heuristic and analytic reasoning.

Introduction

A main theme of cognitive reasoning research over the last decades is that human judgment frequently violates traditional normative standards. In a wide range of reasoning tasks most people do not give the answer that is correct according to logic or probability theory. The discrepancy between normative models and peoples actual performance has been labeled the "normative/descriptive gap" (Stanovich, 1999). The present study focuses on one of the most striking examples of this discrepancy: The Monty Hall Dilemma.

The notorious, counterintuitive Monty Hall Dilemma was adapted from a popular TV game show (Friedman, 1998). Host Monty Hall asks his final guest to choose one of three doors. One of the doors conceals a valuable prize and the other two contain worthless prizes such as goats or a bunch of toilet paper. After the guest makes a selection, the host, who knows where the prize is, opens one of the non-chosen doors to show that it contains a dud. The guests are then asked if they want to stay with their first choice or switch to the other unopened door.

Most people have the strong intuition that whether they switch or not the probability of winning remains 50% either way. However, from a normative point of view, the best strategy is to switch to the other door. Indeed, switching yields a 2/3 chance of winning. The solution hinges on the crucial fact that the host will never open the door concealing the prize, and obviously, he will not open the guest's door either. Taking into account that two thirds of the times the prize will be in one of the non-chosen doors, the non-chosen door that is still closed will hide the prize in two thirds of the trials (Tubau & Alonso, 2003).

Empirical studies of the Monty Hall Dilemma consistently showed that the vast majority of college students fails to give the correct response (switching rates ranging from 9% to 21%, e.g., Burns & Wieth, 2000, 2003; Friedman, 1998; Granberg & Brown, 1995; Krauss & Wang, 2003; Tubau & Alonso, 2003). Likewise, vos Savant (1997) reports that after she discussed the problem in a weekly magazine column she received up to 10,000 letters in response. Ninety-two percent of the writers from the general public disagreed with the switching answer. To paraphrase Friedman (1998), it seems that because of peoples poor MHD reasoning "millions of dollars were left on Monty's table".

Research indicates that the typical MHD response can be attributed to the operation of erroneous but very powerful intuitions or heuristics. For example, Shimojo and Ichikawa (1989) found that most people base their answer on the so called number-of-cases heuristic ("if the number of alternatives is N, then the probability of each one is 1/N"). Thus, since only two doors remain people will automatically assign a 50% chance to each door and fail to take the "knowledgeable host" information into account. Similar claims can be found in Falk (1992) and Johnson-Laird, Legrenzi, Girotto, Legrenzi, and Caverni (1999).

It has been argued that human thinking in general typically relies on the operation of intuitive, prepotent heuristics instead of a deliberate, controlled reasoning process. The primacy of these heuristics has been called the fundamental computational bias in human cognition (Stanovich, 1999). Whereas the fast and undemanding heuristics provide us with useful responses in many situations they can bias reasoning in tasks that require more elaborate, analytic processing (e.g., Evans & Over, 1996; Kahneman, Slovic, & Tversky, 1982;. Sloman, 1996; Stanovich, 1999; Stanovich & West, 2000; Tversky & Kahneman, 1983).

Stanovich and West (e.g., 2000) stressed that although the modal response is often erroneous in many reasoning tasks, a small proportion of the participants does give responses that are in line with the normative standards. Their research on individual differences showed that participants that gave the normative response on classic reasoning tasks such as the conjunction fallacy (Tversky & Kahneman, 1983) and the Wason (1966) selection task were disproportionally those highest in cognitive (working memory) capacity. According to Stanovich and West's dual process framework (see also Evans & Over, 1996; Sloman, 1996) correct normative responding requires that an analytic, controlled reasoning process overrides the prepotent heuristics. The inhibition of the heuristic system and the computations of the analytic system would draw on limited, executive working memory resources. The more resources that are available, the more likely that the analytic system will be successfully engaged and the correct response calculated.

The Stanovich and West (2000) findings suggest that a possible antidote to erroneous MHD reasoning might be a high working memory span. If correct normative reasoning requires executive working memory (WM) resources, then participants with a higher WM-span should be more likely to select the switching response. Bluntly put, the guests that win the prize in the game show will not only be richer but also "smarter"[1]. The link between MHD-reasoning and WM-capacity was examined in Experiment 1A.

The Stanovich and West framework and related dual process theories have been severely criticized (e.g., Stanovich & West, 2000). One important issue concerns the central assumption about the role of controlled, executive resources. Both the claim that correct, normative reasoning depends on the executive system and the characterization of the heuristic system as automatic and independent from executive control have been questioned (e.g., Handley, Feeney, Harper, 2002; Klaczynski, 2000, 2001; Osman, 2002)

Experiment 1B presents a new approach in the dual process field. The experiment adopted secondary task methodology to burden the executive WM-resources while participants were solving the MHD. If correct responding in the MHD draws on WM-resources, performance should decrease under load since less resources will be available for inhibition of the prepotent "50%-heuristic" and subsequent analytic computations. On the other hand, if the heuristic processing would not be automatic and would draw on WM, it will also become harder for people to come up with the "equal probability" answer. The procedure thereby allows a direct, experimental test of the basic executive processing assumptions.

[1] The term « smarter » refers of course to the tight connection between executive WM-capacity and general cognitive ability (e.g., Engle, Tuholski, Laughlin, & Conway, 1999).

Experiment 1A

Method

Participants

A total of 236 first-year psychology students from the Unversity of Leuven, Belgium, participated in return for psychology course credit.

Material

Working memory measure. Participants' working memory capacity was measured using a version of the Operation Span task (Ospan, La Pointe & Engle, 1990) adapted for group testing (Gospan, for details see De Neys, d'Ydewalle, Schaeken, & Vos, 2002). In the Ospan-task participants solve series of simple mathematical operations while attempting to remember a list of unrelated words The main adaptation in the Gospan is that the operation from an operation-word pair is first presented separately on screen (e.g., 'IS (4/2) – 1 = 5 ?'). Participants read the operation silently and press a key to indicate whether the answer is correct or not. Responses and response latencies are recorded. After the participant has typed down the response, the corresponding word (e.g., 'BALL') from the operation-word string is presented for 800 ms. As in the standard Ospan three sets of each length (from two to six operation-word pairs) are tested and set size varies in the same randomly chosen order for each participant. The Gospan-score is the sum of the recalled words for all sets recalled completely and in correct order.

Participants who made more than 15% math errors or whose mean operation response latencies deviated by more than 2.5 standard deviations of the sample mean were discarded (participants already in the bottom quartile of the Gospan-score distribution were not discarded based on the latency criterion). De Neys et al. (2002) reported an internal reliability coefficient alpha of .74 for the Gospan. The corrected correlation between standard Ospan and Gospan-score reached .70.

Monty Hall Dilemma. Participants were presented a standard version of the MHD taken from Krauss and Wang (2003). The formulation tried to avoid possible ambiguities (e.g., the random placement of the prize and duds behind the doors and the knowledge of the host were explicitly mentioned). The text stated (translated from Dutch):

Suppose you're on a game show and you're given the choice of three doors. Behind one door *is the main prize (a car) and behind the other two doors there are dud prizes (a bunch of toilet paper). The car and the dud prizes are placed randomly behind the doors before the show. The rules of the game are as follows: After you have chosen a door, the door remains closed for the time being. The game show host, Monty Hall, who knows what is behind the doors, then opens one of the two remaining doors which always reveals a dud. After he has opened one of the doors with a dud, Monty Hall asks the participant whether he/she wants to stay with his/her first choice or to switch to the last remaining door. Suppose that you chose door 1 and the host opens door 3, which has a dud.*

<u>The host now asks you whether you want to switch to door 2. What should you do to have most chance of winning the main prize?</u>

<u>a. Stick with your first choice, door 1.</u>
<u>b. Switch to door 2.</u>
<u>c. It does not matter. Chances are even.</u>

The MHD was presented on computer. Participants were instructed to carefully read the basic problem information (text in italics), first. When they were finished reading they pressed the ENTER-key and then the question and answer-alternatives (underlined text) appeared on the screen (text in italics remained on the screen). Participants typed their response (a, b, or c) on the keyboard. Instructions stated there were no time limits.

Procedure
The experiment was run on computer. Participants were tested in groups of 21 to 48. Participants completed the Gospan and MHD in a one-hour session, in which they also completed some other tasks not part of the present investigation. The MHD was presented after the Gospan task.

Results and discussion

Six participants were discarded because they did not meet the operation correctness or latency requirements of the WM-measure. Mean Gospan-score of the remaining 230 participants was 32.26 ($SD = 10.45$).

Consistent with previous MHD-studies only a small minority of the participants (5.2%) gave the correct switching answer. The vast majority (85.7%) believed that switching and sticking were equally good strategies. However, the crucial finding is that the participants that did give the correct response had a significantly larger WM-capacity. Mean Gospan-score of the participants that gave the correct response was 38.08 vs. only 31.94 for the incorrect responders, $t(228) = 2$, $n1 = 12$, $n2 = 218$, $p < .05$. In terms of effect sizes, Cohen's d reached .59. Such an effect is classified as "moderate" (Rosenthal & Rosnow, 1991) and corresponds to the effect sizes reported by Stanovich and West (1998a, 1998b) for the impact of executive capacity on the reasoning tasks in their studies.

The present association between MHD-performance and WM-capacity supports Stanovich and West's basic claim concerning the involvement of executive resources in normative reasoning. However, the evidence remains purely correlational. More direct evidence for the mediating role of the executive resources is needed (Klaczynski, 2000). Experiment 1B introduces secondary task methodology to test the basic processing claims experimentally.

Experiment 1B

A major problem for Stanovich and West (2000) and related dual processing frameworks is that the basic processing assumption, the different involvement of controlled, executive resources in heuristic and analytic reasoning, is disputed. On one hand, available (correlational) evidence for the role of executive resources in analytic, normative reasoning has been questioned (e.g., Klaczynski, 2000). On the other hand, the proposed characterization of the heuristic system as automatic and independent from executive control has been challenged (e.g., Handley, Feeney, Harper, 2002; Klaczynski, 2001; Osman, 2002). Experiment 1B presents a new approach to test the basic processing claims.

Participants solved the MHD while they performed a secondary task that burdened the executive WM-resources. If correct responding in the MHD draws on WM-resources, performance should decrease under load since less resources will be available for inhibition of the dominant "50%-heuristic" and subsequent analytic computations. On the other hand, if heuristic processing would not be automatic and would draw on WM, it will also become harder for people to come up with the "equal probability" answer and we would expect a decrease in "50%" responses. In case both the heuristic and normative response would draw on executive resources the computation of any single response should be hindered and we might expect a random guessing pattern under secondary task load.

The secondary task was adopted from Kane and Engle (2000). Participants were requested to continuously tap a novel, complex finger pattern (e.g., index finger/ring finger/ middle finger/ pinkie) with their non-dominant hand while reasoning. The task was selected because previous studies (e.g., Kane & Engle, 2000; Moscovitch, 1994) consistently showed that it put a premium on efficient executive WM-functioning.

Method

Participants
Forty-one first-year psychology students from the Unversity of Leuven, Belgium, participated in return for psychology course credit. None of the participants had participated in Experiment 1A. All participants had taken the Gospan-test prior to Experiment 1B. Between 13 to 46 days intervened between participation in the Gospan-test and Experiment 1B.

Materials
Monty Hall Dilemma. Participants were presented the same version of the MHD as in Experiment 1A.

WM-load task. A program executed by a second computer collected the finger-tapping data. All participants tapped on the "V", "B", "N", and "M" keys on the QUERTY-keyboard of the second computer.

Procedure
All participants were tested individually. Participants were instructed to tap the index-ring-middle-pinkie pattern with their non-dominant hand. The experiment started with five 30s practice tapping trials. Participants always received on-line accuracy feedback: Whenever a wrong finger (key) was tapped the computer emitted a 300 ms, low pitch tone. During the first three practice trials the program also calculated the mean tapping speed for each participant. If any one intertap interval in the subsequent trials was more than 150 ms slower than the established mean, the computer emitted a 600 ms, high pitch tone. The online monitoring served to assure that the tapping task was properly performed.

After the tapping practice, the experimenter explained that the practice tapping speed had to be maintained in the upcoming reasoning task. Participants then read the basic MHD problem information (underlined text) on the screen. When they were finished reading they pressed the ENTER-key and started tapping. Then the question and answer-alternatives (text in bold) were presented and participants continuously tapped the finger pattern (with online response time and accuracy feedback) until they gave their response. Participants said out loud the letter (a, b, or c) corresponding to their answer. Instructions stated there were no time limits.

Results and discussion

Performance of the participants in Experiment 1A was used as a baseline to evaluate the impact of the WM-load. A control analysis established that the WM-capacity of the participants in Experiment 1A (Mean Gospan-score = 32.26, SD = 10.45) and 1B (Mean Gospan-score = 33.1, SD = 10.88) did not differ, $F(1, 269) < 1$.

Burdening the executive resources with the tapping task affected participants' performance. As Table 1 shows the response pattern was clearly not random. The switching rate under the secondary task load decreased to 0%. This decrease in the proportion of correct responses reached marginal significance, $p1 = 5.22\%$, $p2 = 0\%$, $n1 = 230$, $n2 = 41$, $t(269) = 1.50$, $p < .07$, one-tailed. The finding supports the claim that correct normative reasoning in the MHD draws on executive WM-resources. Burdening the executive resources did not decrease the rate of "equal probability" answers. Indeed, there was a slight tendency in the opposite direction. This suggests that the central MHD intuition to assign a 50% chance to the two remaining doors is an automatic, heuristic response that does not

Table 1: Percentage of Different Responses in Experiment 1A and 1B

	Experiment	
Answer	1A: No load	1B: Load
Stick	9.1 (21)	7.3 (3)
Switch	5.2 (12)	0.0 (0)
Equal	85.7 (197)	92.7 (38)

Note. Raw frequencies in parentheses.

involve executive processing.

General Discussion

The present study focused on the Monty Hall Dilemma because it is one of the most striking examples of the "normative/descriptive" gap in the literature (Friedman, 1998). As in previous MHD studies only a small proportion of participants gave the correct, normative switching response. However, Experiment 1A established that the participants that did give the correct response had a significantly larger working memory capacity. This finding complements the work of Stanovich and West (2000) on individual differences in executive resources with related reasoning tasks.

According to the Stanovich and West (2000) framework and associated dual process theories (e.g., Evans & Over, 1996; Sloman, 1996) correct normative responding requires that an analytic, controlled reasoning process overrides prepotent heuristics. The inhibition of the heuristic system and the computations of the analytic system would draw on limited executive, working memory resources. The more resources that are available, the more likely that the analytic system will be successfully engaged and the correct response calculated. Experiment 1B provided experimental evidence for this view. Burdening the executive resources with a secondary task while participants were solving the MHD tended to decrease the rate of correct, switching responses: Although the participants in Experiment 1A and 1B had comparable span sizes, non of the participants in Experiment 1B managed to solve the MHD correctly under WM-load. Indeed, more people tended to commit the intuitive tendency to assign a 50% chance to the two remaining doors. These findings support the basic claim of dual process theories concerning the differential involvement of executive resources in analytic and heuristic reasoning.

As in most MHD-studies, the proportion of correct responses under "standard" conditions in the present study was very low. A consequence of this low figure is that the study inevitably suffers from a floor-effect. The decrease in correct performance under executive load in Experiment 1B still reached marginal significance but the decrease could never be large. One possible solution for the floor-effect is adopting some of the manipulations known to increase MHD performance. Previous studies indicated that practice with the task, training procedures and simple clarifications of the causal structure of the task (e.g., Burns & Wieth, 2000, 2003; Krauss & Wang, 2003; Tubau & Alonso, 2003) increase performance. Thus, testing the impact of the WM-load with such modified MHD versions might suffer less from a floor-effect. In addition, it might be especially enlightening to examine how different span groups benefit from the increased performance manipulations.

The present findings indicate that executive resources are *necessary* for correct, normative reasoning. However, by no means this implies that a large resource pool is also *sufficient* for correct reasoning. The relation between WM-capacity and reasoning performance is not absolute. Although participants that solved the MHD correctly in Experiment 1A had a larger WM-span, numerous "high spans" nevertheless answered erroneously. To illustrate this point MHD-performance of participants in the top and bottom quartile of the WM-capacity distribution in Experiment 1A was compared. Consistent with the previous findings high spans gave significantly more switching responses[2]. But even among the 25% most cognitively gifted college students only 10% gave the correct response. Clearly, factors outside the cognitive WM-ability spectrum will also affect performance (e.g., "epistemic thinking dispositions", see Stanovich, 1999). Thus, in pointing out the necessary role of executive, WM-resources for correct reasoning the present study does not minimize the role of other mediating factors.

The present study demonstrated the potential of a dual task approach to test the central processing claims of dual process theories. In principle, future studies could adopt this procedure with all the classic tasks studied in the field (e.g., conjunction fallacy, base-rate neglect, selection task). Of course, the final empirical evaluation of the executive processing claims of dual process theories will depend on the generalization of the present findings.

[2] $p1 = 10\%$, $p2 = 1\%$, $n1 = 70$, $n2 = 70$, $t(138) = 2.21$, $p < .03$.

Acknowledgements

Preparation of this manuscript was supported by a grant from the Fund for Scientific Research-Flanders (FWO).

References

Burns, B. D., & Wieth, M. (2000). *The Monty Hall dilemma: A causal explanation of a cognitive illusion*. Paper presented at the Forty-First Annual Meeting of the Psychonomic Society, New Orleans, LA.

Burns, B. D., & Wieth, M. (2003). *Causality and Reasoning: The Monty Hall dilemma*. Paper presented at the Twenty-Fifth Annual Meeting of the Cognitive Science Society, Boston, MA.

De Neys, W., d'Ydewalle, G., Schaeken, W., & Vos, G. (2002). A Dutch, computerized, and group administrable adaptation of the operation span test. *Psychologica Belgica, 42,* 177-190.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General, 128*, 309-331.

Evans, J. St. B. T., & Over, D. E. (1996). *Rationality and reasoning*. Hove, UK: Psychology Press.

Falk, R. (1992). A closer look at the probabilities of the notorious three prisoners. *Cognition, 77*, 197-213.

Friedman, D. (1998). Monty Hall's three doors: Construction and deconstruction of a choice anomaly. *American Economic Review, 88*, 933-946.

Granberg, D., & Brown, T. A. (1995). The Monty Hall dilemma. *Personality and Social Psychology Bulletin, 21*, 182-190.

Handley, S. J., Feeney, A., & Harper, C. (2002). Alternative antecedents, probabilities, and the suppression of fallacies in Wason's selection task. *Quarterly Journal of Experimental Psychology, 55A*, 799-818.

Johnson-Laird, P. N., Legrenzi, P., Girotto, P., Legrenzi, M. S., & Caverni, J-P. (1999). Naïve probability: A mental model theory of extensional reasoning. *Psychological Review, 106*, 62-88.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgement under uncertainty: Heuristics and* *biases*. Cambridge, MA: Cambridge University Press.

Kane, M. J., & Engle, R. W. (2000). Working-memory capacity, proactive interference, and divided attention: Limits on long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 336-358.

Klaczynski, P. A. (2000). Is rationality really "bounded" by information processing constraints? *Behavioral and Brain Sciences, 23*, 683-684.

Klaczynski, P. A. (2001). Framing effects on adolescent task representation, analytic and heuristic processing, and decision making: Implications for the normative/descriptive gap. *Applied Developmental Psychology, 22*, 289-309.

Krauss, S., & Wang, X. T. (2003). The psychology of the Monty Hall Problem: Discovering psychological mechanisms for solving a tenacious brain teaser. *Journal of Experimental Psychology: General, 132*, 3-22.

La Pointe, L. B., & Engle, R. W. (1990). Simple and complex word spans as measures of working memory capacity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 1118-1133.

Moscovitch, M. (1994). Cognitive resources and dual-task interference effects at retrieval in normal people: The role of the frontal lobes and medial temporal cortex. *Neuropsychology, 8,* 524-534.

Osman, M. (2002). *Is there evidence for unconscious reasoning processes?*. Paper presented at the Twenty-fourth Annual Meeting of the Cognitive Science Society, Fairfax, VA.

Rosenthal, R., & Rosnow, R. L. (1991). *Essentials of behavioral research: Methods and data analysis* (2nd Edition). New York: McGraw-Hill.

Shimojo, S., & Ichikawa, S. (1989). Intuitive reasoning about probability: Theoretical and experimental analysis of the "problem of three prisoners". *Cognition, 32*, 1-24.

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

Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning*. Mahwah, NJ: Erlbaum.

Stanovich, K. E., & West, R. F. (1998a). Individual differences in framing and conjunction effects. *Thinking and Reasoning, 4*, 289-317.

Stanovich, K. E., & West, R. F. (1998b). Cognitive ability and variation in selection task performance. *Thinking and Reasoning, 4*, 193-230.

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

Tubau, E., & Alonso, D. (2003). Overcoming illusory inferences in a probabilistic counterintuitive problem : The role of explicit representations. *Memory & Cognition, 31*, 596-607.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review, 90,* 293-315.

vos Savant, M. (1997). *The power of logical thinking*. New York: St. Martin's Press.

Wason, P. C. (1966). Reasoning. In B. M. Foss (Ed.), *New Horizons in Psychology: I* (pp. 106-137). Harmandsworth, England: Penguin.