Working Memory Capacity and a Notorious Brain Teaser

The Case of the Monty Hall Dilemma

Wim De Neys and Niki Verschueren

University of Leuven, Belgium

Abstract. The Monty Hall Dilemma (MHD) is an intriguing example of the discrepancy between people's intuitions and normative reasoning. This study examines whether the notorious difficulty of the MHD is associated with limitations in working memory resources. Experiment 1 and 2 examined the link between MHD reasoning and working memory capacity. Experiment 3 tested the role of working memory experimentally by burdening the executive resources with a secondary task. Results showed that participants who solved the MHD correctly had a significantly higher working memory capacity than erroneous responders. Correct responding also decreased under secondary task load. Findings indicate that working memory capacity plays a key role in overcoming salient intuitions and selecting the correct switching response during MHD reasoning.

Keywords: reasoning, individual differences, working memory, Monty Hall Dilemma

The notorious, brain teasing Monty Hall Dilemma (MHD) was adapted from a popular TV game show (Friedman, 1998). At the end of the show, the 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, Monty Hall, who knows where the prize is, opens one of the nonchosen 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. 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. In one third of the trials, the guest will initially select the correct door. In this case, it would be better not to switch. However, in the other two thirds of the cases the nonchosen closed door will hide the prize and switching is advantageous. Hence, switching yields a ⅔ chance of winning.

Empirical studies of the MHD 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, 2004; 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 people's 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 (see Falk, 1992, for a similar claim). Furthermore, some people are simply biased by the general belief that when making a decision one should always stick to one's first choice (e.g., a bias long noted in responses to multiple-choice exams, Geiger, 1997). Gilovich, Medvec, and Chen (1995) clarified that this "stick with your pick" intuition would be based on an anticipation of regret.

DOI: 10.1027/1618-3169.53.2.123

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. Over the last decades, reasoning research indeed showed that in a wide range of reasoning tasks most people do not give the answer that is correct according to logic or probability theory (e.g., Evans & Over, 1996; Kahneman, Slovic, & Tversky, 1982). The primacy of the heuristics has been called the fundamental computational bias in human cognition (Stanovich, 1999). Although the fast and undemanding heuristics can provide us with useful responses in many daily situations, they can bias reasoning in tasks that require more elaborate, analytic processing (e.g., Sloman, 1996; 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 who 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 Epstein, 1994; Evans, 2003; 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 people's limited 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.

Previous MHD studies have tried to characterize erroneous MHD reasoning as a failure of representation (e.g., Krauss & Wang, 2003; Tubau & Alonso, 2004), faulty mental model construction (Johnson-Laird, Legrenzi, Girotto, Legrenzi, & Caverni, 1999), or as an instance of people's difficulty with grasping causal structures (Burns & Wieth, 2004). The present study approaches the MHD from the dual process perspective. The findings of Stanovich and West (2000) suggest that a possible antidote to erroneous MHD reasoning might be a high working memory span. If correct normative reasoning requires working memory resources, then participants with a higher span should be more likely to overcome the heuristic temptations and compute the correct switching response during MHD reasoning.

The possibility that limitations in cognitive capacity might be responsible for people's difficulty with solving the MHD has received little attention in the MHD literature. We suspect that one of the reasons is that some appealing and widely cited anecdotal evidence of ace mathematicians and Nobel Prize winners' erroneous responses seems to argue against the role of cognitive capacity. Burns and Wieth (2004), for example, cite Schechter (1998) who related how Paul Erdõs, one of the greatest mathematicians of the 20th century, failed to solve the MHD, and how it took a fellow mathematician several days to make him understand his error. Likewise, discussions of the MHD, from the front-page of the New York Times (Tierney, 1991) to the top scientific journals, typically stress Marylin vos Savant's infamous observation that numerous readers who disagreed with the switching response and furiously contested her were well-respected scholars. In his popular book *Inevitable Illusions: How Mistakes of Reason Rule Our Mind*, Piattelli-Palmarini (1994) consequently characterized the MHD as the most expressive example of a cognitive illusion in which "even the finest and best-trained minds get trapped" (p. 161).

Despite the appealing examples, the question concerning the role of cognitive resources in MHD reasoning remains an empirical one. It might be very tempting to conclude that cognitive capacity plays no special role in MHD reasoning when even the most gifted scientists fail to solve it. However, logically speaking such a conclusion is not warranted. Indeed, when the army's best trained marines unit would be killed in action we would neither conclude that we should stop training soldiers because it does not increase their survival rate. The death of the ace soldiers would merely suggests that the relation between training and combat survival is not perfect. Therefore, the present study will present an empirical test of the possible involvement of cognitive, working memory capacity in MHD reasoning.

Working memory is typically characterized as a system in which specific storage and maintenance components subserve a central, executive component responsible for the control of information processing (Baddeley & Hitch, 1974). The executive working memory (WM) component is widely recognized as the quintessential constituent of human cognitive capacity (e.g., Engle, Tuholski, Laughlin, & Conway, 1999; Süss, Oberauer, Wittman, Wilhelm, & Schulze, 2002). In three experiments, involving more than 400 participants, we demonstrate that contrary to the popular anecdotal view there is indeed clear empirical support for the role of WM capacity limitations in erroneous MHD reasoning.

Experiments

Experiment 1 and 2 examined the association between WM capacity and MHD performance. Participants were presented a version of the MHD and a measure of WM capacity. If erroneous MHD reasoning results from limitations in WM capacity, we expect that participants who select an erroneous response will have a lower WM capacity than participants who reason correctly. Experiment 3 tests the causal nature of the associations by limiting the available WM resources experimentally.

Experiment 1

Method

Participants

A total of 239 first-year psychology students from the University of Leuven, Belgium, participated in return for psychology course credit.

Material

Monty Hall Dilemma. Participants were presented a version of the MHD based on 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). As in Tubau and Alonso (2003), participants could choose between three answer alternatives (a. Stick – b. Switch – c. Chances are even). The complete problem format is presented in the Appendix.

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

Finally, participants were asked whether they had already read or heard about the problem. Three participants answered positively and were discarded.

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). Ospan is a classic working memory measure that primarily reflects central executive capacity (Engle et al., 1999). Participants solve series of simple mathematical operations while attempting to remember a list of unrelated words. First, an operation is presented 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, a word (e.g., "ball"') is presented for 800 ms. After a number (ranging from two to six) of operation-word pairs have been presented, participants are cued to recall the words. Hence, participants have to solve the operation and simultaneously attempt to remember the previously presented words. Handling such simultaneous information storage and processing is considered one of the key functions of the central executive. The higher one's executive resources, the more words will be correctly recalled.

In the Gospan, 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 (maximum score is 60). Participants who make more than 15 % math errors or whose mean operation response latencies deviate by more than 2.5 standard deviations of the sample mean are discarded. This procedure tries to make sure that participants did not decrease the task burden by simply neglecting the operation processing or extensive word rehearsal. In the present sample six participants did not meet the operation correctness or latency requirements. The mean Gospan score of the remaining 230 participants was 32.26 (SD = 10.45).

Procedure

Participants were tested in groups of 21 to 48 in a one-hour session where the MHD was presented after the Gospan task.

Results and Discussion

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. The remaining 9.1 % believed that sticking to the first choice would be beneficial. The crucial finding is that the participants who did give the correct response had a significantly larger WM capacity than participants who selected one of the erroneous responses. The mean Gospan score of the participants who gave the correct response was 38.08 ($SD = 9.43$) vs. only 31.94 ($SD = 10.43$) for the incorrect responders, $t(228) = 2$, $n1 = 12$, $n2 = 218$, $p < .025$, one-tailed (point biserial correlation coefficient $r = .13$, $n = 230$, $p < .025$, one-tailed). 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.

In the group of incorrect responders, there were no significant WM capacity differences between participants who selected the "stick with your pick" ($M = 31.24$, $SD = 9.42$) and "chances are equal" ($M = 32.01$, $SD = 10.56$) response.

The association between MHD performance and WM capacity supports the claim concerning the involvement of executive resources in correct MHD reasoning. However, because of the floored MHD performance the number of participants in the "correct" group was quite small. Experiment 2 attempted to replicate the findings with a task version that had been shown to boost performance without altering the basic task characteristics.

Experiment 2

Tubau and Alonso (2003) presented participants a modified MHD problem that involved a card game between two players. The goal of the game was to draw an ace from a pool of three cards. One player had to choose one of the three cards without seeing it. The remaining cards were for the opponent. Since the opponent had two cards one of the game's rules stated that he or she had to show one non-ace card to the other player. The player that had chosen the first card then could choose to stick to his initial selected card or to switch to the opponent's remaining card. Tubau and Alonso reasoned that the card game context would make it clearer that (since he or she had an additional card), the opponent would have a higher chance of getting the ace. Tubau and Alonso indeed observed that with this version about 20% of the participants selected the correct switching response. Participants in Experiment 2 received a similar MHD version based on the well-known "cups and ball" game.

Method

Participants

A total of 129 senior high school students (Mean age = 17.11, $SD = .57$) participated voluntarily.

Material

Monty Hall Dilemma. As in Tubau and Alonso (2003, Experiment 2) participants received a version of the MHD situated in a "game of chance" context. The present version was based on Tubau and Alonso but referred to the "cups and ball" game instead of a card game. The complete problem format is presented in the Appendix.

Working memory measure. See Experiment 1. Five participants were discarded because they did not meet the operation correctness or latency requirements of the WM measure. The mean Gospan-score of the remaining 124 participants was 36.23 ($SD = 12.13$).

Procedure

See Experiment 1.

Results and Discussion

Consistent with Tubau and Alonso (2003) about 23% of the participants gave the correct switching answer in the "game of chance" MHD. The vast majority (67%) believed that switching and sticking were equally good strategies and about 10% of the participants preferred to stick to the original choice. As in Experiment 1, we found that participants who did give the correct switching response had a significantly larger WM capacity. Mean Gospan score of correct responders was 39.97 ($SD = 13.23$) vs. only 35.08 ($SD = 11.61$) for the incorrect responders, $t(122) = 1.92$, $n1 = 29$, $n2 = 95$, $p < .03$, one-tailed (point biserial correlation coefficient $r = .17$, $n = 124$, $p < .03$, one-tailed). The effect size, $d = .41$, could also be classified as moderate.

The WM capacity of participants who selected the erroneous "sticking" ($M = 34.58$, $SD = 15.71$) or "chances are equal" ($M = 35.16$, $SD = 11.02$) response did not differ significantly.

Experiment 1 and 2 support the claim concerning the involvement of executive resources in correct MHD reasoning. Experiment 3 introduces secondary task methodology to test the basic processing claims experimentally.

Experiment 3

The rationale for the present study was based on Stanovich and West's influential research program on individual differences in reasoning. One fundamental limitation of the Stanovich and West studies, however, is that they remained purely correlational (see commentaries on Stanovich & West, 2000). The reported correlations do not establish the assumed causality: The findings indicate that selecting a correct, normative response is associated with having a larger executive resource pool, but this does not imply that the resources are necessary for the calculation of the correct response (e.g., Klaczynski, 2000; Newton & Roberts, 2003; Sternberg, 2000). Experiment 3 introduces a secondary task approach to test the claim experimentally. If correct responding in the MHD requires executive resources, performance should decrease under load since fewer resources will be available for the demanding computations.

Participants solved the MHD while they concurrently tried to remember a briefly presented complex dot pattern. Miyake, Friedman, Rettinger, Shah, and Hegarty (2001) established that this specific memorization task burdened the executive resources. The complexity of the dot pattern was manipulated so that storage of the pattern in a control group would be less demanding.

Method

Participants

Participants were 102 undergraduate students from the University of Leuven, Belgium, who participated in return for psychology course credit. None of the participants had participated in Experiment 1 or 2.

Materials

Monty Hall Dilemma. Participants received the "cups and ball" version presented in Experiment 2.

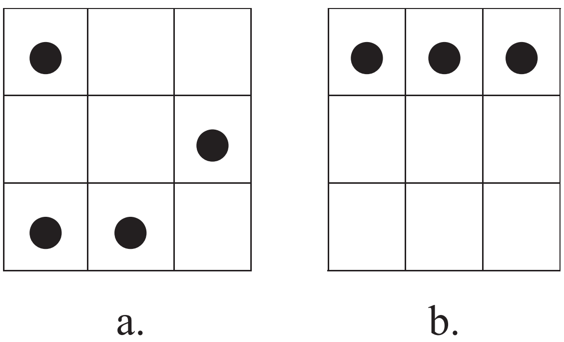

Figure 1. Examples of the dot patterns in the load (a) and control group (b).

Dot memory task. The dot memory task is a classic spatial storage task (e.g., Bethell-Fox & Shepard, 1988; Miyake et al., 2001). For the present study a 3×3 matrix filled with three to four dots was briefly presented for 850 ms. Participants memorized the pattern and were asked to reproduce it afterwards.

In the load group the matrix was filled with a complex 4-dot pattern as presented by Verschueren, Schaeken, and d'Ydewalle (2004, i.e., a "two- or three-piece" pattern based on the work of Bethell-Fox & Shepard, 1988, see Figure 1). Miyake et al. (2001) showed that storage of similar complex dot patterns tapped executive resources. They demonstrated that, in contrast with the verbal domain, temporary storage in the visuospatial domain requires central executive resources. The work of Miyake et al. suggested that in the visuospatial domain the temporary storage system in the original Baddeley and Hitch (1974) working memory framework is closely tied to and might be indistinguishable from the central executive.

In the control group the pattern consisted of three dots on a horizontal line (i.e., a "one-piece" pattern in Bethell-Fox & Shepard's terms). This simple and systematic pattern (Ichikawa, 1981; Miyake et al., 2001) should only place a minimal burden on the executive resources.

Procedure

Participants were tested in groups of 8 to 18 and were randomly assigned to the control or load group. The experiment started with a demonstration of the storage task. On two practice storage items (one with a simple and one with a complex pattern) an empty response matrix was presented 1 s after the pattern had been presented. Participants used the keypad to indicate the location of the dots. Instructions stressed that it was crucial that the dot pattern was reproduced correctly in the upcoming reasoning task.

As in the previous experiments, participants first read the preambles and hit the enter-key when finished. Next, the dot pattern was presented for 850 ms and subsequently the preambles were presented together with the answer-alternatives. Participants typed their response (a, b, or c) on the keyboard. Afterwards, the empty matrix was presented and participants had to reproduce the dot pattern.

Results and Discussion

Dot memory task. Results for the dot memory task indicated that the task was properly performed. The simple dot pattern in the control group was always perfectly recalled. The mean number of correctly localized dots for the complex dot pattern was 3.57 ($SD = .57$). Thus, overall, about 89% of a complex dot pattern was still reproduced correctly. Furthermore, the MHD performance under load was not negatively associated with the mean dot recall score, $r = .22$, $p = .12$. Hence, participants were not simply trading-off dot recall performance to solve the MHD.

MHD reasoning. Burdening the executive resources with the complex dot memory task clearly affected participants' performance. As Table 1 shows, the response pattern was not random: Both in the control and load group the dominant response was the "Chances are equal" answer. The crucial finding is that the switching rate decreased when the secondary task burdened the executive resources. As in Experiment 2, about 22% of the participants gave the correct response in the control group. The figure was more than halved in the load group, $n1 = 51$, $n2 = 51$, $t(100) = 1.98$, $p < .03$, one-tailed; $d = .39$ (point biserial correlation coefficient $r = .17$, $n = 102$, $p < .05$, one-tailed). Results were also replicated in a nonparametric analysis with the Fisher Exact Probability Test, $p < .05$, one-tailed.

The load findings support the postulated involvement of executive resource limitations in erroneous MHD reasoning. However, it will be clear that the present results should be interpreted with some caution. The experiment is the first to introduce the secondary task approach in the MHD field. Further experimentation will need to fine-tune the findings. One possible limitation of the present study is that we cannot guarantee that there were no a priori differences in WM capacity between the two groups. Results can be confounded when the participants in the control group would have a higher capacity than participants in the load group. Although such a confound is not likely given the sample size and random allocation of participants to groups, it cannot be ruled out completely. It would be advisable to measure participants' WM capacity first or to adopt a within-subjects design in future studies.

Table 1. Percentage of different responses in Experiment 3.

Answer	Group	
	Control	Load
Stick	5.9% (3)	13.7% (7)
Switch	21.6% (11)	7.8% (4)
Even	72.5% (37)	78.4% (40)

Note. Raw frequencies in parentheses.

Table 1 also suggests that the impact of the executive burden tended to be more pronounced on the "stick" than on the "equal" responses. Although the difference was not significant one possible reason for the trend could be that the two types of erroneous responses have somewhat different computational demands. One could suggest that whereas the number-of-cases heuristic is based on a cognitive probability estimation (albeit a simple one), the "stick with your pick" heuristic has a more basic, less demanding, emotional basis. Hence, an executive burden would specifically trigger the computationally less demanding "stick with your pick" intuition. With hindsight one might note that participants in Experiment 1 and 2 who selected the "sticking" response also tended to have a somewhat lower WM capacity than those who responded that the "chances are equal". However, given that none of the effects reached significance these claims remain speculative. A final remark in this respect is that the present study, as many individual differences studies, focused exclusively on the reasoning performance of educated, young adults. Consequently, the variation in WM capacity may have been restricted what may have blurred these trends somewhat. It is possible that in the population at large the two types of erroneous responses might be further differentiated.

General Discussion

In this study we presented an empirical test of the involvement of WM resources in MHD reasoning. Experiment 1 and 2 established that participants who managed to give the correct switching response had a significantly larger WM capacity than participants who reasoned erroneously. This finding supports the research program of Stanovich and West (e.g., 1998a, 1998b, 2000) on individual differences in reasoning. The presents MHD results indicate that one of the most notorious reasoning problems in the literature is no exception to the general rule they established: Although the modal response is often erroneous in many reasoning tasks, a small proportion of the participants does manage to respond correctly. These participants will be specifically those highest in WM capacity.

The correlational nature of the Stanovich and West studies has been severely criticized (e.g., Klaczynski, 2000; Newton & Roberts, 2003; Sternberg, 2000). Stanovich and West showed that selecting a correct, normative response is associated with having a larger resource pool, but this does not imply that the WM resources are necessary for the calculation of the correct response. Some other factor (e.g., education or motivation) might account for the positive association.

The introduction of a secondary task approach in the present study circumvented this critique. Experiment 3 showed that correct switching rates decreased when available executive resources were experimentally limited by an attention demanding secondary task. This finding suggests that limitations in executive resources play indeed a key role in erroneous MHD reasoning.

It should be clear that our study does not imply that the executive resource pool is the only factor affecting MHD reasoning. Obviously, the relation between WM capacity and reasoning performance is not absolute. For example, correct solution rates were never higher than 20% in our experiments. This already suggests that, just like Paul Erdõs, even people in the higher levels of the WM capacity distribution sometimes fail. Clearly, factors outside the cognitive ability spectrum can also affect performance (e.g., Stanovich, 1999). Hence, whereas the present focus on WM stresses the impact of limitations in executive resources it does not discard the role of other mediating factors.

The possible role of cognitive capacity limitations in erroneous MHD reasoning has been neglected in the MHD literature. We believe that the eagerly cited examples of ace mathematicians' MHD failures contributed to the implicit view that MHD reasoning would be "immune" to cognitive capacity. It might indeed be tempting to conclude that cognitive capacity plays no special role in MHD reasoning when even the most gifted scientists fail to solve it. However, as we noted in the introduction, logically speaking such a conclusion is not warranted: When the army's best trained marines unit would be killed in action we would neither conclude that we should stop training soldiers because it does not increase their survival rate. As argued above, the point is not that MHD reasoning is completely determined by WM capacity. The crucial stipulation we wanted to make is that despite some appealing examples, there is clear empirical support for the role of WM capacity limitations in erroneous MHD reasoning.

Previous MHD studies have proposed numerous psychological mechanisms behind the MHD. For example, Krauss and Wang (2003) and Tubau and Alonso (2003) have focused on the MHD as failure of representation. They showed that presenting the MHD in ways that help participants form the correct problem representation (e.g., by leading them to focus less on the two doors remaining) improved performance. Likewise, Johnson-Laird et al. (1999) argued that people fail to differentiate the options because they create the wrong set of mental models. Burns and Wieth (2004) argued that a clarification of the causal problem structure (i.e., a so-called collider principle in which two independent causal factors influence a single outcome) boosted performance. The present study approached the MHD from a dual process theory perspective (e.g., Epstein, 1994; Evans, 2003; Stanovich & West, 2000). This perspective hinges on the idea that the main problem in the MHD is that, as in many classic reasoning tasks, tempting heuristics impede correct reasoning. 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 people's limited WM resources. Consequently, in the MHD, participants with a higher WM capacity should be more likely to overcome the heuristic temptations and compute the correct switching response. Note that although the present findings support the dual process view they do not contradict the previously proposed mechanisms. In dual process terms these mechanisms might be considered to be more fine-grained specifications of the nature of the analytic reasoning process itself. The findings do imply, however, that whatever the specific proposed mechanism behind the MHD might be, its application will depend on the available WM resources. Taking the WM mediation into account might thereby help MHD researchers to link their work with more general research on the role of WM in higher-order cognition.

Acknowledgements

Preparation of this manuscript was supported by grants from the Fund for Scientific Research (FWO). Parts of this study were presented at the Twenty-Sixth Annual Conference of the Cognitive Science Society, Chicago, IL.

References

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. Bower (Ed.), *Recent advances in learning and motivation* (Vol. 8, pp. 47–90). New York: Academic Press.

Bethell-Fox, C. E., & Shepard, R. N. (1988). Mental rotation: Effects of stimulus complexity and familiarity. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 12–23.

Burns, B. D., & Wieth, M. (2004). The collider principle in causal reasoning: Why the monty hall dilemma is so hard. *Journal of Experimental Psychology: General, 133,* 434–449.

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 ap-

proach. *Journal of Experimental Psychology: General, 128,* 309–331.
Epstein, S. (1994). Integration of the cognitive and psychodynamic unconscious. *American Psychologist, 49,* 709–724.
Evans, J. St. B. T (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences, 7,* 454–459.
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.
Geiger, M. A. (1997). Educators' warnings about changing examination answers: Effects on students perceptions and performance. *College Student Journal, 31,* 429–432.
Gilovich, T., Medvec, V. H., & Chen, S. (1995). Commission, omission, and dissonance reduction: Coping with regret in the "Monty Hall" problem. *Personality and Social Psychology Bulletin, 21,* 185–190.
Granberg, D., & Brown, T. A. (1995). The Monty Hall dilemma. *Personality and Social Psychology Bulletin, 21,* 182–190.
Ichikawa, S. (1981). Rated ease of memorization of dot-in-matrix patterns: Multiple regression analysis by physical variables which describe configuration of dots. *Japanese Psychological Research, 25,* 173–180.
Johnson-Laird, P. N., Legrenzi, P., Girotto, P., Legrenzi, M. S., & Caverni, J-P (1999). Naı⁄ve 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.
Klaczynski, P. A. (2000). Is rationality really "bounded" by information processing constraints? *Behavioral and Brain Sciences, 23,* 683–684.
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.
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.
Newton, E. J., & Roberts, M. J. (2003). Individual differences transcend the rationality debate. *Behavioral and Brain Sciences, 26,* 530–531.
Piattelli-Palmarini, M. (1994). *Inevitable illusions: How mistakes of reason rule our minds*. New York: Wiley.
Rosenthal, R., & Rosnow, R. L. (1991). *Essentials of behavioral research: Methods and data analysis* (2nd Ed.). New York: McGraw-Hill.
Schechter, B. (1998). *My brain is open: The mathematical journeys of Paul Erdõs*. New York: Simon & Schuster.
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.
Sternberg, R. J. (2000). The ability is not general, and neither are the conclusions. *Behavioral and Brain Sciences, 23,* 697–698.
Süss, H., Oberauer, K., Wittmann, W. W., Wilhelm, O., & Schulze, R. (2002). Working-memory capacity explains reasoning ability and a little bit more. *Intelligence, 30,* 261–288.
Tierney, J. (1991, July 21). Behind Monty Hall's doors: Puzzle, debate and answer? New York Times, pp. 1, 20.
Toms, M., Morris, N., & Ward, D. (1993). Working memory and conditional reasoning. *Quarterly Journal of Experimental Psychology, 46A,* 679–699.
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.
Verschueren, N., Schaeken, W., & d'Ydewalle, G. (2004). Everyday conditional reasoning with working memory preload. *Proceedings of the Twenty-Sixth Annual Meeting of the Cognitive Science Society*. Mahwah: Erlbaum.
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, UK: Penguin.

Received April 6, 2005
Revision received May 30, 2005
Accepted May 31, 2005

Wim De Neys

Lab Experimentele Psychologie
K. U. Leuven
Tiensestraat 102
B-3000 Leuven
Belgium
Tel. +32 16 326 143
Fax +32 16 326 099
E-mail Wim.Deneys@psy.kuleuven.be

Appendix

MHD version Experiment 1

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.

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?

a. Stick with your first choice, door 1.

b. Switch to door 2.

c. It does not matter. Chances are even.

MHD version Experiment 2 and 3

Two players, a boy and a girl, play a game. Three cups are placed on a table. One cup hides a marble whereas the other two hide nothing. The boy and the girl do not know which cup hides the marble. The rules of the game are as follows: First, the boy picks one cup randomly. His opponent, the girl, gets the remaining two cups. The girl then checks her two cups. Next, she must lift one cup and show it to the boy. The girl must always lift a cup that hides nothing. Hence, the boy and the girl are both left with one cup. The boy could see that the marble was not under the third cup. Now, he has to indicate which one of the two remaining cup hides the marble.

Either the boy picks the cup that he initially chose or he chooses the girl's cup. What should he do to have most chance of winning the game?

a. Stick with the cup he initially chose.

b. He should switch and choose the girl's cup.

c. It does not matter which cup he chooses. Chances are even