THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY
2006, 59 (6), 1070–1100

Psychology Press
Taylor & Francis Group

Automatic–heuristic and executive–analytic processing during reasoning: Chronometric and dual-task considerations

Wim De Neys
University of Leuven, Leuven, Belgium

Human reasoning has been shown to overly rely on intuitive, heuristic processing instead of a more demanding analytic inference process. Four experiments tested the central claim of current dual-process theories that analytic operations involve time-consuming executive processing whereas the heuristic system would operate automatically. Participants solved conjunction fallacy problems and indicative and deontic selection tasks. Experiment 1 established that making correct analytic inferences demanded more processing time than did making heuristic inferences. Experiment 2 showed that burdening the executive resources with an attention-demanding secondary task decreased correct, analytic responding and boosted the rate of conjunction fallacies and indicative matching card selections. Results were replicated in Experiments 3 and 4 with a different secondary-task procedure. Involvement of executive resources for the deontic selection task was less clear. Findings validate basic processing assumptions of the dual-process framework and complete the correlational research programme of K. E. Stanovich and R. F. West (2000).

One of the main themes of cognitive reasoning research over the last few decades is that human judgement frequently violates traditional normative standards. In a range of reasoning tasks most educated people fail to give the answer that is correct according to logic or probability theory. One of the most celebrated examples of this failure is the notorious conjunction fallacy (e.g., the "Linda problem", Kahneman, Slovic, & Tversky, 1982; Tversky & Kahneman, 1983). In this task people typically read a short personality sketch, for example: "Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in antinuclear demonstrations." Participants are then asked to rank several hypotheses according to their probability, including "(A) Linda is active in the feminist movement", "(B) Linda is a bank teller", and "(C) Linda is a bank teller and is active in the feminist movement".

The conjunction rule, the simplest and most fundamental law of probability (Tversky &

Correspondence should be addressed to Wim De Neys, Lab Experimentele Psychologie, K.U. Leuven, Tiensestraat 102, 3000 Leuven, Belgium. Email: Wim.Deneys@psy.kuleuven.be

This research was supported by the Fund for Scientific Research Flanders (FWO-Vlaanderen). The author would like to thank Walter Schroyens, Walter Schaeken, and Niki Verschueren for their helpful comments.

http://www.psypress.com/qjep DOI:10.1080/02724980543000123

Kahneman, 1983), holds that the probability of a conjunction of two events cannot exceed that of either of its constituents—that is, $p(A\&B) \leq p(A), p(B)$. Thus, there should always be more individuals that are simply bank tellers than individuals that are bank teller and in addition also active in the feminist movement. However, Kahneman and Tversky, and numerous others since, found that the vast majority (over 80%) of university students violate the conjunction rule and rate Statement C as more probable than Statement A or B.

A further classic demonstration of people's difficulty with adhering to logical, normative standards comes from Wason's (1966) selection task. The selection task is probably the single most investigated task in the whole psychological literature on reasoning (Evans, 2002). In the standard task people are shown four cards with a letter on one side and a number on the other. The four cards displayed might have the values A, T, 4, and 7 on their visible sides. Participants have to check whether a rule, for example "If there is an A on one side, then there is a 4 on the other side" applies to the cards. Participants have to indicate which cards need to be turned over in order to decide whether the rule is being followed. The normatively correct solution hinges on the standard logical falsification principle: It is necessary to turn over cards whose hidden values might falsify the rule: thus, cases in which an "A" is not coupled with a "4". The only cards that might lead one to find such a case are the "A" and "7" cards. As with the conjunction rule, the logical falsification principle is massively violated. Typically, most people simply select the cards that match the lexical content of the rule (e.g., "A" and "4" cards). Less than 10% of university students manage to select the correct cards (e.g., Evans, Newtead, & Byrne, 1993; Manktelow, 1999).

Faced with the impressive gap between people's actual performance and normative standards (i.e., the so-called normative/descriptive gap), advocates of the dual-process framework of thinking (e.g., Epstein, 1994; Evans & Over, 1996; Goel, 1995; Kahneman, 2000; Sloman, 1996; Stanovich, 1999) have posited the existence of two different cognitive reasoning systems. Dual-process theories of reasoning come in various flavours, but in general one of the reasoning systems is typically characterized as automatic, associative, unconscious, and undemanding of computational working memory capacity. This system, often termed the heuristic system, will respond rapidly and is biased toward judgements based on overall similarity to stored prototypes (Sloman, 1996).

The second system (often termed the analytic system) is typified as consciously controlled, deliberate, and effortful. Analytic processing is assumed to be serial, time-consuming, and heavily demanding of our limited computational working memory capacities. The analytic system would operate on "decontextualized" representations in which the underlying structure of a task is decoupled from superficial content.

The typical failure to provide the normatively correct answer on standard reasoning tasks has been attributed to the pervasiveness of the heuristic system. Whereas the fast and undemanding heuristics provide us with useful responses in many situations they may also bias reasoning in tasks that require more elaborate, analytic processing. That is, both systems will sometimes cue different responses. In these cases the analytic system will need to override the response generated by the heuristic system (Stanovich & West, 2000).

Tversky and Kahneman (1983) attributed the conjunction fallacy, for example, to the operation of the so-called representativeness heuristic. Representativeness refers to an automatic assessment of the correspondence between an instance and a category. The personality description of Linda is very representative of an active feminist but not of a bank teller. Adding feminism to the profession of bank teller improves the match with Linda's personality description. The representativeness heuristic thereby prompts us to conclude that Linda is a feminst bank teller rather than a mere bank teller. Evans and Lynch (1973, see Evans, 1998b, for a review) showed that a matching bias or heuristic, an automatic tendency to focus attention to items explicitly named in the rule, is responsible for the typical erroneous selection task response.

Although the dual-process framework has been very influential (and with the work of A. Tversky & D. Kahneman even indirectly awarded a Nobel prize) the posited processing characteristics of the heuristic and analytic system have been severely questioned (Gigerenzer & Regier, 1996; see also commentaries on Evans & Over, 1997, and Stanovich & West, 2000). The present study focuses on one of the framework's most fundamental processing assumptions: The differential involvement of executive, working memory resources in heuristic and analytic processing (e.g., Evans, 2002, 2003; Evans & Over, 1996; Feldman Barrett, Tugade, & Engle, 2004; Kokis, Macpherson, Toplak, West, & Stanovich, 2002; Stanovich & West, 2000): Heuristic operations are assumed to be completely automatic (i.e., undemanding of executive resources) whereas analytic operations would heavily draw on the executive resources.

Executive working memory resources are widely considered as the quintessential component of cognitive capacity. Working memory (WM) is typically characterized as a hierarchically organized system in which specific storage and maintenance components subserve a central component responsible for the control of information processing (e.g., Baddeley & Hitch, 1974; Cowan, 1995; Miyake & Shah, 1999). The controlling component or "central executive" is conceived of as a limited-capacity system that regulates the allocation of attentional resources. It has been shown that general cognitive ability tests primarily reflect central executive capacity (e.g., Engle, Tuholski, Laughlin, & Conway, 1999; Kane, Bleckley, Conway, & Engle, 2001).

The posited differential recruitment of executive resources in heuristic and analytic reasoning received substantial support from the research programme of Stanovich and West (e.g., 1998a, 1998b, 1998c, 2000, 2003). These authors started a systematic examination of the impact of individual differences in cognitive capacity on classic reasoning problems. Given that the computations of the analytic system would draw on limited, executive WM resources one can expect that limitations in executive resources are a primary cause of erroneous heuristic responses: The more resources that are available, the more likely that the analytic system will be successfully engaged and the correct response calculated. Stanovich and West found that participants higher in cognitive capacity (as assessed by standard cognitive ability tests) were indeed more likely to give the correct normative response (see also Klaczynski, 2001; Newstead, Handley, Harley, Wright, & Farrelly, 2004; Torrens, Thompson, & Cramer, 1999; Valentine, 1975).

One fundamental limitation of the Stanovich and West findings, however, is that they remain purely correlational (see Stanovich & West, 2000). The reported correlations do not establish the assumed causality: The findings indicate that selecting the correct, analytic response is associated with having a larger WM resource pool, but this does not imply that the WM resources are necessary for the calculation of the correct response. Thus, some other factor might account for the positive associations (e.g., Klaczynski, 2000; Newton & Roberts, 2003; Sternberg, 2000). The present study uses a secondary-task approach to test the central dual-process claim experimentally. The rationale behind the secondary-task experiments was suggested by Sloman (1996, p. 17). In commenting on the lack of explicit evidence for the automaticity assumption in dual-process theories Sloman argued that the crucial prediction would be that a cognitive load should differentially affect analytic and heuristic responses. If correct analytic responding draws on executive resources, performance should decrease under load since fewer resources will be available for inhibition of the prepotent heuristic response and subsequent analytic computations. If the heuristic system operates automatically, heuristic responses should not decrease under secondary-task load. On the contrary, since it will be harder for the analytic system to override the prepotent heuristics one expects a specific increase in heuristic responses under secondary load.

The present experiments focus on conjunction fallacy and selection task problems since studies with these tasks lie at the core of dual-process theorizing in cognitive science (e.g., Evans, 1984; Evans & Over, 1996; Gigerenzer & Regier, 1996; Kahneman, 2000; Sloman, 1996) and were most extensively studied by Stanovich and West (e.g., 1998a, 1998b). Both tasks also stem from two somewhat separated research programmes (i.e., the deductive-reasoning field and the heuristics and biases field), which are brought together in the dual-process framework (Evans, 2002; Stanovich & West, 1998c). Furthermore, the degree of stability of the results across the two different tasks will be indicative of the generality of the findings. The response that is typically characterized as processed by the heuristic system will be referred to as the heuristic response (i.e., making the conjunction fallacy or selecting the matching card pattern in a standard selection task). The response that is considered to be computed by the analytic system is referred to as the analytic response (i.e., the responses that are correct according to traditional normative standards).

A second general dual-process claim, associated with the differential-resource assumption, concerns the differential-processing speed of both reasoning systems. The heuristic system is assumed to operate faster than the analytic system (e.g., Epstein, 1994; Evans & Over, 1996; Goel, 1995; Kahneman, 2000; Sloman, 1996; Stanovich, 1999). The claim makes sense from a theoretical viewpoint since research on executive functioning clearly established that in contrast with automatic processing, executive processing is serial and time consuming (e.g., Cowan, 1995; Kane et al., 2001; Logan, Taylor, & Etherton, 1996; McElree, 2001; Shiffrin & Schneider, 1977). One thus expects that heuristic responses require less processing time than analytic responses. However, there is surprisingly little direct chronometric evidence for this claim. As has reasoning and memory research in general, dual-process-related studies have been mainly concerned with response accuracy data and not with response latencies (Kahana & Loftus, 1999; Thompson, Striemer, Reikoff, Gunter, & Campbell, 2003). Despite over 25 years of experimenting with the conjunction fallacy a basic latency analysis (i.e., comparing inference times for correct vs. heuristic responses) has not yet been presented. Experiment 1 attempts to fill this time gap. Chronometric studies of the selection task have been somewhat more widespread (e.g., Evans, 1996; Osman, 2002; Roberts, 1998; Roberts & Newton, 2001). However, these studies have typically modified the standard task and/or have focused on more specific predictions. For example, Osman (2002) presented the cards serially and measured the time that people needed to decide whether an individual card needed to be turned over or not. Evans (1996) measured "inspection times" or the time people spend considering each of the four presented cards. Findings have been mixed, and overall the validity of the assumptions underlying the task modifications has been questioned (e.g., Roberts, 1998; Roberts & Newton, 2001; but see Evans, 1998a). Therefore, Experiment 1 did not focus on individual card selection latencies or inspection times but compared the inference latencies for the final selected card pattern. This more general analysis avoids possible complications associated with previous latency studies and still allows a test of the crucial assumption under consideration.

It was argued that the output of the heuristic system is not always erroneous. Dual-process theories have underlined that in many situations both the analytic and heuristic system cue the same, normatively correct response. In these cases the heuristic system provides us with a correct, "fast and easy" problem solution. One paradigmatic example comes from selection tasks with deontic rules. Contrary to the original, so-called "indicative" selection task that describes an arbitrary conditional rule (e.g., "If there is an A, then there is a 7" or "If P, then Q"), the deontic task asks people to reason about known regulations (e.g., "If a person is drinking beer, the person must be over 21 years of age"). The task describes a scenario that requires people to search for rule violations. In case of the "drinking age" problem (Griggs & Cox, 1982), for example, people are told that they are a policeman that needs to check whether the law is respected in a specific bar. The cards are said to display the age and the drinks of customers in the bar. Participants are shown four cards with an equivalent logical status as in the standard selection task (i.e., two cards that affirm the antecedent and consequent of the rule, and two cards that negate the antecedent and consequent). In the deontic task the cards would display, for example, "drinking coke", "drinking beer", "age 22", "age 14". Such task content has a spectacular facilitatory effect on the performance. With the deontic version of the selection task about 80% of the participants select the correct falsification answer (see Evans et al., 1993; Manktelow, 1999, for a review).

This facilitatory deontic effect has been attributed to the fact that the heuristic system triggers the correct response in these tasks (Evans, 2002; Evans & Over, 1996; Sloman, 1996; Stanovich & West, 1998a). Even without doing any analytic computations our semantic knowledge about drinking regulations would make it readily clear that it is necessary to "check the youngster!": The tendency of the heuristic system to "contextualize" a problem with prior semantic knowledge would automatically result in selection of the appropriate cards (Stanovich & West, 2000). Consistent with the claim that deontic versions of the selection task do not require analytic, WM-dependent reasoning, Stanovich and West (1998a) showed that for these tasks there was no association between cognitive capacity and performance (but see also Newstead et al., 2004).

The present study adopted both indicative and deontic selection task versions. This allowed an additional test of the processing assumptions of the dual-process theories. If selection of the correct cards in the deontic version is indeed heuristic based we do not expect a decrease in correct card selections under secondary-task load. In addition, differential latency findings are predicted. The assumed heuristic-based correct card selection in the deontic task should be faster than the analytic-based correct card selection in the indicative version.

In sum, Experiment 1 presents a chronometric comparison of the analytic and heuristic responses. Experiment 2 imposes an executive complex tapping task while participants solve conjunction and selection problems. Experiments 3 and 4 attempt to replicate the findings with a different secondary-task procedure. The tested central processing assumption amounts to the claim that the heuristic response is given fast and automatically whereas the analytic response requires more time-consuming executive processing.

EXPERIMENT 1

Experiment 1 tests the claim that the automatic–heuristic system operates faster than the executive–analytic system (Evans & Over, 1996; Sloman, 1996; Stanovich & West, 2000). Participants were presented with conjunction fallacy problems and an indicative and deontic selection task. The time needed to give a response was recorded together with the time participants needed to read the preambles in order to separate inference time from mere reading time. The dual-process framework predicts longer inference times for the analytic response than for the heuristic response. In addition, it is predicted that the assumed heuristic-based correct card selection in the deontic selection task should be made faster than the analytic-based correct card selection in the indicative version.

To make as few assumptions about the timing registration as possible the analysis measured the total time needed to give the complete answer (i.e., the complete ordering of hypotheses or the complete selected card pattern). The nature of the latency analysis implies that on each task, every participant only provides a single registration. In order to get a reliable measurement a large number of participants were tested, and the analysis was restricted to the most frequently generated responses.

Method

Participants

The 189 participants were all first-year students at the University of Leuven, Department of Social

Sciences, who either participated voluntarily or received course credit for taking part.

Materials

Participants were presented a total of four problems: two conjunction fallacy problems and two selection tasks. All problems were translated into Dutch and were adapted for computerized presentation. The conjunction fallacy tasks were the "Linda" problem and an adaptation of the analogous "Bill" problem (Tversky & Kahneman, 1983):

Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in antinuclear demonstrations.

Please rank the following statements by their probability:

A. Linda is a bank teller

B. Linda is active in the feminist movement

C. Linda is a bank teller and is active in the feminist movement

Most probable statement:_

Second most probable statement:_

Least probable statement:_

Bill is 34 years old. He is intelligent, very punctual, but unimaginative, rather compulsive, and generally lifeless. In school, he was strong in mathematics but weak in social studies and humanities.

Please rank the following statements by their probability:

A. Bill plays in a rock band for a hobby

B. Bill is an accountant

C. Bill is an accountant and plays in a rock band for a hobby

Most probable statement:_

Second most probable statement:_

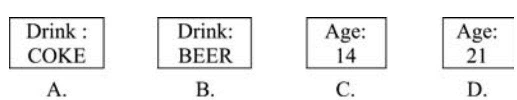
Least probable statement:_

Participants only ranked the two constituent hypotheses (A, B) and the conjoint (C) hypothesis. Thus, no filler statements were presented (see Hertwig & Gigerenzer, 1999).

Both the deontic ("drinking age problem") and indicative ("destination problem") versions of the selection task were adapted from Stanovich & West (1998a):

Imagine that you are an American police officer on duty, walking through a local bar. It is your job to ensure that the drinking laws are in effect in this bar. The law states that if a person is drinking beer, then the person must be over 21. There are four persons in the bar. On a form two pieces of information have been reported for every person. Whether or not a person is drinking beer is on one side of the form and the person's age is on the other side. You will get to see those four forms below. You'll only get to see one side of the form. So, either you get to see a person's age or you get to see what the person is drinking. Your task is to decide whether or not the law is being broken in the bar. You have to indicate which forms would need to be turned over in order to decide whether the law is being violated. You must only turn the forms that you definitely need to check.

Which of the following forms do you have to turn to find out whether the rule "If a person is drinking beer, then the person needs to be over 21 years of age" is being violated?



—Type down the letters of the cards that need to be turned. Press space bar when finished—

The indicative task was composed of realistic but arbitrary content. Both destinations referred to nonexisting places to make sure that participants had no prior knowledge about the relation that the conditional rule expressed:

Below you will get to see 4 cards. Each card has a destination on one side and a mode of travel on the other side. Your task is to decide which cards you have to turn over in order to decide whether or not a specific rule is being followed. The rule states that if the destination is CRANSHIL, the mode of transportation is CAR.

Which of the following cards do you have to turn to find out whether the rule "If CRANSHIL is on one side of the card, then CAR is on the other side" is being violated?

Destination: VELPLACE	Destination: CRANSHIL	Transport: CAR	Transport: TRAIN
A.	B.	C.	D.

—Type down the letters of the cards that need to be turned. Press space bar when finished—

For every problem the preambles (italicized text) were first presented on the computer screen. When participants finished reading they pressed the enter key and then the actual question and the answer alternatives appeared (the italicized text remained on the screen). Participants typed down their responses on the keyboard and pressed the space bar when finished.

For the conjunction problems participants typed a letter (a, b, or c) corresponding to the most, second most, and least probable statements. By pressing the enter key participants could freely cycle through the three answer statements and type their answers. When participants were satisfied with their ranking they pressed the spacebar. In the selection tasks participants typed the letter (a, b, c, or d) of the cards that needed to be turned. When a letter was entered the corresponding card on the screen was shaded, and the statement "TURN" appeared below the card. When participants typed the same letter once more the card was deselected. Participants pressed the space bar when they had finished the card selection.

Participants were instructed to press the enter key immediately after they had finished reading the preambles. The time that elapsed between the item presentation and the enter key press was defined as *reading time*. Instructions clarified that subsequently participants could take all the time they wanted to think about the problem question. The time that elapsed between the enter key press and the pressing of the spacebar was defined as the *inference time*. The labels "reading" time and "inference" time and the precise splitting point are of course somewhat arbitrary. The rationale was simply that the inference phase would start with the presentation of the problem question and the answer alternatives. More crucial is the fact that the split-up allowed the sidestepping of possible complications due to the different length of the preambles. The inference phase in both conjunction problems and both selection problems presents the same amount of information and requires the same type of response. Thus, especially with respect to the selection tasks, difference in inference times cannot be attributed merely to superficial item characteristics.

Procedure

Participants were tested in groups of 16 to 47 at the same time in a large computer room with an individual booth for every participant. The different response alternatives (i.e., hypotheses or cards) were always presented in the same randomly determined arrangement. All participants also received the tasks in the same fixed order (i.e., "Linda", "Bill", "drinking age", and "destination") to minimize any measurement error due to a participant by order interaction.

Participants were instructed about the specific task and response format of the conjunction problems and received one practice item. In the practice item participants simply rank ordered the probability of three events about which they had received likelihood information (e.g., see the "base-rate only" tasks for children in Kokis et al., 2002). A similar item was presented as filler between the "Linda" and "Bill" problems. After the two conjunction problems were solved, item presentation was paused until the participant decided to continue. Next, general instructions and a practice item for the selection task were presented. The practice item was loosely based on the "cholera" problem (e.g., Cheng & Holyoak, 1985).

Results and discussion

Conjunction problems

Results for the conjunction problems are presented in Table 1. The response accuracy pattern replicated well the classic Tversky and Kahneman (1983) findings. Both on the "Linda" (85%) and on the "Bill" (81%) problems the vast majority of participants committed the conjunction fallacy and ranked the conjunction as more probable than any of its constituents.

The dual-process framework claims that making the incorrect conjunction fallacy results from a fast heuristic reasoning process whereas the correct response is assumed to be computed by a slow analytic process. As Table 1 shows, inference latencies supported the basic prediction. Participants who gave the correct response on the "Linda" problem had significantly longer inference times than participants who committed the conjunction fallacy, $t(187) = 2.07$, $p < .025$, one-tailed.[1] Although participants overall tended to speed up on the subsequent "Bill" problem, a similar trend was observed: Correct responders had longer inference times than incorrect responders, $t(187) = 3.83$, $p < .001$, one-tailed.

The present latency analysis assumes that longer response times during the inference phase reflect a slower operating reasoning process. A straightforward alternative explanation is that people who solve the problem correctly (i.e., presumably the more cognitively gifted people, e.g., Stanovich & West, 1998b) are generally more cautious. They could, for example, take more time because they read the answer alternatives in the inference phase more carefully. Thus, longer latencies would not be associated with a slower operating inference process per se. However, such a "general cautiousness" factor should also show up during the processing of the preambles in the reading phase. If participants who solve the problem correctly have longer inference times because they are generally more cautious, the cautiousness should also result in longer latencies for the reading phase. However, on both the "Linda", $t(187) = 0.87$, $p = .2$, one-tailed, and the "Bill" problems, $t(187) = -0.05$, $p = .43$, one-tailed, reading times did not differ for those who solved the problem correctly and those who solved it incorrectly.

Indicative selection task

The presentation of the selection data adopts the common selection task nomenclature. Thus, the cards that correspond to the antecedent and consequent of the conditional rule are referred to as the *P* and *Q* cards, respectively. Cards that correspond to the negation of the antecedent and consequent are labelled the *not-P* and *not-Q* cards, respectively. Table 2 presents an overview of the results.

About 18% of the participants correctly selected the P and not-Q card pattern on the "destination" problem. Note that the correct solution rate for indicative selection tasks typically hovers around 10% (Evans et al., 1993). The slightly higher figure in the present experiment is probably due to the prior presentation of the deontic "drinking age" problem (e.g., see Stanovich & West, 1998a). The distribution of the response patterns further followed classic findings, with the majority of participants (53%) selecting the matching pattern (P and Q cards) and a substantial number of participants selecting the P card only (22%). These three patterns together accounted for 92.6% of the selections. The remaining patterns were only rarely selected (i.e., none of them was selected by more than 4 out of 189 participants). Because the low number of observations did not allow a reliable latency estimation the latency data for these infrequently selected patterns were not analysed. The data are included in Table 2 for completeness.

[1] All reported latency tests were also analysed with nonparametric tests. Results for parametric and nonparametric analyses were completely consistent.

Table 1. *Mean inference and reading times[a] for correct and incorrect conjunction responses*

	Correct					Incorrect				
		Inference time		Reading time			Inference time		Reading time	
	n	*M*	*SD*	*M*	*SD*	*n*	*M*	*SD*	*M*	*SD*
Linda problem	28	57.52	21.42	11.14	4.43	161	47.23	24.75	10.33	4.55
Bill problem	36	44.14	21.89	8.36	3.7	153	29.93	19.59	8.39	3.17

[a]In s.

The dual-process framework claims that selecting the matching P and Q pattern results from a fast heuristic reasoning process whereas the correct P and not-Q response is assumed to be computed by a slow analytic process. As Table 2 indicates, inference latencies supported the basic prediction. Inference times for the P and not-Q pattern were significantly longer than those for the matching, P and Q, pattern, $t(132) = 2.48$, $p < .01$, one-tailed.

In addition to these two main patterns of interest, latency data for the "P only" (i.e., participants who selected the P card only) pattern were also compared. This pattern was selected by a substantial number of participants. The pattern is of special interest since Margolis (1987) argued that, theoretically, the "P only" selection on indicative selection tasks might be viewed as an additional normatively correct, analytic response. Such suggestions have been uttered for other possible selection patterns (e.g., Evans, 1977). However, consistent with Margolis' claim, Stanovich and West (1998a) observed that, as with the P and not-Q response, participants who gave the "P only" response were significantly higher in cognitive capacity than were matching responders. An analysis of variance (ANOVA) established that the inference times for the P and not-Q, P only, and matching response patterns differed significantly, $F(2, 172) = 3.94$, $MSE = 216.09$, $p < .025$. The latency data supported the Stanovich and West findings: Inference times of participants who selected "P only" were indeed longer than those of matching selectors, $t(139) = 2.09$, $p < .02$, one-tailed.

Table 2. *Selection frequency and mean inference and reading times[a] for the different response patterns in the indicative and deontic selection tasks*

	Indicative					Deontic				
		Inference time		Reading time			Inference time		Reading time	
Card pattern	*n*	*M*	*SD*	*M*	*SD*	*n*	*M*	*SD*	*M*	*SD*
Correct	34	27.91	13.26	20.07	6.34	168	19.33	11.88	41.43	14.98
Matching	100	21.08	14.07	19.15	7.13	4	43.65	28.77	57.37	27.96
P only	41	26.88	17.15	19.26	8.98	6	40.08	22.23	45.36	17.80
All	4	49.42	11.58	16.88	7.23	1	34.82	–	3.40	–
P, Q, not-Q	3	24.31	11.46	14.24	3.59	2	21.64	1.09	62.73	11.34
Q	3	31.20	19.09	31.01	12.06	1	22.30	–	31.86	–
Not-P, not-Q	3	66.13	29.04	14.56	8.75	1	62.45	–	48.94	–
P, not-P	1	33.94	–	12.74	–	0	–		–	
Not-Q	0	–		–		6	15.20	10.36	39.26	7.44

[a]In s.

The reading times of participants who selected the matching, P only, or P and not-Q patterns did not differ, $F(2, 172) < 1$.

Deontic selection task

About 88% of the participants selected the P and not-Q card pattern on the "drinking age" problem. This solution rate is close to the solution rate of 85.7% observed by Stanovich and West (1998a) and establishes that the present version obtained the full facilitatory deontic effect. As could be expected, other card selection patterns were only rarely selected (e.g., no pattern was selected by more than 6 participants). Therefore, comparing the latency data for different selection patterns in the deontic task is not fully warranted. Note that the following analyses involve within-subject comparisons and are restricted to the group of participants who gave the correct deontic P and not-Q card response.

The deontic task was included in the experiment to test the general dual-process claim that the correct P and not-Q card selection in the deontic task is based on heuristic processing. It follows that the assumed heuristic-based correct P and not-Q card selection in the deontic task should be made faster than the analytic-based correct P and not-Q card selection in the indicative task. A within-subjects t test supported the prediction; for the 32 participants who solved both the indicative and deontic tasks correctly, the indicative task (M = 28.56 s, SD = 13.27) demanded more inference time than did the deontic task (M = 16.83 s, SD = 7.10), $t(31) = 4.90$, $p < .0001$, one-tailed. Note that because the indicative task was presented after the deontic task any speeding up due to familiarization with the task format should have played against the hypothesis.

One could also compare the inference times for the assumed heuristic response in both selection tasks (i.e., the correct P and not-Q pattern in the deontic task and the matching P and Q pattern in the indicative task). Interestingly, a within-subjects t test on the inference times of the 92 participants who solved the deontic task correctly (M = 19.41 s, SD = 11.83) but matched in the indicative task (M = 20.49 s, SD = 13.88) showed no significant differences, $t(91) = -0.71$, $p = .24$, one-tailed. Thus, inference times for two selection patterns that are both assumed to be computed by the heuristic system do not seem to differ. However, the validity of the conclusion is somewhat restricted since one cannot exclude the possibility that familiarization with the task format decreased inference times on the indicative task.

For completeness, the deontic and indicative inference latencies in function of the response on the indicative task were also entered in an ANOVA. This resulted in a 2 (task, within subjects)×2 (indicative correct vs. match, between subjects) design. There was a main effect of task with longer inference times for the indicative task, $F(1, 122) = 18.82$, $MSE = 103.47$, $p < .001$, but no main effect of response pattern, $F(1, 122) = 1.78$, $MSE = 202$. Both factors also interacted, $F(1, 122) = 12.99$, $MSE = 103.47$, $p < .001$. Whereas inference times for indicative matchers and correct responders differed on the indicative task, $F(1, 122) = 8.20$, $MSE = 188.64$, $p < .005$, both groups solved the deontic problem equally fast, $F(1, 122) = 1.34$, $MSE = 117.13$. As the above findings implied, the increased indicative latencies were only apparent when the indicative task was solved correctly.

A comparison of reading times for the deontic and indicative selection task is not very informative since the preambles of both problems differ in length considerably. Nevertheless, for completeness we reran the above ANOVA on the reading times. Given the lengthier preambles of the deontic problem it is not surprising that overall participants needed more time to read them, $F(1, 122) = 354.87$, $MSE = 65.56$, $p < .0001$. More important is that the reading times did not differ as a function of the response pattern, $F(1, 22) = 2.2$, $MSE = 120.31$, and the two factors did not interact, $F(1, 122) = 2.78$, $MSE = 65.56$.

Analytic index

A final latency analysis focused on the correlation between an analytic composite index and response latencies. The analytic index consisted of the three problems for which giving the correct response was assumed to require analytic reasoning. One point was given for every correct conjunction response and one point for the selection of the P and not-Q pattern on the indicative selection task. The analytic index scores consequently ranged between 0 and 3. The correlation between the index and the total inference time for the three problems reached .24, $n = 189$, $p < .001$. Thus, overall a higher number of analytic responses is associated with longer inference times. The correlation with the mean reading time for the three items did not reach significance, $r = .05$, $n = 189$, $p = .47$.

Conclusion

A chronometric analysis of the classic conjunction fallacy problems and indicative selection task supported the dual-process claim concerning the differential operating speed of the heuristic and analytic system. Responses that were assumed to be processed by the heuristic system required less inference time than did assumed analytic responses. The time needed to read the preambles did not differ for analytic and heuristic responses. The latency difference was only apparent for the inference phase that started with the presentation of the problem question and the answer alternatives. This indicates that the measured latency differences are not caused by a general cautiousness factor but are specifically tied to differences in the operation speed of systems that support problem solving.

Findings also supported the claim that in contrast to indicative tasks, the heuristic system provides the correct response in deontic tasks. Inference times for the correct, P and not-Q card combination were much faster in the deontic task than in the indicative task.

Experiment 1 validated the operation speed assumption of dual-process theories. The differential processing speed assumption is linked with the processing resource assumption. It is widely acknowledged that in contrast with automatic processing, executive processing is serial and time consuming (e.g., Cowan, 1995; Kane et al., 2001; Logan et al., 1996; Shiffrin & Schneider, 1977). The reaction time data are thus consistent with the assumed differential involvement of executive WM resources in heuristic and analytic reasoning. The remaining experiments present a conclusive test of the processing resource assumption.

EXPERIMENT 2

Experiment 2 presents an experimental test of the claim that the analytic system draws on executive WM resources whereas the heuristic system operates automatically without executive resources. Participants solved the conjunction fallacy and selection task problems of Experiment 1 while they performed a secondary task that burdened the executive WM resources.

The experiment attempts to complete the influential correlational research programme of Stanovich and West (e.g., 1998a, 1998b, 1998c, 2000, 2003). Stanovich and West's findings suggested that the number of available executive resources were associated with the probability of analytic responding. If correct analytic responding draws on executive resources, the frequency of analytic responses should decrease under executive load since fewer resources will be available for the analytic computations. On the other hand, a system that is assumed to operate automatically should not be hindered by an executive load. The heuristic response in the selection and conjunction tasks is considered as the prepotent, default response that will need to be overridden by the analytic system. If the heuristic system is indeed the default system, one therefore expects an increase in heuristic responses when the analytic computations become harder under executive load.

The experiment is one of the first to introduce secondary-task methodology in the heuristics and biases and the dual-process research fields. There is one specific previous experiment that might also speak to the issue. In a study designed to assess the impact of mood states on selection task performance, Oaksford, Morris, Grainger, and Williams (1996) presented a dual-task experiment to validate one of their mood manipulations. Results indicated that when executive resources were burdened, correct P and not-Q selection responses decreased whereas matching responses increased. Hence, the findings might seem to validate the dual-process framework's basic claim. However, in the present context, one fundamental limitation of the study is that Oaksford et al. used a special "no rationale" (Cheng & Holyoak, 1985), deontic selection task. The dual-process framework claims that the high number of correct responses in a standard deontic task is based on automatic–heuristic processing. Therefore, a secondary-task impact on the standard deontic task would be problematic. However, the no-rationale version is not a standard problem since it merely presents a deontic rule without a clear scenario or explicit rationale that emphasizes the detection of transgressors of the rule (e.g., the policeman scenario in the drinking age problem). No-rationale versions do not show the almost perfect selection performance that was the very ground for the deontic automaticity claim. Presentation of the rationale is paramount in order to derive the full facilitatory deontic effect (Cheng & Holyoak, 1985; Pollard & Evans, 1987; Stanovich & West, 1998a). Hence, the status of the no-rationale task with respect to the involvement of the heuristic and analytic systems is somewhat unclear. The present study sidestepped this problem by using both a standard indicative and a standard deontic task. The assumed analytic-based selection of the correct P and not-Q cards is expected to decrease under executive load on the indicative task. If the correct response on the deontic task is indeed processed by the heuristic system one expects no secondary-task impact here.

The secondary task was adopted from Kane and Engle (2000) and Moscovitch (1994). Participants were requested to tap a complex, novel finger sequence (e.g., index finger–ring finger–middle finger–little finger) with their nondominant hand. The Kane and Engle and Moscovitch studies consistently showed that the task put a premium on efficient executive WM functioning.

Method

Participants

A total of 42 first-year psychology students from the University of Leuven, Belgium, participated in return for course credit. Performance of the 189 participants in Experiment 1 served as a baseline for the effect of introducing the secondary task.

Materials

Reasoning tasks. Participants were presented the same problems as those in Experiment 1, in the same order, and with the same instructions and practice items. With respect to the nature of the deontic selection task it is important to note that the "drinking age" problem adopted in the present study always included the facilitating policeman scenario. Experiment 1 showed that this resulted in the expected (88%) high number of correct responses (e.g., vs. 31% correct responses observed by Stanovich & West, 1998a, for the drinking rule without scenario). Oaksford et al. (1996) observed only 26% correct responses in their control group for the specific deontic selection task that they adopted (i.e., the no-rationale version of the "cholera" problem, see Cheng & Holyoak, 1985). One could argue that such a low number of correct responses indicates that the task did not trigger the typical automatic–heuristic deontic processing. Hence, the relevance of the findings could be questioned based on the suboptimal deontic task characteristics. The selection of a deontic problem that showed full deontic facilitation circumvented this problem in the present study.

Executive tapping 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. The experiment started with a tapping-practice phase. The tapping procedure was based on that of Kane and Engle (2000). Participants were asked to continuously tap the complex index finger/ring finger/middle finger/little finger sequence with their nondominant hand. The experimenter demonstrated the tapping sequence and instructed participants to tap it at a "comfortable and consistent" rate.

The choice for the executive tapping task was inspired by pilot work (e.g., De Neys, 2003) that indicated that this secondary task had an appropriate difficulty level in a reasoning context. One should note that before a reasoner can start reasoning about a problem, the reasoner will have to read and mentally represent the problem information first. Such reading or comprehension processes may also demand WM capacity (e.g., Just, Carpenter, & Keller, 1996). One thus needs a secondary task that interferes with analytic computations but that leaves the more elementary representational processes unaffected. Indeed, if the secondary task would be so demanding that participants would not be able to represent the premises, the findings would not be very informative (e.g., the load would only result in a random guessing pattern). De Neys found that, at least with a conditional inference task, the random-response problem was minimal for Kane and Engle's complex tapping task. In addition, the present experiment only imposed the secondary task during the inference phase and not during the reading of the preambles.

Participants began with five 30-s practice trials of tapping. Participants always received online accuracy feedback: Whenever a wrong finger (key) was tapped the computer emitted an "error" tone (300 ms, low pitch). For the last three practice trials participants also received response time feedback (a 600-ms, high-pitch "speed" tone). Participants received examples of the "error" and "speed" tones, and their different meaning was explained. The computer determined the feedback cut-off times for each participant individually: During the second 30-s practice trial, the computer calculated the mean intertap interval and added 150 ms to it. This became the feedback cut-off for the next practice trial. Thus, if any one intertap interval was more than 150 ms slower than the established mean from the prior practice trial, the computer immediately emitted a "speed" tone (600 ms, high pitch).

During all tapping-practice trials participants were instructed to focus a fixation cross presented at the centre of the computer screen in front of them. Thus, participants could not watch their fingers while tapping.

After the final 30-s practice trial participants received the instructions for the conjunction task. The experimenter explained that the practice tapping speed had to be maintained in the upcoming reasoning tasks. For all reasoning tasks participants always tapped with accuracy and response time feedback. During the reasoning task the "speed" tapping tone was only given for the final finger (i.e., little finger). Thus, no tones were emitted if the first three fingers were tapped too slowly.

All responses for the reasoning task were given orally. The oral responses were entered by the experimenter on the keyboard connected to the computer that ran the reasoning task. For all reasoning tasks participants read the preambles without secondary-task load. Participants indicated orally (i.e., by saying "Ready!") that they had finished reading, upon which the experimenter pressed the enter key. Next, the participant started to tap the finger pattern, and after a 2-s delay the problem question and the answer alternatives were presented. Participants were instructed that they could take as much time as they wanted to solve the problems but practice tapping speeds had to be maintained during the whole inference phase.

Participants said out loud their response (i.e., the letters), and these were entered by the experimenter. Participants said "Ready" to indicate that their answer was final. Participants then stopped tapping, and the next problem was presented. The procedure was clarified with the practice items before the conjunction and selection test problems were presented.

Results and discussion

Conjunction problems

Conjunction performance. Performance on both conjunction problems was combined into a composite total score (i.e., 1 point for every correct conjunction response, scores ranging from 0 to 2) to test the executive load impact on the conjunction fallacy. Table 3 presents the results. Consistent with the dual-process prediction, correct conjunction responses tended to decrease when the executive resources were burdened. However, the decrease on the total conjunction score only reached marginal significance, Mann–Whitney $U = 3{,}511$, $n_1 = 42$, $n_2 = 189$, $p < .06$, one-tailed. Nevertheless, as Table 3 shows, the decrease in correct analytic responding was apparent for both conjunction problems. Although performance on the conjunction problems was already quite floored in the no-load group, the proportion of correct responses was almost halved when the secondary task burdened the executive resources.

The impressive number of participants who fell prey to the conjunction fallacy under secondary-task load suggests that the load specifically promoted heuristic processing and not merely random responding. Note that with the present item format there are only six possible orderings of the three answer statements. Two of these (i.e., 33%) respect the conjunction rule. Thus, any random-response tendency under executive load would have actually resulted in an increase in correct responses.

One could further argue that the above claim only holds under the assumption that all reasoners would start giving random responses under load. It could also be suggested that a random-response effect under load would not affect the people who already reasoned heuristically. Hence, in order to determine the effect of random responding one should only consider the 17% of responses (overall) that were correct in the absence of the load. If only these people (or some of them, sometimes) would start giving random answers under load, the expected number of correct responses would be closer to the observed percentage. However, one can test the random-response hypothesis in this case by examining the distribution of the erroneous responses. If one considers the six possible orderings of the answer alternatives one finds that in one third of the orderings the conjunction rule is respected, in one third the conjunction is erroneously rated as more probable than one of its constituents, and in one third the conjunction is erroneously rated as more probable than both of its constituents. Of the two possible types of erroneous ordering, the first one is typically the most popular (i.e., the "typical" pattern, 86% of the erroneous orderings in the no-load group, see also Reeves & Lockhart, 1993). Now if the analytic reasoners would merely give random responses under load, they should select both erroneous orderings with equal probability. Consequently, one would expect that the strong preference for the typical pattern decreases under load. If the load specifically promotes heuristic processing, the dominance of the typical ordering should not be affected or it may even increase. The load data showed that more than 91% of the erroneous responses under executive load were of the "typical" type. This figure further contradicts the random-responding hypothesis.

Table 3. *Percentage of correct conjunction responses and mean total score with and without concurrent executive load*

	Load group							
	No load (Experiment 1)				*Load (Experiment 2)*			
Measure	*% correct*	*n*	*M*	*SD*	*% correct*	*n*	*M*	*SD*
Linda problem	14.8	28			7.1	3		
Bill problem	19.1	36			12	5		
Total score			0.34	0.62			0.19	0.51

Tapping performance. Due to equipment failure the tapping data of one participant were lost. Analyses are performed on the data of the remaining 41 participants. In the final 30-s practice tapping trial (where participants tapped without reasoning) participants tapped a mean number of 2.30 ($SD = 0.64$) correct taps per second. During the inference phases of the conjunction problems, this figure decreased to a mean number of 1.72 ($SD = 0.60$), $t(40) = 7.50$, $p < .0001$, one-tailed. Thus, concurrent conjunction problem solving also affected the tapping performance. However, the decrease in tapping performance (i.e., mean correct taps per second during final practice minus mean correct taps per second during conjunctive reasoning) was not associated with conjunction performance. Indeed, participants with higher conjunction scores even tended to show less tapping decrease, $r = -.15$, $n = 40$, $p = .37$. This indicates that participants were not simply trading off reasoning and tapping performance.

Indicative selection task

Selection performance. Table 4 presents an overview of the selection task results. Findings clearly corroborated the dual-process predictions. Under executive load the proportion of participants who correctly selected the P and not-Q card pattern on the indicative task was more than halved to about 7%. The decrease in correct responses was accompanied by a specific increase in matching responses to about 69% of card selections. Comparing the frequencies of these two responses in the no-load and load group showed that the cross-over interaction was marginally significant, $\chi^2(1, N = 166) = 3.82, p < .051$.[2] The selection rate of the P only pattern also tended to decrease under executive load but the cross-over interaction with the matching pattern did not reach significance, $\chi^2(1, N = 176) = 2.04$, $p < .16$. The selection rate of other patterns was hardly affected. Thus, card selections under load were not simply random but followed a systematic pattern: Analytic card selections decreased, whereas heuristic responding increased.

The selection task findings provide further support for the dual-process assumption. Note that on a conjunction problem one either gives the correct response or commits the conjunction fallacy. Thus, a decrease in correct responses will always be accompanied by an increase in conjunction fallacies. However, on the selection task one can select numerous responses (i.e., a total of 16 possible card combinations) besides the heuristic matching pattern. The fact that the decrease in correct responses is specifically accompanied by an increase in matching responses is clear evidence for the claim that the heuristic system is the dominant, default reasoning system (Stanovich & West, 2000).

To act further as a confirmation of the findings Pollard indices (Pollard & Evans, 1987) were used as dependent variable in parametric, planned comparisons. Pollard indices are calculated in the following way. The analytic response is to turn the P and not-Q cards. A score of one is assigned to each card turned and zero to each card not turned. The falsification index (FI) is FI = (P + not-Q) − (not-P + Q), which ranges between −2 and +2. A high FI indicates that participants are responding analytically. The confirmation index (CI) indicates a tendency to make matching P and Q card selections. CI is calculated in a similar way to FI: CI = (P + Q) − (not-P + not-Q). A high CI indicates that participants are responding heuristically. Note that because FI and CI are not independent they cannot be compared with each other within a task.

Table 4 also shows the means of the Pollard indices on the indicative selection task in the no-load and load groups. As predicted, the FIs decreased under executive load, $t(229) = 2.39$, $p < .01$, one-tailed. The CIs showed the opposite trend but the effect did not reach significance,

[2] Analyses involving the χ^2 test were also analysed with the Fisher Exact Probability Test since the expected cell frequencies were sometimes small. Results for both tests were always consistent.

Table 4. *Proportion of card selections and mean falsification and confirmation indices with and without concurrent executive load*

		Indicative								Deontic							
		No load				Load				No load				Load			
Measure		%	*Freq.*	*M*	*SD*	%	*Freq.*	*M*	*SD*	%	*Freq.*	*M*	*SD*	%	*Freq.*	*M*	*SD*
Card pattern	Correct	18	34			7.1	3			88.9	168			73.8	31		
	Matching	52.9	100			69.1	29			2.1	4			9.5	4		
	P only	21.7	41			14.3	6			3.2	6			11.9	5		
	Other	7.4	14			9.5	4			5.8	11			4.8	2		
Index	FI			0.58	0.78			0.26	0.63			1.85	0.48			1.62	0.68
	CI			1.28	0.90			1.45	1.01			0.05	0.43			0.29	0.67
	CCI			0.35	0.79			0.60	0.70			−0.90	0.38			−0.67	0.65

Note: FI = falsification index. CI = confirmation index. CCI = consequent confirmation index.

$t(229) = 1.13$, $p = .13$, one-tailed. The failure to find an effect on the CIs may be due to the special status of the P only responses. The confirmation index does not take into account that selection of the P only pattern on indicative tasks may be based on analytic computations. A participant who selects the pattern gets a CI of +1. Consequently, a less frequent P only selection under load might restrict the CI increase. To bypass this complication a consequent confirmation index was calculated (Oaksford & Stenning, 1992). Oaksford and Stenning argued that because all explanations of the indicative selection task agree that participants should select the P card, attention should focus on the consequent cards. For the consequent confirmation index (CCI), +1 was added if a participant selected the Q card and −1 if they selected the not-Q card. As expected the mean CCI in the load group was significantly higher than that in the no-load group, $t(229) = 1.86$, $p < .035$, one-tailed.

Tapping performance. Mean tapping performance during the final 30-s practice tapping trial ($M = 2.30$, $SD = 0.64$) was compared with the tapping performance during the inference phase of the indicative selection task. The mean number of correct taps per second tended to decrease during indicative reasoning ($M = 2.16$, $SD = 0.74$) but the effect was not significant, $t(40) = 1.27$, $p = .11$, one-tailed. Remember that the indicative task was presented as the last reasoning task. The increased tapping performance during the indicative task probably results from a practice effect. More important is that, as with the conjunction problems, the decrease in tapping performance (i.e., mean correct taps per second during final practice minus mean correct taps per second during indicative reasoning) was not associated with the selection performance. Participants who selected the P and not-Q cards under load even tended to show a smaller decrease in tapping performance, $r = -.23$, $n = 41$, $p = .16$.

Deontic selection task

The dual-process framework claims that correct responding in the deontic selection task is based on mere heuristic processing. Given that the heuristic system is assumed to operate automatically, one expects no impact of the secondary task on the deontic "drinking age" problem. Table 4 shows the results. Contrary to the prediction, and as in Oaksford et al. (1996), the proportion of correct P and not-Q card selections decreased under load. The decreases in correct responses were compensated for by an increase in matching responses, cross-over interaction with matching pattern, $\chi^2(1, N = 207) = 6.49$, $p < .02$, but contrary to the indicative task also by an increased number of P only responses, cross-over interaction with correct pattern, $\chi^2(1, N = 210) = 6.55$, $p < .02$.

The analysis of the Pollard indices further confirmed the findings: FIs decreased under executive load, $t(229) = 2.55$, $p < .01$, one-tailed, whereas the CIs increased under load, $t(229) = 2.90$, $p < .001$, one-tailed.

It could be noted that although the heuristic system provides the correct response on the deontic task, one cannot exclude that some reasoners (e.g., the cognitively most gifted reasoners) might nevertheless rely on an analytic reasoning process to solve the task (Stanovich & West, 1998a, 2000). Thus, one might suggest that the executive load impact merely results from these participants. However, the dual-process framework conceptualizes the heuristic system as the default reasoning system. Even if some reasoners would use the analytic system to solve the deontic task when the executive resources are not burdened, an automatically operating heuristic "back-up" system would still provide them with the correct response under load. Thus, the decrease in correct deontic responses under load still conflicts with the assumed automatic nature of the deontic card selection.

Tapping performance. Compared with the final 30-s practice trial ($M = 2.30$, $SD = 0.64$), tapping performance decreased during the inference phase of the deontic selection task ($M = 2.06$, $SD = 0.78$), $t(40) = 2.07$, $p < .025$, one-tailed. Participants who selected the P and not-Q cards under load tended to show somewhat larger drops in tapping performance but the effect was not significant, $r = .06$, $n = 41$, $p = .72$. Thus, it is unlikely that correct reasoning on the deontic task results from a disproportional tapping-task neglect.

Analytic index

The analytic composite index consisted of the three problems for which giving the correct response was assumed to require analytic reasoning. One point was given for every correct conjunction response and one point for the selection of the P and not-Q pattern on the indicative selection task. The mean composite score in the no-load group was 0.52 ($SD = 0.75$). The composite score clearly decreased under executive load ($M = 0.26$, $SD = 0.57$), Mann–Whitney $U = 3{,}209$, $n_1 = 42$, $n_2 = 189$, $p < .015$, one-tailed.

Conclusion

Burdening the executive resources while participants solved the conjunction problems and indicative selection task resulted in a clear and systematic pattern: The rate of responses that were assumed to require analytic computations decreased under load, whereas the rate of responses that were assumed to be based on heuristic processing increased. This supports the basic dual-process framework claim that the operations of the analytic system draw on executive resources, whereas the heuristic system operates automatically. The increase in heuristic responses is consistent with the claim that the heuristic system is the dominant, default reasoning system. When analytic reasoning becomes too demanding, a reasoner will not merely make a random guess but will fall prey to the prepotent heuristic response.

Findings for the deontic selection task conflicted with the predictions. As in Oaksford et al. (1996), the executive load also affected deontic reasoning performance. The present experiment made sure that a standard deontic problem was adopted that was almost perfectly solved without executive load. Hence, the findings cannot merely be discarded based on suboptimal deontic task characteristics. This indicates that selection of the correct card pattern in the deontic task is not based on a purely automatic process.

However, it will be clear that the present findings should be interpreted with some caution. The experiment is the first one that adopted dual-task methodology with conjunction problems and indicative and standard deontic selection tasks. Procedural complications may always limit the contributions of the findings. Therefore, Experiment 3 (conjunction problems) and Experiment 4 (selection tasks) examine whether the present findings can be replicated with a different dual-task procedure.

EXPERIMENT 3

Experiment 2 showed that there was a marginally significant decrease in correct conjunction responses when a finger tapping task burdened the executive resources. Experiment 3 used a spatial storage task as a secondary task (e.g., Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001). Participants solved conjunction problems while they concurrently tried to remember a briefly presented visual dot pattern. Miyake et al. established that this spatial storage task burdened the central executive.

Experiment 2 compared the conjunction performance under executive task load with performance in a baseline group where participants did not perform a secondary task. Experiment 3 introduces a more appropriate control condition. Both the control and experimental groups had to remember a visual dot pattern. The complexity of the dot pattern was manipulated so that storage of the pattern in the control group would be less demanding.

Since the "Linda" and "Bill" problems have very similar tasks contents (i.e., judgements about personality traits of a specific individual) performance on the secondly presented "Bill" problem could have been affected by a practice effect (e.g., Reeves & Lockhart, 1993). Experiment 3 tried to minimize this problem. The "Bill" problem was replaced by the "job" problem (e.g., Reeves

& Lockhart). The “job” problem describes the distribution of event categories and bears no superficial relation to the “Linda” problem. Although reasoners make fewer conjunction fallacies on the “job” problem (Reeves & Lockhart, 1993; Stanovich & West, 1998a), Stanovich and West established that performance on this conjunction problem is also linked with individual differences in cognitive capacity.

Method

Participants
A total of 44 first-year psychology students from the University of Leuven, Belgium, participated in return for course credit. None of them had participated in the previous experiments. Participants were randomly selected for the present experiment but had previously participated in a screening for working memory capacity with the “gospan” task (see De Neys, d’Ydewalle, Schaeken, & Vos, 2002).

Materials
Conjunction tasks. Participants solved two conjunction problems: The “Linda” and “job” problems. The “Linda” problem was identical to the problem presented in Experiment 1, and the “job” problem was adapted from Reeves and Lockhart (1993) and Stanovich and West (1998b):

> *Lisa is in her twenties and jobless. She applied for three different part-time jobs. For the dress shop job, there are 7 other applicants, for the bookstore job, there are 5 other applicants, and for the job in the shoe-store, there is only 1 other applicant.*
>
> Please rank the following statements by their probability:
>
> A. Lisa will be offered the job in the shoe-store
>
> B. Lisa will be offered the dress shop job and the job in the shoe-store
>
> C. Lisa will be offered the dress shop job
>
> Most probable statement:_
>
> Second most probable statement:_
>
> Least probable statement:_

The problems were presented with the same instructions, practice, and filler item as those in Experiment 1. The “Linda” problem was always presented before the “job” problem. Reeves and Lockhart (1993) found that prior presentation of the “Linda” problem did not affect performance on the “job” problem.

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 reproduction phase an empty matrix was presented on the screen, and participants used the keypad to indicate the location of the dots. Each number key corresponded to a square in the matrix (from 1 = lower left, to 9 = upper right). When a number key was pressed, a dot appeared in the corresponding location. Pressing the same key once more removed the dot. Participants pressed the space bar when they finished reproducing the pattern.

In the load group the matrix was filled with complex four-dot patterns (i.e., “three-piece” patterns based on the work of Bethell-Fox & Shepard, 1988, and Verschueren, Schaeken, & d’Ydewalle, 2004, see Figure 1 for an example). Miyake et al. (2001) established that storage of similar complex dot patterns tapped executive resources. Verschueren et al. (2004) successfully used the task to burden executive resources in a

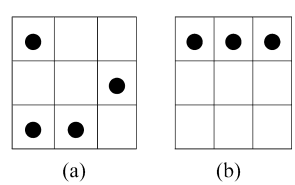


Figure 1. *Examples of the dot patterns in (a) the load group and (b) the control group.*

thinking-aloud study on everyday conditional reasoning.[3]

In the control group the patterns 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 at the same time in a large computer room with an individual booth for every participant. Participants 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) the response matrix was presented 1 s after the pattern had been presented. Participants next received the instructions for the conjunction task. The instructions stressed that it was crucial that the dot patterns were reproduced correctly in the upcoming reasoning task.

As in the previous experiments participants first read the preambles (italicized text) 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 ranking question and the answer alternatives. Participants entered the ranking themselves as in Experiment 1. Afterwards, the empty matrix was presented, and participants had to reproduce the dot pattern. Participants received feedback on whether the pattern had been reproduced correctly and were reminded that they had to try to remember the complete pattern correctly. The procedure was clarified with a practice item (i.e., the practice item of Experiment 1 and 2 presented with an easy three-dot pattern).

Results and discussion

Dot memory task

Results for the dot memory task indicated that the task was properly performed. The simple dot patterns in the control group were perfectly recalled by all participants. In the load group the mean number of correctly localized dots for the two complex patterns was 3.43 ($SD = 0.85$, range = 2 to 4). Thus, overall, about 86% of a complex dot pattern was reproduced correctly. Furthermore, the mean conjunction score in the load group was not associated with the mean dot recall score, $r = .12$, $n = 22$, $p = .61$. This indicates that participants were not merely trading off reasoning and dot recall performance.

Conjunction performance

Results for the conjunction tasks are summarized in Table 5. As expected, the conjunction fallacy was made less frequently for the "job" problem than for the "Linda" problem but the executive load clearly decreased the number of correct analytic responses on both problems. Thus, the basic finding of Experiment 2 was replicated. Performance on the two conjunction problems was combined into a composite total score (i.e., 1 point for every correct conjunction response, scores ranging from 0 to 2) for the statistical analysis of the load impact. Results established that correct analytic responses were less frequent in the load than in the control group, Mann–Whitney $U = 175$, $n_1 = 22$, $n_2 = 22$, $p < .045$, one-tailed.

Response latencies

Together with participants' rankings, response latencies were recorded. Because of the low number of correct responses in the load group the latencies were analysed independently from response accuracy. For every participant the mean reading and inference time on the two

[3] Based on the original Baddeley and Hitch (1974) working memory model one might expect that a spatial storage task would burden the visuospatial storage or "slave" system and not the central executive. However, recently it is becoming clear that, contrary to the verbal domain, the visuospatial storage system is closely tied to and might be indistinguishable from the central executive (e.g., Baddeley, Cocchini, Della Sala, Logie, & Spinnler, 1999; Miyake et al., 2001; Quinn & McConnell, 1996).

Table 5. *Percentage of correct conjunction responses, mean total score, and mean response latencies[a] with low and high concurrent executive load*

	Load group							
	Control				*Load*			
Measure	*% correct*	*n*	*M*	*SD*	*% correct*	*n*	*M*	*SD*
Linda problem	18	4			5	1		
Job problem	68	15			45	10		
Total score			0.86	0.71			0.50	0.58
Latency								
Reading time			17.50	6.01			18.20	5.86
Inference time			40.80	10.94			46.80	19.11

Note: Control = low load. Load = high load.

conjunction problems were calculated. These means were subjected to a 2 (phase, within subjects)×2 (load, between subjects) ANOVA. Table 5 presents the results. Participants spend less time reading than inferring, $F(1, 42) = 135$, $MSE = 109.3$, $p < .0001$. More important is that neither the load factor, $F(1, 42) = 1.48$, $MSE = 168.6$, $p = .023$, nor its interaction with the phase factor, $F(1, 42) = 1.40$, $MSE = 109.3$, $p = .24$, reached significance. Thus, the higher executive burden in the load group did not affect reading or inference times. The equal reading times in the load and control group establish that the decrease in analytic responses in the load group was not hampered by any strategic attempts to minimize the secondary-task burden.

Working memory

Participants were randomly assigned to the control and load group. Since there are only a limited number of participants in both groups one could argue that the randomization does not exclude group differences in WM capacity. Because differences in WM capacity are associated with the frequency of analytic responding this could confound the findings. However, participants' executive WM capacity had been previously measured with a version of the operation span task ("ospan", La Pointe & Engle, 1990) adapted for group testing ("gospan", for details see De Neys et al., 2002). Scores of 43 (out of 44) participants could be retrieved. An analysis on the mean scores in the load ($M = 36.68$, $SD = 10.55$) and control ($M = 33.52$, $SD = 9.37$) groups indicated that the groups did not differ in WM capacity, $t(41) = 1.05$, $p = .15$, one-tailed. Thus, the decrease in analytic responses in the load group cannot be attributed to group differences in executive capacity.

A final analysis examined the association between WM capacity and the conjunction performance. The analysis was not the focus of the present study and is reported for completeness. A proper correlational analysis would require a far greater number of participants. Nevertheless, it is interesting to note that, consistent with Stanovich and West (1998b), for the 21 participants in the control group, WM capacity tended to be positively associated with the total conjunction score, $r = .28$, $p = .23$. For the participants in the load group the correlation was somewhat less clear, $r = .19$, $n = 22$, $p = .4$. A smaller correlation in the load group would make sense since more cognitively gifted people will also start to reason heuristically under load, hence blurring the distinction with the less gifted reasoners.

Conclusion

Experiment 3 showed that analytic responding decreased when the executive resources were burdened by the memorization of a complex dot pattern. Several methodological changes were introduced, and a new secondary-task procedure was used. Nevertheless, the basic conjunction task findings of Experiment 2 were replicated. This indicates that the results are robust.

EXPERIMENT 4

Experiment 2 supported the predictions of the dual-process framework on the indicative selection task: Burdening the executive resources with the complex tapping task boosted heuristic, matching responses and decreased correct, analytic responses. However, contrary to the dual-process prediction the secondary-task load also affected deontic reasoning performance. The present experiment adopted the dot memory task procedure in an attempt to replicate the selection findings.

Participants in the control group memorized the systematic three-dot pattern while reasoning, whereas the load group had to memorize the complex four-dot patterns. In Experiment 4 the indicative and deontic selection tasks were also presented to two different groups of participants in order to minimize possible training effects.

Method

Participants

The participants were 186 first-year psychology students from the University of Leuven, Belgium, who participated in return for course credit. None of them had participated in the previous experiments. Participants were randomly selected for the present experiment but had previously participated in a screening for working memory capacity with the gospan task (see De Neys et al., 2002).

Materials

Approximately half of the participants were presented the deontic "drinking age" task as used in Experiment 1. The other half were presented an indicative selection task based on Stanovich and West's (1998a) "coin" problem. The indicative task had a similar structure to that of the indicative tasks in the previous experiments but a different content. The task referred to the names of non-existing painters to make sure that participants had no prior knowledge about the relation that the conditional rule expressed:

> *Below you will get to see 4 cards. Each card has the name of a painter on one side and a building on the other side. Your task is to decide which cards you have to turn over in order to decide whether or not a specific rule is being followed. The rule states that if the painter is BELTRAN, the building is a CHURCH.*
>
> Which of the following cards do you have to turn to find out whether the rule "If BELTRAN is on one side of the card, then CHURCH is on the other side" is being violated?
>
> | Painter: CUNZZI | Painter: BELTRAN | Building: CHURCH | Building: HOUSE |
> |---|---|---|---|
> | A. | B. | C. | D. |
>
> —Type down the letters of the cards that need to be turned. Press space bar when finished—

The tasks were presented with the same instructions and practice item as those in the previous selection experiments.

Procedure

Participants were tested in groups of 12 to 21 at the same time in a large computer room with an individual booth for every participant. Participants were randomly assigned to the control or load group. Approximately half of the participants in each group solved the indicative task whereas the other half solved the deontic task. The procedure was similar to Experiment 3. Participants started with the two practice storage items followed by the practice selection task. As in the previous experiments, participants first read the preambles (italicized text) 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 selection rule and cards. Participants selected the cards as in Experiment 1. Afterwards, the empty matrix was presented, and participants had to reproduce the dot pattern.

Results and discussion

The indicative selection task was presented to 46 participants in the control group and 47 participants in the load group. The deontic problem was presented to 47 participants in the control group and 46 participants in the load group. Participants' executive WM capacity had previously been measured with a version of the operation span task (ospan, La Pointe & Engle, 1990) adapted for group testing (gospan, for details see De Neys et al., 2002). Test scores of 184 participants were available. A first control analysis established that the mean WM capacity scores did not differ in the four experimental groups, $F(3, 180) = 1.64$, $MSE = 112.96$, $p = .18$. Mean score of the complete sample was 35.67 ($SD = 10.68$).

Indicative selection task

Dot memory task. Results for the dot memory task indicated that participants managed to memorize the dot pattern during indicative reasoning. In the control group ($n = 46$) a mean number of 2.48 dots ($SD = 1.15$, range 0–3), or 83% of the 3-dot pattern, were reproduced correctly. In the load group ($n = 47$) the mean number of correctly localized dots was 3.45 ($SD = 0.90$, range = 1 to 4), or 86% of the 4-dot pattern. Although participants who selected the correct P and not-Q pattern tended to remember somewhat fewer dots for the easy control patterns, $r = -.21$, $n = 46$, $p = .17$, the negative association did not reach significance and was absent in the crucial load group, $r = .09$, $n = 47$, $p = .54$. This indicates that participants were not merely trading off reasoning and dot recall performance.

Card selections. Table 6 shows that about 11% of the participants in the control group selected the correct P and not-Q pattern on the indicative selection task. When the executive resources were burdened by the memorization of the complex dot pattern, only 2% gave the correct response. As Table 6 shows, the decrease in analytic P and not-Q responses under load was specifically compensated for by an increase in matching responses. Comparing the frequencies of both responses in the control and load group showed that this cross-over interaction was marginally significant, $\chi^2(1, N = 46) = 3.49$, $p < .065$. As in Experiment 2, the selection rate of the P only pattern also tended to decrease under load but the cross-over interaction with the matching responses did not reach significance, $\chi^2(1, N = 81) = 1.01$, $p = .32$. There were no clear trends for any of the other selection patterns.

The analysis of the Pollard indices further replicated the findings of Experiment 2. As Table 6 shows, the FIs decreased in the load group, $t(91) = 1.99$, $p < .025$, one-tailed. There was also an increase in the CIs under load but the effect did not reach significance, $t(91) = 1.06$, $p = .15$, one tailed. As in Experiment 2, the tendency towards increased heuristic responding under load was clearer for the CCIs, $t(91) = 1.79$, $p < .04$, one-tailed.

An examination of the effect sizes (Cohen's d) on the Pollard indices gave a final indication of the consistency of the findings. In Experiment 2, d reached .41 on the FIs, .19 on the CIs, and .32 on the CCIs, which is very close to the .41 (FIs), .22 (CIs), and .37 (CCIs) values observed in the present experiment.

Response latencies. A 2 (phase, within subjects) $\times$ 2 (load, between subjects) ANOVA on the indicative response latencies indicated that none of the factors reached significance, all $F(1, 91) < 1$. Thus, as in Experiment 3, the higher executive burden in the load group did not affect reading or inference times.

WM capacity correlations. Consistent with Stanovich and West's (1998a) findings, overall WM capacity was positively associated with the indicative falsification index, $r = .24$, $n = 92$, $p < .025$. As with the conjunction tasks in Experiment 3, the correlation tended to be

Table 6. *Proportion of card selections, mean falsification and confirmation indices, and mean response latencies*[a] *with low and high concurrent executive load*

		Indicative								*Deontic*							
		Control				*Load*				*Control*				*Load*			
Measure		%	*Freq.*	*M*	*SD*	%	*Freq.*	*M*	*SD*	%	*Freq.*	*M*	*SD*	%	*Freq.*	*M*	*SD*
Card pattern	Correct	10.9	5			2.1	1			76.6	36			71.7	33		
	Matching	36.9	17			48.9	23			4.3	2			2.2	1		
	P only	47.8	22			40.4	19			19.1	9			19.6	9		
	Other	4.4	2			8.6	4			0	0			6.5	3		
Index	FI			0.67	0.70			0.38	0.71			1.77	0.48			1.70	0.51
	CI			1.20	0.81			1.36	0.71			0.23	0.48			0.17	0.57
	CCI			0.26	0.68			0.49	0.55			−0.77	0.48			−0.76	0.48
Latency	Reading			21.04	9.14			21.35	8.93			38.99	10.48			43.11	11.59
	Inference			20.59	12.94			20.08	12.23			22.17	14.31			26.79	14.78

Note: Control = low load. Load = high load. FI = falsification index. CI = confirmation index. CCI = consequent confirmation index.
[a]In s.

somewhat higher in the control, $r = .25$, $n = 46$, $p < .1$, than in the load group, $r = .18$, $n = 46$, $p = .24$.

Deontic selection task

Dot memory task. As with the previous reasoning tasks, participants who were presented the deontic selection task properly performed the secondary dot memory task. In the control group ($n = 47$) a mean number of 2.94 dots ($SD = 0.44$, range 0–3), or 98% of the 3-dot pattern, was reproduced correctly. In the load group ($n = 46$) the mean number of correctly localized dots was 3.63 ($SD = 0.65$, range = 1 to 4), or 91% of the 4-dot pattern. Participants were not trading off reasoning and dot recall performance: As with the indicative selection task, participants who selected the correct P and not-Q pattern tended to remember somewhat fewer dots for the easy control patterns, $r = -.08$, $n = 47$, $p = .59$, but the negative trend did not reach significance and was absent in the crucial load group, $r = .24$, $n = 46$, $p = .11$.

Card selections. Results are presented in Table 6. Card selections were clearly not random; the vast majority of participants selected the correct pattern, but the selection pattern hardly differed in the control and load groups. Although there was a slight decrease of about 5% in correct, P and not-Q responses under load, the decrease was not accompanied by an increase in matching responses, and the cross-over interaction was not significant, $\chi^2(1, N = 72) = 0.24$, $p = .66$.

The analysis of the Pollard indices further confirmed that the deontic card selections in the present experiment were not affected by the executive load. Neither the FIs, $t(91) = 0.69$, $p = .25$, one-tailed, $d = .14$, nor the CIs, $t(91) = 0.55$, $p = .28$, one-tailed, $d = .11$, or CCIs, $t(91) = 0.05$, $p = .48$, one-tailed, $d = .01$, differed in the two groups. Not surprisingly, effect sizes also differed considerably in comparison with Experiment 2 where d reached .44 on the FIs, and .50 on the CIs.

Response latencies. Deontic response latencies were subjected to a 2 (phase, within subjects) × 2 (load, between subjects) ANOVA. Not surprisingly, given the lengthy preambles, reading times were longer than inference times, $F(1, 91) = 122.42$, $MSE = 104.27$, $p < .0001$. There was a marginal main effect of the load factor with a tendency for slower responses in the load group, $F(1, 91) = 3.87$, $MSE = 229.19$, $p < .055$, and the two factors did not interact, $F(1, 91) < 1$. The marginal effect of the load factor indicates that reading tended to be somewhat slower in the load group. However, since participants received only a single item (with the dot pattern presented after the reading phase) it is highly unlikely that the longer reading times in the load group would follow from a strategy to minimize the secondary-task burden.

WM capacity correlations. Consistent with Stanovich and West (1998a) there was no clear correlation between WM capacity and the deontic falsification index in the control group, $r = .02$, $n = 46$, $p = .92$. However, contrary to the pattern on the conjunction and indicative tasks, the correlation was stronger in the load group, $r = .26$, $n = 46$, $p < .085$. Thus, when the executive resources were heavily burdened, individual differences in the number of available resources tended to become important for deontic reasoning. A possible explanation for this finding is presented in the General Discussion. The overall correlation between WM capacity and the deontic falsification index was $r = .13$, $n = 92$, $p = .22$.

Conclusion

Results of Experiment 4 were consistent with the dual-process predictions. Analytic responding on the indicative selection task decreased under executive load whereas matching responses increased. Contrary to Experiment 2, burdening the executive resources during deontic reasoning had no effect on the card selections.

GENERAL DISCUSSION

The present study aimed to validate basic assumptions about the processing characteristics of the heuristic and analytic reasoning systems in dual-process theories of reasoning. Four experiments tested the claim that the heuristic reasoning system operates fast and automatically whereas the operations of the analytic system are time consuming and draw on executive resources. Experiment 1 presented direct chronometric evidence for the processing-speed assumption. Responses that were assumed to require analytic computations took more inference time than responses that were assumed to be based on heuristic processing: Inferences on the conjunction problems were slower for people who responded correctly than for people who fell prey to the conjunction fallacy. Likewise, participants needed less inference time for matching card selections than for selection of the correct P and not-Q cards on the indicative selection task. Consistent with dual-process predictions, correctly selecting the P and not-Q cards also took more time in the indicative selection task than in the deontic selection task.

Experiment 2 adopted secondary-task methodology to test the processing resources claim. Burdening the executive resources with an attention-demanding secondary task while participants solved the conjunction problems and indicative selection task resulted in a clear and systematic pattern: Responses that were assumed to require analytic computations decreased under load, whereas responses assumed to be based on heuristic processing increased. The executive burden decreased correct responding and boosted the rate of conjunction fallacies and matching card selections. The same pattern was observed with a different secondary-task procedure in Experiments 3 and 4. This provides new support for the basic dual-process framework claim that the operations of the analytic system draw on executive resources, whereas the heuristic system operates automatically. The increase in heuristic responses is consistent with the claim that the heuristic system is the dominant, default reasoning system. When analytic reasoning becomes too demanding, a reasoner will not merely make a random guess but will fall prey to the prepotent heuristic response. The findings may serve to complete the Stanovich and West (e.g., 1998a, 1998b, 1998c, 2000, 2003) research programme on individual differences. The present study indicates that making erroneous, heuristic inferences is not merely associated with but directly caused by limitations in executive resources.

It might be noted that the secondary-task effects in the individual experiments were not very strong. One should bear in mind, however, that performance on the conjunction and indicative selection tasks is already quite floored in the control groups. One of the very reasons that the tasks have occupied such a central place in the field is precisely the dramatic low rate of correct responses. Thus, by definition, the decrease in performance can never be spectacular. What is important concerning the reliability of the results is that the effects are consistently replicated. The fact that the same trends were observed in two independent experiments provides additional support for the findings.

There were no consistent secondary-task effects on the deontic selection task in the present study. Burdening the executive resources in Experiment 2 did affect deontic reasoning performance, but the effect was not replicated in Experiment 4. The dual-process framework generally assumes that correctly solving the deontic task does not require analytic operations. The heuristic tendency to contextualize a problem with prior semantic knowledge would automatically trigger the selection of the correct cards. The inconsistent deontic load effects indicate that this claim should be interpreted with caution.

As in Stanovich and West (1998a), Experiment 4 showed there was no correlation between WM capacity and deontic reasoning performance in the control condition. Contrary to the indicative selection task and conjunction problems, however, the correlation was boosted under executive load. Together with the inconsistent nature of the secondary-task impact this finding might indicate that correctly solving the deontic problem does actually require a minimal amount of executive involvement. The involvement would be minimal in the sense that in a normal population of university students, under normal reasoning conditions, nearly everyone would have sufficient resources to solve the task. Therefore, individual differences in executive resources would not predict performance. However, a minimal resource pool will not suffice when a demanding secondary task also usurps executive resources. Thus, for reasoners with minimal executive capacity the load might nevertheless affect deontic reasoning.[4] This idea is supported by the recent work of Newstead et al. (2004). In a normal sample of university students Newstead et al. replicated the finding that there was no correlation between cognitive ability and deontic selection performance. The correlation returned, however, in a sample that was overall lower in cognitive capacity (i.e., lower than the test norms for university students). Thus, when the relative executive burden is increased, either by reducing the overall capacity of the sample in Newstead et al. or by demanding concurrent secondary-task performance in the present study, individual differences in executive resources do seem to become important for deontic reasoning. If correctly solving the deontic reasoning task would be completely automatic, such mediation would not be expected. Thus, a possible contribution of executive resources to correct deontic card selection should not be discarded although it is clear that contrary to the indicative task the role of the executive resources will only be minimal. The limited executive demands would still result in faster responses than would the indicative task but will make it harder to detect a consistent secondary-task effect.

From a theoretical point of view it is not unwarranted to assume that correctly solving the deontic task might require some executive resources. Dual-process theories take it for granted that reasoning on the basis of stored background knowledge is completely automatic. However, even in the "drinking age" problem reasoners will still have to access their long-term memory to retrieve the knowledge that 14-year-olds are not supposed to drink. Memory studies show that although much of the long-term memory retrieval is spontaneous and automatic, retrieval efficiency can be boosted by an active, WM-dependent strategic search (e.g., De Neys, Schaeken, & d'Ydewalle, 2005; Moscovitch, 1995; Rosen & Engle, 1997). Thus, one possible contribution of the executive resources in deontic reasoning might lie in an efficient retrieval of the necessary background knowledge. This hypothesis remains speculative of course and will require further testing.

The present study bears some relevance for more general issues in the WM field. It has been argued that WM capacity and general reasoning ability represent one and the same construct in our cognitive architecture (e.g., Engle et al., 1999; Kyllonen & Cristal, 1990; Süss, Oberauer, Wittmann, Wilhelm, & Schulze, 2002). There is indeed clear evidence (e.g., see Engle et al.; Kyllonen & Cristal; Süss et al.) for a tight link between WM capacity and performance in such basic reasoning tasks as analogy solving (e.g., analogies of the form "A:B = C: ?" in which the elements are geometric patterns or words), mental arithmetic (e.g., "If 3X = 18, then X = ?", or detecting the rule in a number sequence), relational reasoning (e.g., "Dick is better than Pete. Jake is worse than Pete. Who's best?"), or solving nonsense syllogisms (e.g., "All trees are fish. All fish are horses. Thus, are all trees horses?"). A number of secondary-task studies (e.g., Gilhooly, Logie, & Wynn, 1999; Halford, Bain, & Maybery, 1984; Meiser, Klauer, & Naumer, 2001; Toms, Morris, & Ward, 1993) also confirmed the crucial role of the WM resources in these tasks. Together with Stanovich

[4] The suggestion assumes that especially the lower WM spans should show some secondary-task impact on the deontic task. Therefore, deontic performance of participants in the top half (high span) and bottom half (low span) of the WM capacity distribution in Experiment 4 was compared. Interestingly, consistent with the prediction, high span's FI was similar in the control ($M = 1.8$, $SD = 0.42$) and load groups ($M = 1.8$, $SD = 0.53$). For the low spans, however, the executive burden in the load group ($M = 1.5$, $SD = 0.62$) did tend to decrease the FIs compared to control ($M = 1.8$, $SD = 0.43$), $t(38) = 1.31$, $p < .095$, one-tailed.

and West's (e.g., 1998c) correlational findings the present study made it clear that the WM capacity involvement generalizes to more complex reasoning tasks such as the conjunction and selection tasks typically studied in the heuristics and biases literature and the dual-process framework. The hallmark of these tasks is that there is a clear, dominant heuristic response that massively biases performance and that will need to be overridden by the analytic system in order to give the correct response. The present data showed that the specific errors that people make under executive load are not merely random but precisely the responses triggered by the heuristic system. Hence, it seems that the executive burden hampered an efficient blocking of the heuristic response. Thereby the findings also support the tendency to link aspects of executive functioning with notions of inhibitory control (e.g., Barkley, 1998; Dempster & Corkill, 1999; Harnishfeger & Bjorklund, 1994; Kokis et al., 2002).

As in most dual-process theories, the present study referred to the responses that are sanctioned by the traditional normative standards (e.g., probability theory or propositional logic) as correct, analytic responses. The validity of these norms for the kind of everyday reasoning evoked in conjunction and selection tasks has been the topic of intense debate (see Oaksford & Chater, 1998; Stanovich & West, 2000; Stein, 1996, for discussions). The present study took a neutral stance on this issue. The experiments show that heuristic and analytic responses have a different time course and draw on different processing resources. No claims are made about the quality of the reasoning process. The fact that Stanovich and West's individual differences studies showed that it are precisely people higher in executive ability that do adhere to these norms has been interpreted as support for the norm validity (Evans, 2002; Stanovich & West, 2000). Although the present study did not focus on the norm validity debate it might be noted that the experimental evidence for the role of executive resources can strengthen Stanovich and West's position. Given the current findings it is unlikely that the link between adhering to the normative standards and executive capacity would merely arise, for example, from an educational confound (pace, Schneider, 2000). This must at least give pause for thought before completely discarding the notion of logic-based normative rationality (Evans, 2002).

The present study presented new evidence for the posited processing characteristics of the analytic and heuristic systems. Establishing the basic processing characteristics of the two reasoning systems is an important first step toward an algorithmic-level specification of the dual-process framework (Stanovich & West, 1998c, 2000). Based on the present work more fine-grained future analyses can focus on which specific executive functions are involved in which part of the analytic process in a given reasoning task. One example concerns the precise locus of the executive burden. Dual-process theories assume that executive resources are required both for the inhibition of the default heuristics and subsequent analytic computations. Whereas the presently observed decrease in analytic responses under executive load supported this claim the study did not yet distinguish the precise load impact on each of the two components. By stimulating such specific analyses the present findings can help pave the way for a further development of dual-process theories of reasoning.

Original manuscript received 26 August 2004
Accepted revision received 1 March 2005
PrEview proof published online 8 August 2005

REFERENCES

Baddeley, A. D., Cocchini, G., Della Sala, S., Logie, R. H., & Spinnler, H. (1999). Working memory and vigilance: Evidence from normal aging and Alzheimer's disease. *Brain and Cognition, 41*, 87–108.

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.

Barkley, R. A. (1998). Behavioral inhibition, sustained attention, and executive functions: Constructing a unifying theory of ADHD. *Psychological Bulletin, 121*, 65–94.

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.

Cheng, P. W., & Holyoak, K. J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology, 17*, 391–416.

Cowan, N. (1995). *Attention and memory: An integrated framework.* New York: Oxford University Press.

Dempster, F. N., & Corkill, A. J. (1999). Individual differences in susceptibility to interference and general cognitive ability. *Acta Psychologica, 101*, 395–416.

De Neys, W. (2003). *Memory search for stored counterexamples during conditional reasoning.* Unpublished doctoral dissertation, University of Leuven, Belgium.

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.

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2005). Working memory and counterexample retrieval for causal conditionals. *Thinking and Reasoning, 11*, 123–150.

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.

Epstein, S. (1994). Integration of the cognitive and psychodynamic unconscious. *American Psychologist, 49*, 709–724.

Evans, J. St. B. T. (1977). Toward a statistical theory of reasoning. *Quarterly Journal of Experimental Psychology, 29*, 297–306.

Evans, J. St. B. T. (1984). Heuristic and analytic processes in reasoning. *British Journal of Psychology, 75*, 451–468.

Evans, J. St. B. T. (1996). Deciding before you think: Relevance and reasoning in the selection task. *British Journal of Psychology, 87*, 223–240.

Evans, J. St. B. T. (1998a). Inspection times, relevance, and reasoning: A reply to Roberts. *Quarterly Journal of Experimental Psychology, 51A*, 811–814.

Evans, J. St. B. T. (1998b). Matching bias in conditional reasoning: Do we understand it after 25 years? *Thinking and Reasoning, 4*, 45–82.

Evans, J. St. B. T. (2002). Logic and human reasoning: An assessment of the deduction paradigm. *Psychological Bulletin, 128*, 978–996.

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., & Lynch, J. S. (1973). Matching bias in the selection task. *British Journal of Psychology, 64*, 391–397.

Evans, J. St. B. T., Newstead, S. E., & Byrne, R. M. J. (1993). *Human reasoning: The psychology of deduction.* Hove, UK: Lawrence Erlbaum Associates Ltd.

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

Evans, J. St. B. T., & Over, D. E. (1997). Rationality in reasoning: The problem of deductive competence. *Current Psychology of Cognition, 16*, 3–38.

Feldman Barrett, L., Tugade, M. M., & Engle, R. W. (2004). Individual differences in working memory capacity and dual-process theories of the mind. *Psychological Bulletin, 130*, 553–573.

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

Gilhooly, K. J., Logie, R. H., & Wynn, V. (1999). Syllogistic reasoning tasks, working memory, and skill. *European Journal of Cognitive Psychology, 11*, 473–498.

Goel, V. (1995). *Sketches of thought.* Cambridge, MA: MIT Press.

Griggs, R. A., & Cox, J. R. (1982). The elusive thematic materials effect in the Wason selection task. *British Journal of Psychology, 72*, 407–420.

Halford, G. S., Bain, J. D., & Maybery, M. T. (1984). Does a concurrent memory load interfere with reasoning? *Current Psychological Research & Reviews, 3*, 14–23.

Harnishfeger, K. K., & Bjorklund, D. F. (1994). A developmental perspective on individual differences in inhibition. *Learning and Individual Differences, 6*, 331–356.

Hertwig, R., & Gigerenzer, G. (1999). The "conjunction fallacy" revisited: How intelligent inferences look like reasoning errors. *Journal of Behavioral Decision Making, 12*, 275–305.

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.

Just, M. A., Carpenter, P. A., & Keller, T. A. (1996). The capacity theory of comprehension: New frontiers of evidence and arguments. *Psychological Review, 103*, 773–780.

Kahana, M., & Loftus, G. (1999). Response time versus accuracy in human memory. In R. J. Sternberg (Ed.), *The nature of cognition* (pp. 323–340). Cambridge, MA: MIT Press.

Kahneman, D. (2000). A psychological point of view: Violations of rational rules as a diagnostic of mental processes. *Behavioral and Brain Sciences, 23*, 681–683.

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

Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of working memory capacity. *Journal of Experimental Psychology: General, 130*, 169–183.

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). Analytic and heuristic processing influences on adolescent reasoning and decision making. *Child Development, 72*, 844–861.

Kokis, J. V., Macpherson, R., Toplak, M. E., West, R. F., & Stanovich, K. E. (2002). Heuristic and analytic processing: Age trends and associations with cognitive ability and cognitive styles. *Journal of Experimental Child Psychology, 83*, 26–52.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working memory capacity. *Intelligence, 14*, 389–433.

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.

Logan, G. D., Taylor, S. E., & Etherton, J. L. (1996). Attention in the acquisition and expression of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 620–638.

Manktelow, K. I. (1999). *Reasoning and thinking.* Hove, UK: Psychology Press.

Margolis, H. (1987). *Patterns, thinking, and cognition.* Chicago: University of Chicago Press.

McElree, B. (2001). Working memory and focal attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 817–835.

Meiser, T., Klauer, K. C., & Naumer, B. (2001). Propositional reasoning and working memory: The role of prior training and pragmatic content. *Acta Psychologica, 106*, 303–327.

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.

Miyake, A., & Shah, P. (Eds.). (1999). *Models of working memory: Mechanisms of active maintenance and executive control.* Cambridge, MA: Cambridge University Press.

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.

Moscovitch, M. (1995). Models of consciousness and memory. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 1341–1356). Cambridge, MA: MIT Press.

Newstead, S. E., Handley, S. J., Harley, C., Wright, H., & Farrelly, D. (2004). Individual differences in deductive reasoning. *Quarterly Journal of Experimental Psychology, 57A*, 33–60.

Newton, E. J., & Roberts, M. J. (2003). Individual differences transcend the rationality debate. *Behavioral and Brain Sciences, 26*, 530–531.

Oaksford, M., & Chater, N. (1998). *Rationality in an uncertain world: Essays on the cognitive science of human reasoning.* Hove, UK: Psychology Press.

Oaksford, M., Morris, F., Grainger, B., & Williams, J. M. G. (1996). Mood, reasoning, and central executive processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 476–492.

Oaksford, M., & Stenning, K. (1992). Reasoning with conditionals containing negated constituents. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 835–854.

Osman, M. (2002). Is there evidence for unconscious reasoning processes? *Proceedings of Annual Meeting of the Cognitive Science Society, 24*, 732–737.

Pollard, P., & Evans, J. St. B. T. (1987). Content and context effects in reasoning. *American Journal of Psychology, 100*, 41–60.

Quinn, J. G., & McConnell, J. (1996). Irrelevant pictures in visual working memory. *Quarterly Journal of Experimental Psychology, 49A*, 200–215.

Reeves, T., & Lockhart, R. S. (1993). Distributional versus singular approaches to probability and errors in probabilistic reasoning. *Journal of Experimental Psychology: General, 122*, 207–226.

Roberts, M. J. (1998). Inspection times and the selection task: Are they relevant? *Quarterly Journal of Experimental Psychology, 51A*, 781–810.
Roberts, M. J., & Newton, E. J. (2001). Inspection times, the change task, and the rapid-response task. *Quarterly Journal of Experimental Psychology, 54A*, 1031–1048.
Rosen, V. M., & Engle, R. W. (1997). The role of working memory capacity in retrieval. *Journal of Experimental Psychology: General, 126*, 211–227.
Schneider, S. L. (2000). An elitist naturalistic fallacy and the automatic-controlled continuum. *Behavioral and Brain Sciences, 23*, 695–696.
Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84*, 127–190.
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: Lawrence Erlbaum Associates, Inc.
Stanovich, K. E., & West, R. F. (1998a). Cognitive ability and variation in selection task performance. *Thinking and Reasoning, 4*, 193–230.
Stanovich, K. E., & West, R. F. (1998b). Individual differences in framing and conjunction effects. *Thinking and Reasoning, 4*, 289–317.
Stanovich, K. E., & West, R. F. (1998c). Individual differences in rational thought. *Journal of Experimental Psychology: General, 127*, 161–188.
Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate. *Behavioral and Brain Sciences, 23*, 645–726.
Stanovich, K. E., & West, R. F. (2003). The rationality debate as a progressive research program. *Behavioral and Brain Sciences, 26*, 531–534.
Stein, E. (1996). *Without good reason: The rationality debate in philosophy and cognitive science.* Oxford, UK: Oxford University Press.
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.
Thompson, V. A., Striemer, C. L., Reikoff, R., Gunter, R. W., & Campbell, J. I. D. (2003). Syllogistic reasoning time: Disconfirmation disconfirmed. *Psychonomic Bulletin & Review, 10*, 184–189.
Toms, M., Morris, N., & Ward, D. (1993). Working memory and conditional reasoning. *Quarterly Journal of Experimental Psychology, 46A*, 679–699.
Torrens, D., Thompson, V. A., & Cramer, K. M. (1999). Individual differences and the belief bias effect: Mental models, logical necessity, and abstract reasoning. *Thinking and Reasoning, 5*, 1–28.
Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review, 90*, 293–315.
Valentine, E. R. (1975). Performance on two reasoning tasks in relation to intelligence, divergence and interference proneness. *British Journal of Educational Psychology, 45*, 198–205.
Verschueren, N., Schaeken, W., & d'Ydewalle, G. (2004). Everyday conditional reasoning with working memory preload. *Proceedings of the Annual Meeting of the Cognitive Science Society.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
Wason, P. C. (1966). Reasoning. In B. M. Foss (Ed.), *New horizons in psychology: I* (pp. 106–137). Harmandsworth, UK: Penguin.