

Methodological pitfalls of the Unconscious Thought paradigm

Laurent Waroquier*, David Marchiori, Olivier Klein and Axel Cleeremans

Université Libre de Bruxelles

Abstract

According to Unconscious Thought Theory (UTT: Dijksterhuis & Nordgren, 2006), complex decisions are best made after a period of distraction assumed to elicit “unconscious thought”. Over three studies, respectively offering a conceptual, an identical and a methodologically improved replication of Dijksterhuis et al. (2006), we reassessed UTT’s predictions and dissected the decision task used to demonstrate these predictions. We failed to find any evidence for the benefits of unconscious decision-making. By contrast, we found some evidence that conscious deliberation can lead to better decisions. Further, we identified methodological weaknesses in the UTT decision task: (a) attributes weighting was neglected although attributes were seen as different in importance; (b) the material was not properly counterbalanced; and (c) there was some confusion in the experimental instructions. We propose methodological improvements that address these concerns. Keywords: unconscious thought, conscious thought, decision-making.

1 Introduction

Choices are fundamental in human existence. Throughout life we make decisions that range from mundane everyday choices, such as selecting a particular brand of cereal or a route to go to work, to life-changing choices in selecting a partner, a house, or a career. Decision making, of course, is not limited to individuals. Rather, it is also central to the structure of societies, at different levels of organization. Governments make choices that engage entire nations. Courts of law render judgments that influence the lives of many. Corporations make “business decisions” that affect the welfare of entire regions.

According to classical approaches to decision making (e.g., Simon, 1955), in situations where decision makers face difficult choices, they should think thoroughly about the different alternatives and ponder the positive and negative aspects of each option. However, a different perspective on decision making was proposed in “Unconscious Thought Theory” (UTT: Dijksterhuis & Nordgren, 2006). UTT suggests that when people are dealing with complex choices, they should simply “sleep on it” and

thus avoid engaging in conscious consideration of the different alternatives. In this way, the unconscious would process the information more efficiently than would be possible through conscious reflection. Specifically, UTT makes the following recommendation:

“When faced with complex decisions such as where to work or where to live, do not think too much consciously. Instead after a little initial conscious information acquisition, avoid thinking about it consciously. Take your time and let the unconscious deal with it” (Dijksterhuis, 2004, p. 596).

On the one hand, UTT’s advice is intuitively appealing. Almost everybody has experienced the beneficial effects of idling before taking important decisions. On the other hand, this suggestion flies in the face of the Cartesian tradition that would have people ponder and think the problem through with great care. For these reasons, UTT has attracted considerable interest, as evidenced by research featured in the most prominent journals (Dijksterhuis, 2004; Dijksterhuis, Bos, Nordgren & Van Baaren, 2006).

Although some empirical findings support the intuition that an unconscious process takes place when attention is directed elsewhere and improves complex decisions’ accuracy (e.g., Dijksterhuis, 2004, Dijksterhuis, et al. 2006; Lerouge, 2009), several studies have failed to replicate this effect or proposed alternative interpretations (e.g., Calvillo & Penaloza, 2009 ; Thorsteinson & Withrow, 2009 ; Waroquier, Marchiori, Klein, Cleeremans, in press; for a meta-analysis see: Acker, 2008; see also Strick, Dijksterhuis, Bos, Sjoerdma, van Baaren, & Nordgren, 2009).

*This research was supported by a mini-ARC grant from the Université Libre de Bruxelles to L.W. and O.K., by grant BFR 07/052 from the “Ministère luxembourgeois de la Culture, de l’Enseignement Supérieur et de la Recherche” to D. M. by Concerted Research Action 06/11–342 titled “Culturally modified organisms: What it means to be human in the age of culture”, financed by the Ministère de la Communauté Française — Direction Générale de l’Enseignement non obligatoire et de la Recherche scientifique (Belgium) to AC and OK., an institutional grant from the Université Libre de Bruxelles to A.C., and by European Commission Grant #043457 “Mindbridge — Measuring Consciousness” to A.C. Address: Laurent Waroquier, Unité de Psychologie Sociale, Université Libre de Bruxelles 50 avenue Franklin Roosevelt 1050 Brussels, Belgium. E-mail: lwaroqui@ulb.ac.be.

Across three studies, respectively offering a conceptual, an identical and a methodologically improved replication of Dijksterhuis et al. (2006), we reassessed UTT's predictions and dissected the decision task used to demonstrate these predictions. We failed to find any evidence for the benefits of unconscious decision-making. By contrast, we found some evidence that conscious deliberation can improve complex decisions. Further, we identified methodological weaknesses in the UTT decision task. In a nutshell, attributes weighting was neglected although attributes were seen as different in importance, the material was not properly counterbalanced and there was some confusion in the experimental instructions. We propose methodological improvements that circumvent these problems.

1.1 What is unconscious thought?

UTT distinguishes two modes of thought. Conscious thought refers to processes that occur while the problem at hand is the focus of conscious attention. By contrast, unconscious thought is said to take place while conscious attention is diverted from the problem. This active process is said to improve decision quality when dealing with complex problems (e.g., Dijksterhuis & Nordgren, 2006). During the period of unconscious processing, the information would be organized, weighted, and integrated in memory resulting in a clearer and more polarized evaluation of the different alternatives (e.g., Dijksterhuis & Nordgren, 2006). Contrary to previous work (e.g., Kahneman & Frederick 2002), unconscious thought is described as a complex, time consuming and goal-dependent mechanism (Bos, Dijksterhuis & van Baaren, 2008). Further, according to UTT, deliberation could occur without attention (Dijksterhuis et al., 2006, see also Evans, 2008) while conscious thought would be of heuristic nature (Dijksterhuis & Nordgren, 2006).

The purported superiority of unconscious thought in matter of complex decisions is related to the capacity and weighting principles described in the UTT. The capacity principle states that unconscious thought has much larger capacity than conscious thought. In this respect, performing a distraction task rather than deliberating consciously would allow a person to take more attributes into account and therefore enhance the quality of complex decisions. Since this principle has not received strong empirical support (see: Dijksterhuis, 2004), it has been examined in Experiments 2 and 3. The weighting principle states that while the unconscious naturally weights the relative importance of alternatives' attributes, conscious thought disturbs this natural process. The literature does not offer strong empirical support for this principle either (see: Dijksterhuis, 2004). In fact, most studies addressing this question suggest that thinking consciously leads

to similar or superior weighting than performing a distraction task (Newell, Wong, Cheung & Rakow, 2009; Thorsteinson & Withrow, 2009; Payne, Samper, Bettman & Luce, 2008).

1.2 Overview of experiments

In three experiments, decisions made after a period of distraction were compared to those made after a period of conscious deliberation of the same duration.

In all experiments, the decision task was similar to the classical UTT decision task. Before information acquisition, participants were instructed to form an impression of the different alternatives. Then information about three or four alternatives, one of which was characterized by more positive features than the others (typically 9, 6, 6 and 3 positive features), was presented to participants. After reading the information, participants judged the alternatives either after a distraction task, hypothesized to elicit "unconscious thought", or after a fixed period of conscious deliberation (typically four minutes during which they could not consult the information).

More specifically, in the first study we offer a conceptual replication of UTT's findings with a different material. With this first study, we also tested whether distraction reduces stereotyping. In the second experiment, to ensure maximal comparability between the studies, we used the same material and procedure as Dijksterhuis et al. (2006). Experiment 3, follows the same design, but we used more stringent controls on the materials and the methods.

2 Experiment 1

On the basis of previous studies showing that stereotyping increases when cognitive capacity is constrained (e.g., Bodenhausen, 1988; Macrae, Milne & Bodenhausen, 1994), UTT theorists have argued that "unconscious thought" should reduce stereotyping (Dijksterhuis & Nordgren, 2006). Indeed, according to Dijksterhuis and colleagues, as constraining cognitive capacity during encoding enhances stereotyping, relying on a mode of thought (i.e., conscious thought) that has limited capacity during decision-making should enhance stereotyping. Since UTT assumes that "unconscious thought" has very large processing capacity, the theory predicts that, by contrast, "thinking unconsciously" should reduce stereotyping. For example, after having received information about job candidates, a period of distraction would help to reduce discrimination. Thus the same type of manipulation (i.e., performing a concurrent task) would have a different effect during information acquisition and decision-making. Some experiments supporting these predictions

are mentioned in Dijksterhuis & Nordgren (2006), however the details of that work are not reported.

It should be noted that submitting participants to a concurrent task is classically viewed as a way to constrain cognitive capacity (e.g., Fiske & Neuberg, 1990; Gilbert & Hixon, 1991) rather than promote *unconscious thought*. In this respect, performing a distraction task while making a decision could simply be viewed as a way to limit cognitive resources and therefore enhance stereotyping rather than a way allow the unconscious to deploy its processing capacity. Congruently with this view, previous literature states that people are more likely to rely on heuristics when their cognitive resources are limited (Gigerenzer & Selten, 2001) and that cognitive busyness facilitates stereotype application (Gilbert & Hixon, 1991). However, cognitive resources are also required to maintain stereotypes in the face of disconfirmation via the subtyping process (Yzerbyt, Coull & Rocher, 1999).

Given that experiments relative to UTT's predictions concerning stereotyping were not reported in the literature, Experiment 1 offers a first thorough report of data relative to these predictions.

2.1 Method

2.1.1 Participants and design

There were 109 participants (54 women and 55 men), mostly students following various curricula at the Université Libre de Bruxelles, ranging in age from 18 to 44 (mean: 20.95). They were randomly assigned to one of four conditions resulting from crossing two factors: Decision Mode (deliberation vs. distraction) and Stereotype-consistency of the Target (stereotype-consistent vs. stereotype-inconsistent). They received course credits or €2 in return for their participation. Data from eight participants were dropped because they were incomplete.

2.1.2 Procedure and materials

Participants were each seated in front of a computer in separate cubicles. They were told that they had to take the role of a psychologist in charge of staff recruitment and that they would have to assess four candidates applying for an engineer position. After a description of the company and of the job, they received information about four candidates, each of them characterized with 12 attributes. Each attribute was presented for 3500 ms in the centre of the screen. There was a pause of 500 ms between two subsequent attributes.

The attributes used to describe the candidates pertained to 12 criteria (e.g., experience in management) and were either positive or negative. (e.g., David has already managed a team during his career vs. Sophie has not yet

managed a team in her career). One candidate was described with eight positive attributes ("the best candidate"); two candidates were described with six positive attributes ("the average candidates"); one candidate was described with only four positive attributes ("the worst candidate"). Given that the competence of each candidate was determined by the number of positive attributes that characterized him or her as in previous research (e.g., Dijksterhuis, 2004; Dijksterhuis et al., 2006), the criteria were pretested to ensure that they were perceived as similar in terms of importance. Criteria were rated on a 5-points scale by 21 university students that did not take part to the actual experiment. We then selected those criteria that had elicited similar rankings. They had ranged from 2.6 to 3.5 in importance. Before receiving information, participants were told that all the criteria were equally important and that the desired candidate should be as good as possible on each of them. Information was organized by candidate: participants first received information about the first "average candidate", then about the "worst candidate", then about the "best candidate" and finally about the second "average candidate". The order in which attributes pertaining to each candidate were presented was randomized for each participant.

As the profession of engineer is stereotypically more associated with men than with women (Cjeka & Eagly, 1999), we manipulated the first name of the candidates. In the stereotype-consistent condition, the "best candidate" was described with a male first name, and the "worst candidate" with a female first name. By contrast, in the stereotype-inconsistent condition, the "best candidate" was described with a female first name, and the "worst candidate" with a male first name. Independently of the previous manipulation, two additional controls were implemented: For half of the participants, the first "average candidate" had a male first name and the second a female first name, whereas, for the other half, the first "average candidate" had a female first name and the second a male first name. Moreover two sets of first names were used to dub candidates.

After information about each candidate had been presented, participants in the deliberation condition were instructed to think about the four candidates for three minutes. In the distraction condition, they were informed that they would have to assess the four candidates after they had performed another task. In this task participants had to memorize numbers and perform simple mathematical operations on them. Five numbers between one and nine were sequentially presented for two seconds each. Each number was associated with a letter: the first was associated with A, the second with B, the third with C, the fourth with D and the last with E. Following the presentation of the numbers, participants were asked to add two of them (e.g., C + E). They had a maximum of ten seconds

Table 1: Differentiation index and rank order correlation as a function of decision mode and stereotype-consistency of the target.

Condition	Differentiation index		Rank order corr.	
	Mean	SE	Mean	SE
Stereotype consistent				
Deliberation	1.81	.34	.66	.10
Distraction	1.67	.32	.44	.10
Stereotype inconsistent				
Deliberation	1.89	.36	.61	.11
Distraction	1.25	.38	.42	.11

to provide their answer. Afterwards, they assessed the four candidates on a 9-points scale by judging “to what extent do the different candidates fit the proposed job?” At the end of the experiment they provided demographic information and were thanked and debriefed.

2.2 Results

As in previous research (e.g., Lassiter, Lindberg, González-Vallejo, Bellezza & Phillips, 2009), we used the difference between the attitude toward the best candidate and the mean attitude toward the others as an indicator of decision quality (differentiation index). Larger values of this index reflect a stronger preference for the best candidate. Decision quality was then examined as a function of Decision Mode and Stereotype-consistency of the Target.

The critical comparison between Decision Modes failed to reach significance, $F(1,97) = 1.23, p > .2, \eta^2 = .013$. As can be seen in Table 1, the data were numerically opposite to what had been predicted: participants performing slightly better in the deliberation condition than in the distraction condition. Stereotype-consistency of the Target did not have a significant impact on decision quality, $F(1,97) = .23, p > .6, \eta^2 = .002$. The predicted interaction between Decision Mode and Stereotype-consistency of the Target was not significant either, $F(1,97) = .50, p > .4, \eta = .005$. Again, the data were numerically opposite to what had been predicted: in the deliberation condition, decision quality was roughly equal whether the “best candidate” was stereotype-consistent or stereotype-inconsistent whereas in the distraction condition, decision quality was poorer when the “best candidate” was stereotype-inconsistent than when he or she was stereotype consistent.

The difference between the attitude toward the best

candidate and the mean attitude toward the others assesses only whether the best alternative is differentiated from the others. We thus computed a second index that permitted us to examine whether participants correctly rank the four alternatives. To compute this index, we recoded evaluations in a ranking for each participant and computed the rank order correlation between this ranking and the normative ranking (based on the percentage of positive attributes characterizing each alternative).¹

Rank order correlations were examined as a function of Decision Mode and Stereotype-consistency of the Target. This analysis revealed a main effect of Decision Mode, $F(1,94) = 3.89, p = .051, \eta^2 = .40$. The ranking was more appropriate in the deliberation condition than in the distraction condition. Neither the effect of Stereotype-consistency of the Target, $F(1,94) = .11, p > .7, \eta^2 = .001$, nor the interaction with Decision Mode was significant, $F(1,94) = .04, p > .8, \eta^2 = .000$.

2.3 Discussion

Our results did not support the idea that an unconscious process, more efficient than conscious deliberation, occurs during the distraction period. In fact the data suggest that conscious deliberation leads to superior performance in ranking alternatives correctly. The analysis of rankings of alternatives provides more statistical power than analysis of whether or not the best alternative is preferred to the others.

In view of our failure to demonstrate a conceptual replication of earlier findings, two types of explanations come to mind: It may be the case that procedural differences between the present experiment and experiments supporting UTT are sufficient to account for divergent results, or it may be the case that the different materials used induce different effects. Indeed, the material we used in Experiment 1 was partially numeric. This fact could explain why we obtained discrepant results, because UTT states that unconscious thought is unable to perform mathematical operations. However, as in previous research (e.g., Dijksterhuis et al., 2006), all attributes were dichotomous (e.g., “has already managed a team during his career” vs. “has not yet managed a team in his career”). In our view, it is equivalent to say that the test result (e.g., English level) was 4/5 versus 2/5 or to say that the result was good versus bad. Nevertheless, we decided to conduct a second experiment that replicated the original study (Dijksterhuis et al., 2006, Study 2) more exactly.

Regarding Stereotype-consistency of the Targets, our results do not support UTT predictions: Participants did not perform better after distraction than after deliberation

¹ A correlation could not be computed for participants who assessed the four candidates as equivalent.

in the stereotype-inconsistent condition. However, in our sample, participants did not apply the stereotype in any experimental condition, preventing firm conclusions on this aspect.

3 Experiment 2

In this study, as in the original study, the complexity of the task, operationalized as the number of attributes characterizing alternatives, was manipulated and participants had to evaluate cars. According to UTT, distracted participants should perform better when dealing with complex choices whereas those who have time to consciously consider their decision should perform better when dealing with simple choices. These predictions derive from two characteristics of modes of thought proposed by UTT: *unconscious thought* has much larger capacity than *conscious thought* whereas *conscious thought* can follow rules and is more precise than *unconscious thought* (e.g., Dijksterhuis et al., 2006). The present experiment's procedure was as close as possible to the procedure used in the original study (Dijksterhuis et al., 2006). To achieve consistency, the online supporting material provided by Dijksterhuis et al. (2006) was used after translation into French.

In order to investigate the capacity principle, after participants had evaluated alternatives, we asked how many attributes they had taken into account to perform their evaluations. The capacity principle states that more attributes could be taken into account when processing information unconsciously rather than consciously. The ability of unconscious thought to take many attributes into account is said to be related to the processing capacity of the unconscious. To estimate this capacity, Dijksterhuis and colleagues endorse the following reasoning: as 40 to 60 bits per second can approximately be processed consciously whereas the entire human system can process about 11,200,000 bits, the amount of information that can be processed unconsciously should be enormous.

The capacity principle has been only indirectly tested in a study by asking participants whether they based their choice on a global impression or on a few specific attributes: Distracted participants more often reported that they based their choice on a global impression (Dijksterhuis, 2004, Experiment 2). In that experiment no mediation of the effect of Decision Mode by the number of attributes taken into account was evidenced and this number was inferred from an indirect question. The capacity principle would be one of the advantages of unconscious thought (Dijksterhuis & Nordgren, 2006). As it has not been thoroughly tested, we offer additional data pertaining to this principle in Experiments 2 and 3.

3.1 Method

3.1.1 Participants and design

Participants were 65 people (55 women and 10 men), mostly students following various curricula at the Université Libre de Bruxelles, ranging in age from 17 to 29 (mean, 19.49). They were randomly assigned to one of four conditions resulting from crossing two factors: Decision Mode (deliberation vs. distraction) and Complexity (4 aspects vs. 12 aspects). They received €5 in return for their participation.

3.1.2 Procedure and materials

Participants were seated in front of computers in separate cubicles. The study was described as an experiment on decision making. Participants were told that they would receive information about four hypothetical cars and were instructed to carefully read this information in order to form an impression of these cars. Information about the 4 cars was then presented. Each attribute was presented for 8000 ms in the centre of the screen. There was a pause of 800 ms between two successive attributes. Depending on the condition, each car was characterized by 4 attributes or by 12 attributes. Each hypothetical car was associated with an imaginary name: Hatsdun, Kaiwa, Dasuka or Nabusi. The attributes used to describe the cars were relative to 12 criteria (e.g., handling) and were either positive or negative. (e.g., The Hatsdun has good handling vs. The Kaiwa has poor handling). One car was described with 75% positive attributes ("the best car"); two were described with 50% positive attributes ("the average cars") and one was described with only 25% positive attributes ("the worst car"). We also ensured that the names of the cars were counterbalanced: Each type of car (as defined by a specific proportion of positive attributes) was associated with each of the four possible names across participants. In the simple condition, the set of four criteria used to describe the cars was randomly selected and the prescribed percentage of positive attributes for each car was respected. Order of presentation was randomized with the following constraints:

- Attributes of the same car (e.g., the Hatsdun) were never presented successively;
- Attributes relative to the same criteria (e.g., handling) were never presented successively.

After information had been presented, participants in the distraction condition had to solve anagrams for four minutes. By contrast, participants in the deliberation condition were instructed to think about the cars for four minutes. All participants were then asked to evaluate each

Table 2: Differentiation index and rank order correlation as a function of decision mode and complexity.

Condition	Differentiation index		Rank order corr.	
	Mean	SE	Mean	SE
4 aspects				
Deliberation	1.55	0.45	0.36	0.10
Distraction	1.25	0.46	0.36	0.10
12 aspects				
Deliberation	1.88	0.46	0.54	0.11
Distraction	1.71	0.46	0.48	0.11

car on a 9-points scale. After this evaluation phase, participants were asked about how many attributes they had taken into account to perform their evaluations. Then they were asked to rate the importance of the different criteria used to describe the cars. Finally, they provided personal information and were thanked and debriefed.

3.2 Results

3.2.1 Decision quality

Decision quality was computed by subtracting from the evaluation of the “best car” the mean of the evaluations of the other cars (differentiation index). It was then examined as a function of Decision Mode and Complexity. The critical interaction between these factors was absent, $F(1,61) = .02$, $p > .8$, $\eta^2 = .000$. As can be seen in Table 2, in the simple condition, participants who had time to consciously consider their decision performed slightly better than distracted participants and the same trend was observed in the complex condition.

As another indicator of decision quality, we also examined the Rank order correlation between participants' rankings and normative rankings as a function of Decision Mode and Complexity (see Table 2). As in the previous analysis, neither Decision Mode, $F(1,59) = .13$, $p > .7$, $\eta^2 = .002$, nor Complexity, $F(1,59) = 2.08$, $p > .1$, $\eta^2 = .034$, nor the interaction between these factors had a significant effect, $F(1,59) = .08$, $p > .7$, $\eta^2 = .001$.

3.2.2 Capacity principle

The number of attributes that participants reported having taken into account did not differ as a function of Decision Mode, $F_s(1,61) = .47$, $p > .5$, $\eta^2 = .008$. However, as expected, this number was significantly higher in the complex condition than in the simple condition, $F = 18.16$, $p < .001$, $\eta^2 = .229$. The correlation between the number of

Table 3: Criteria importance (on a 9-points scale) and 95% confidence intervals.

Criteria	Mean	95% CI	
		Lower	Upper
Handling	7.68	7.30	8.07
Easiness of gears shifting	7.05	6.63	7.47
Environment respect	7.02	6.46	7.58
Quality of the service	6.61	6.12	7.11
Available legroom	6.53	6.08	6.98
Trunk size	6.42	6.07	6.78
Mileage	6.33	5.88	6.78
Sound system quality	5.25	4.68	5.81
Recency of the model	5.05	4.52	5.58
Number of available colors	4.26	3.71	4.82
To have a sunroof	2.96	2.52	3.41
Presence of cup-holders	1.93	1.62	2.24

attributes taken into account and decision quality (differentiation index) was also computed, in both simple and complex conditions. Surprisingly these two correlations were not significant ($r_s = .241$ and $.069$, $p_s > .17$).²

3.2.3 Criteria importance

The criteria used to describe the four cars were seen as very different in importance; criteria importance ranged from 7.68 for the handling to 1.93 for the presence of cup-holders on a 9-points scale (see Table 3).

3.3 Discussion

In spite of the fact that we used an identical design and the same material as Dijksterhuis et al. (2006), the results of Experiment 2 failed to replicate the results of the original study. Indeed, the expected interaction between Decision Mode and Complexity was not observed: considered decisions had roughly the same accuracy as decisions made after a period of distraction for both levels of complexity. As in Experiment 1, results do not support the idea that an unconscious process, more efficient than conscious consideration when dealing with complex decisions, occurs during the distraction period.

Compared to the only set of data relative to processing capacity (Dijksterhuis, 2004, Experiment 2), Experiment 2 offers a more direct test of the number of attributes that both Decision Modes allow taking into account. In

²Correlations were also computed between the number of attributes taken into account and the Rank order correlation. These correlations were not significant either.

Dijksterhuis (2004), participants were asked if their choice was based on a global judgment or on only one or two specific attributes and this measure was taken as an indicator of the number of attributes that participants were able to take into account. Moreover, processing capacity of each Decision Mode was inferred on the basis of this measure. In our view, reporting that the choice was based on a global impression does not necessarily imply that many attributes have been taken into account. Moreover if participants are able to correctly estimate how many attributes they took into account, using a dichotomic question (global vs. specific) necessarily considerably decreases the statistical power.

In Experiment 2, we directly asked participants how many attributes they took into account to make their decision. When using this more direct phrasing, no difference between Decision Modes was found. This finding does not support the idea that performing a distraction task increases the number attributes that are taken into account when making a decision. However, it should be noted that the self-report measures used in both the previous and the present experiments may not be valid estimators of the number of attributes that have been taken into account, especially given the fact that processing is supposed to be unconscious in the distraction condition.

No correlation was found between the number of attributes that participants reported to have taken into account and decision quality. This lack of correlation may indicate that the used measure is not reliable. Another possibility is that the lack of correlation is related to the weightings that participants gave to the attributes used to describe the cars. As these attributes varied enormously in perceived importance, taking more attributes into account did not necessarily lead to a better decision. For example, if the handling of the car strongly influenced its evaluation, and if the presence of cup-holders had a negligible influence on this evaluation, basing the evaluation on two rather than on one attribute would also have a negligible influence on evaluations and decision quality. This finding raises a methodological concern since, in UTT's standard decision task (e.g., Dijksterhuis et al., 2006), the quality of the alternatives is defined by simply counting the number of positive features of each of them. One way to solve this problem would be to describe the cars with attributes that are more similar to one another in terms of importance (Experiment 3). Another method would be to take into account the relative importance of each attribute (e.g., Newell et al., 2009) as in a weighted-additive model (Dawes & Corrigan, 1974; Gigerenzer & Goldstein, 1996; Payne, Bettman & Johnson, 1993). Still, given that one of the rationales for the advantage of "unconscious thought" in multi-factorial decisions, is its ability to take many attributes into account (e.g., Dijksterhuis & Nordgren, 2006), this lack of corre-

lation is problematic from a theoretical point of view.

4 Experiment 3

Given that the material used in previous experiments (e.g., Experiment 2; Acker, 2008; Dijksterhuis, Bos et al., 2006; Rey, Goldstein & Perruchet, 2009) raised methodological problems, we decided to run a pre-test to select a material in which the importance of attributes was more similar than in the original material, presuming that this would enhance the correlation between the number of attributes taken into account and decision quality. Using attributes of various importance would be valuable to the extent that a weighted additive model is used to define alternatives quality. However, because we were investigating the capacity principle, we designed a decision task involving attributes of similar importance in order to maximize the number of attributes that needed to be taken into account to make a correct decision. Indeed, when using attributes of various importance, a correct decision can be made when only considering a few important attributes. So in this case, the "take the best" heuristics (e.g., Gigerenzer & Goldstein, 1999) that implies to consider important attributes first and to base the choice on the first discriminant attribute encountered could perform well. UTT does not assume that the nature of the unconscious treatment is heuristic but rather that unconscious thought should perform better because it is capable to base the decision on a large number of attributes. To investigate the capacity principle, it is thus necessary to use a decision task that requires taking many attributes into account to be correctly performed.

We decided to focus on complex decisions, because the prediction made by UTT is more crucial for such decisions. In addition, when inspecting the original paper, it was not very clear whether participants had to choose the best car or their favorite car:

*"In the **conscious thought** condition, participants were asked to think about the cars for 4 min before they **chose their favorite car** ... In the **unconscious thought** condition, participants were distracted for 4 min (they solved anagrams) and were told that after the period of distraction they would be asked to **choose the best car**."* (Dijksterhuis et al., 2006, p. 1006).

Hence, we decided to manipulate this factor in this third experiment. Indeed we wished to investigate whether the use of different instructions could explain the observed difference between Decision Modes. Indeed, because, decision quality was defined normatively, choosing the best car (this is normative) could lead to better performance than choosing one's favorite car (which is subjective).³

³This distinction may be congruent with the distinction between

4.1 Method

4.1.1 Participants and design

Participants were 100 undergraduate psychology students (83 women and 17 men) at the Université Libre de Bruxelles, ranging in age from 18 to 38 (mean: 20.03). They were randomly assigned to one of four conditions resulting from the crossing of two factors: Instruction Type (favorite vs. best) and Decision Mode (deliberation vs. distraction). They received course credits for their participation.

4.1.2 Procedure and materials

The procedure was very similar to that of previous experiments (e.g., Acker, 2008; Dijksterhuis et al., 2006; Newell et al., 2009; Rey et al., 2009; Experiment 2) except for a few changes. Because attributes used in previous studies were seen as very different in importance by participants, we ran a pretest to select a new set of attributes in which the importance of each item was more similar than in the original material. In the pretest, 43 participants had to rate the influence of several positive and negative attributes on the choice of a car. Selected attributes ranged from 5.6 to 7.33 for the positive items and from -6.95 to -5.17 for the negative items on a 21 points (ranging from -10 to 10) scale whereas the original material ranged from 1.93 to 7.68 on a nine points scale.

We also introduced another modification in the information presentation method. In all previous studies, a profile was associated with each car. This means that there were four sets of 12 attributes and that each set was always associated with the same car. When using this method, it is fundamental that each of the 12 attributes has exactly the same importance because the quality of the cars is defined by the number of positive attributes. To avoid this concern we used an additional control in the present study: we randomized the profile associated with each car. When a set of four profiles was created, the valence of the attributes associated with each car was randomized while respecting the prescribed percentage of positive information for each car (75% for the “best car” 50% for the two “average cars” and 25% for the “worst car”). In previous experiments, the three positive attributes associated with the “worst car” were always about the legroom, the sunroof and the number of available colors; with the new method the positive attributes associated with this car could be about the service, the power of the engine and the alarm system, or the acceleration, the air conditioning and the sound system, or any other combination of attributes. Thanks to this method, even if all items do not exactly have the same importance,

“want” and “should” judgments developed by Bazerman, Tenbrunsel, & Wade-Benzoni (1998).

the hierarchy between the different cars is inevitably respected and the results cannot be explained by a specific pattern of information as in previous experiments.

We also made another methodological improvement: we used the same set of profiles, the same order of presentation for all attributes, all other items, and questions for one participant in each experimental condition. So each participant was paired with three other participants in the three other conditions.

Contrary to Experiment 2, we kept the complexity of the decision constant: each car was described with 12 attributes. In this study, we manipulated the Type of Decision: half of the participants had to objectively evaluate the cars whereas the other half had to rely on their personal preferences. Finally the evaluation was made on a continuous scale (recoded on 100-points) rather than a 9-points scale.

4.2 Results

4.2.1 Decision quality

Decision quality was computed by subtracting from the evaluation of the “best car” the mean of the evaluations of the other cars (differentiation index). It was then examined as a function of Decision Mode and Instruction Type. The analysis of variance revealed no effect of Decision Mode, $F(1,96) = .06, p > .8, \eta^2 = .001$, no effect of Instruction Type, $F(1,96) = .27, p > .6, \eta^2 = .003$, and no interaction between these factors, $F(1,96) = .24, p > .6, \eta^2 = .002$. As shown in Table 4, decision quality was roughly equal in all experimental groups.

We also examined Rank order correlations as a function of Decision Mode and Instruction Type (see Table 4). As in the previous analysis, neither Decision Mode, $F(1,96) = .25, p > .6, \eta^2 = .003$, nor Instruction Type, $F(1,96) = .13, p > .7, \eta^2 = .001$, nor the interaction between these factors had a significant effect, $F(1,96) = .85, p > .3, \eta^2 = .009$.

4.2.2 Capacity principle

The number of attributes that participants reported having taken into account failed to differ as a function of Decision Mode, $F_s(1,98) = .18, p > .6, \eta^2_s = .002$. Contrary to the previous experiment, the number of attributes that participants reported to have taken into account was significantly correlated with decision quality (differentiation index) ($r = .32, p < .001$).⁴

⁴A correlation of .29, $p < .005$ was obtained when considering the rank order correlation index.

Table 4: Differentiation index and Rank order Correlation as a function of Decision Mode and Instruction Type

Condition	Differentiation index		Rank order corr.	
	Mean	SE	Mean	SE
Favorite				
Deliberation	20.88	5.55	0.44	0.11
Distraction	19.53	5.55	0.28	0.11
Best				
Deliberation	21.04	5.55	0.38	0.11
Distraction	25.12	5.55	0.43	0.11

4.2.3 Criteria importance

The importance that participants gave to the criteria used to describe cars ranged from 5.25 to 6.7 on a 9-points scale.

4.3 Discussion

Even when using more controlled material and several random combinations of information, distraction did not improve complex decisions. Decision quality did not differ either as a function of the Instruction Type. Descriptively, decision quality was slightly better in the best condition than in the favorite condition, which is not surprising given that decision quality was computed normatively.

As in Experiment 2, the results do not support the idea that more information can be taken into account after distraction than after deliberation. Contrary to Experiment 2, a significant correlation was found between the number of attributes that participants reported to have taken into account and decision quality. This suggests that a more controlled selection of material helps to enhance the link between the number of attributes taken into account and decision quality and that the ability to take more attributes into account improves the quality of decisions based on numerous aspects. It also supports the reliability of the self report measure used to estimate the number of attributes that have been considered. Indeed we failed to obtain a significant correlation in a choice task that can be successfully completed by considering only a few important attributes (e.g., Experiment 2; Dijksterhuis et al., 2006) whereas we obtained a significant correlation in a choice task that cannot be successfully completed without considering many attributes (because attributes were of similar importance).

5 General discussion

The three studies presented here did not support UTT's claims regarding the merits of "unconscious thought" in complex decision making. Indeed in Experiment 1, we observed a marginally significant effect in the opposite direction (when considering rank order correlation): decisions were better after deliberation than after distraction. A similar pattern of means was obtained when examining the differentiation index. However decision modes did not differ significantly. Experiments 2 and 3 also yielded null results but the pattern of means was again incongruent with UTT.

Whereas these findings appear puzzling when compared to studies that support UTT (e.g., Dijksterhuis, 2004; Dijksterhuis et al., 2006), they lead to similar conclusions as other attempts of replication (for a meta-analysis see Acker, 2008).

Importantly, the present line of experiments identifies concerns regarding the UTT paradigm. First we showed that, although participants perceived attributes characterizing the cars as very different in terms of importance, studies supporting UTT have neglected the weighting of the attributes in determining their quality. Moreover, when using the original material, we found no correlation between the number of attributes that participants reported to have taken into account and the quality of their decisions, which appears problematic since the superiority of "unconscious thought" should be dependent on its processing capacity, according to the theory. By contrast when using a more controlled material we indeed found a correlation between the number of attributes taken into account and decision quality. Second, we noted that the experimental instructions used by Dijksterhuis et al. (2006) were unclear about whether participants had to choose the best car or their favorite one. Third, we have suggested that, to avoid artifacts, it would be more rigorous to use several sets of information to describe the alternatives rather than always describing the same car with the same advantages and drawbacks (see Experiment 3 for more details). Finally Dijksterhuis and colleagues (2006) did not mention whether they counterbalanced the name associated with each car in their experiments. Obviously, the name of the cars should be counterbalanced.

In sum, across three studies, we found no support for the superiority of "unconscious thought" in complex decision-making and identified methodological problems that should be taken into account in further investigation of UTT.

References

Acker, F. (2008). New findings on unconscious versus conscious thought in decision making: additional em-

- irical data and meta-analysis. *Judgment and Decision Making*, 3, 292–303.
- Bazerman, M. H., Tenbrunsel, A. E. & Wade-Benzoni, W. B. (1998). Negotiating with yourself and loosing: Making decision with competing internal preferences. *Academy of Management Review*, 23, 225–241.
- Bodenhausen, G. V. (1988). Stereotypic biases in social decision making and memory: Testing process models of stereotype use. *Journal of Personality and Social Psychology*, 55, 726–737.
- Bos, M. W., Dijksterhuis, A., & van Baaren, R. B. (2008). On the goal-dependency of unconscious thought. *Journal of Experimental Psychology*, 44, 1114–1120
- Cejka, M. A., & Eagly, A. H. (1999). Gender-Stereotypic Images of Occupations Correspond to the Sex Segregation of Employment. *Personality and Social Psychology Bulletin*, 25, 413–423.
- Dawes, R. M., & Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin*, 81, 95–106.
- Dijksterhuis, A. (2004). Think different: The merits of unconscious thought in preference development and decision making. *Journal of Personality and Social Psychology*, 87, 586–598.
- Dijksterhuis, A., Bos, M. W., Nordgren, L. F., & van Baaren, R. B. (2006). On making the right choice: The deliberation-without attention effect. *Science*, 311, 1005–1007.
- Dijksterhuis, A., & Nordgren, L. F. (2006). A theory of unconscious thought. *Perspectives on Psychological Science*, 1, 95–109.
- Evans, J. St. B. T. (2008). Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition, *Annual Review of Psychology*, 59, 255–278
- Fiske, S. T., & Neuberg, S. L. (1990). A continuum model of impression formation from category-based to individuating processes: Influences of information and motivation on attention and interpretation. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 23, pp. 1–74). San Diego, CA: Academic Press.
- Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, 103, 650–669.
- Gigerenzer, G., & Goldstein, D. G. (1999). Betting on one good reason: The take-the-best heuristics. In G. Gigerenzer, P. M. Todd, & the ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 75–96). New York: Oxford University Press.
- Gigerenzer, G., & Selten, R. (2001). *Bounded rationality: The adaptive toolbox*. Cambridge/MA: MIT Press.
- Gilbert, D. T., & Hixon, J. G., (1991). The trouble of thinking: Activation and application of stereotypical beliefs. *Journal of Personality and Social Psychology*, 60, 509–517.
- Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases* (pp. 49–81). New York: Cambridge University Press.
- Lassiter, G. D., Lindberg, M. J., González-Vallejo, C., Bellezza, F. S. & Phillips N. D. (2009) The Deliberation-Without-Attention Effect: Evidence for an Artifactual Interpretation, *Psychological Science*, 20, 671–675.
- Lerouge, D. (2009). Evaluating the benefits of distraction on product evaluations: The mindset effect. *Journal of Consumer Research*, 36, 367–379.
- Macrae, C. N., Milne, A. B., Bodenhausen G. V. (1994). Stereotypes as energy-saving devices: A peek inside the cognitive toolbox. *Journal of Personality and Social Psychology*, 66, 37–47.
- Newell, B. R., Wong, K. Y., Cheung, J. C., & Rakow, T. (2009) Think, Blink or Sleep on it? The impact of Modes of Thought on Complex Decision Making, *The Quarterly Journal of Experimental Psychology*, 62, 707–732.
- Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. Cambridge: Cambridge University Press.
- Payne, J. W., Samper, A., Bettman, J. R., & Luce, M. F. (2008). Boundary condition on unconscious thought in complex decision making. *Psychological Science*, 19, 1118–1223
- Rey, A., Goldstein, R. M., & Perruchet, P. (2009). Does unconscious thought improve complex decision making?, *Psychological Research*, 73, 372–379.
- Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, 69, 99–118.
- Strick, M., Dijksterhuis, A., Bos, M. W., Sjoerdma, A., van Baaren, R. B., Nordgren, L. F. (2009). A meta-analysis on unconscious thought effects. Manuscript in preparation
- Thorsteinson, T. J. & Withrow, S. (2009). Does unconscious thought outperform conscious thought on complex decisions? A further examination. *Judgment and Decision Making*, 4, 235–247.
- Waroquier, L., Marchiori, D., Klein, O., Cleeremans, A. (in press) Is it better to think unconsciously or to trust your first impression? A reassessment of Unconscious Thought Theory. *Social Psychological and Personality Science*.
- Yzerbyt, V., Coull, A. & Rocher, J. (1999). Fencing Off the Deviant : The Role of Cognitive Resources in the Maintenance of Stereotypes. *Journal of Personality and Social Psychology*, 77, 449–462.