Judgment and Decision Making, Vol. 3, No. 4, April 2008, pp. 292–303

New findings on unconscious versus conscious thought in decision making: additional empirical data and meta-analysis.

Felix Acker*
School of Psychological Science
La Trobe University

Abstract

Ninety-eight Australian students participated in a functional replication of a study published by Dijksterhuis et al. (2006). The results indicated that unconscious thought does not necessarily lead to better normative decision making performance than conscious thought, which is contrary to the results found in Dijksterhuis et al. Since other studies showed a positive, though statistically not significant, effect for unconscious thought, a meta-analysis comprising a total of 17 experiments was conducted. It suggests that there is little evidence for an advantage to normative decision making using unconscious thought. However, a discussion of potential moderators shows that further study would help to identify situations in which unconscious thought is truly helpful and those in which it is not.

Keywords: unconscious thought, meta-analysis, normative decision making.

1 Introduction

In a series of studies, Dijksterhuis and colleagues (e.g., Dijksterhuis, 2004b; Dijksterhuis, Bos, Nordgren, & van Baaren, 2006; Dijksterhuis & Meurs, 2006) established the surprising and counterintuitive finding that unconscious thought leads to better decision making performance for complex problems than conscious thought. Unconscious thought is defined as "cognitive and/ or affective task-relevant processes that take place outside conscious awareness" (Dijksterhuis, 2004b, p. 586). A second postulate is that the reverse is true for simple decision: consciously thinking about them is better. This paper is concerned only with the first.

A common denominator of all experiments described by Dijksterhuis and colleagues is the experimental technique. Participants are split in two or three groups and provided with a number of pieces of information about a number of options to choose from, for instance four apartments or flatmates. Each option is described by the same number of attributes, and usually these are attributes of the choice option. For example, an attribute of all choice options for "apartments" might be the size. Typically these attributes are conceptualized dichotomously — either an apartment is spacious or it is not. All pieces of information about all choice options are presented either as lists (e.g., Newell, Wong, Cheung, & Rakow, in press), individually and randomized (e.g., this study), or individually in a fixed order (e.g., Phillips et al., 2007) for a fixed amount of time. Prior to the information presentation, participants are informed that they will have to a) choose one of the options (e.g., Dijksterhuis, 2004b, Experiment 1) or b) rate each option (e.g., Lerouge, submitted). Then, following the presentation of the information, participants either immediately make a choice between options or think about their choice for a fixed amount of time (except in Payne, Samper, Bettman, & Luce, 2007), or are distracted for the same amount of time before making their decision.

The principal claim is that, when the number of factors that ought to be considered is high (as indicated by the number of attributes), and the decision is therefore complex, unconscious thought will lead to better decision performance than conscious deliberation. Better decision making (which I call "normative" here) is defined as choosing, or ranking higher, options with more positive features. However, although most of Dijksterhuis and colleagues' experiments found an effect in the hypothesized direction, such effects often failed to reach statistical significance when comparing the unconscious and conscious group directly. For example, in the first reported experiment in Dijksterhuis' 2004 study the critical comparison did not reach statistical significance ($p <$ .08).

In this context it is important to point out that in the sequence of experiments carried out by Dijksterhuis, only the first two to be published (Dijksterhuis, 2004, Experi-

*I am indebted to Mary Omodei and Geoff Cumming (both La Trobe University), Ben Newell (UNSW) as well as one other anonymous reviewer for helpful comments and critique. Address: Felix Acker, School of Psychological Science, La Trobe University, 1 Kingsbury Drive, Melbourne, 3078. Email: f.acker@latrobe.edu.au

ments 1 and 2) were exclusively directed at establishing differences between decision success under unconscious versus conscious thought. The other experiments primarily investigated further aspects of what was to become the *theory of unconscious thought* (Dijksterhuis & Nordgren, 2006). For example, Experiments 1 and 2 of Dijksterhuis, Bos, Nordgren, and van Baaren's 2006 study dealt with the *deliberation without attention effect*, which required two conditions: a complex and a simple one, which were differentiated by the number of information bits participants received. However, even though these studies did not primarily contrast conscious with unconscious thought, they frequently provided data supporting the notion that unconscious thought is a superior form of integrating a large amount of information.

Conscious thought, according to Dijksterhuis, is very useful for monitoring information with respect to a particular criterion, for example the minimal amount of space that must be available in an apartment. It can engage in logical operations and work with quantities in a precise fashion, as is the case in mathematical operations. On the other hand, it has limited capacity and is thus unsuitable to integrate large amounts of information. Unconscious thought, on the other hand, is claimed to possess virtually infinite processing capacity, to process information divergently and to employ a natural weighing mechanism for acquired information. The alleged differences between the two forms of thought have been described by six principles and some additional characteristics in Dijksterhuis and Nordgren (2006), who re-stated the basic tenet that unconscious thought works well for complex decision situations with multiple sources of relevant information. Despite its empirical successes and its advanced conceptual development, the theory of unconscious thought has not been without criticism.

Shanks (2006), for example claimed that the unconscious thought effects were an artefact of differential rates of forgetting rather than the product of different forms of processing. The theory was also criticised for limited sample size in supporting studies, and concern was voiced about the applicability of the theory to the clinical context (Bekker, 2006) for which, as Dijksterhuis and colleagues pointed out, the theory was not intended. In addition to these criticisms, the currently published data are fairly limited in that they almost exclusively include studies from one laboratory.

The purpose of the present article is two-fold. The first aim is to provide replication data for the unconscious decision making effect with an English-speaking sample and English stimulus material. The study was also designed to gather additional information to rule out potential alternative explanations for the unconscious thought effect. Secondly, this article aims to provide a statistically more robust estimate of the population effect size for the unconscious thought effect by meta-analytically synthesizing data from a subset of studies on unconscious thought.

Table 1: Valence for the 12 attributes for each choice option. A value of "1" implies positive, "0" implies negative valence.

Attributes	Hatsdun	Kawai	Dasuka	Nabusi
Environment	1	1	0	0
Cupholders	1	0	1	1
Many colours	1	1	0	1
Sound system	0	0	1	0
Service	1	1	0	0
Handling	1	0	1	0
Milage	1	1	0	0
Leg Space	0	1	0	1
Trunk size	1	1	0	0
Sunroof	1	0	1	1
Gear shifting	0	1	1	0
Age	1	0	1	0
Sum	9	7	6	4

2 Empirical Study

2.1 Method

2.1.1 Participants

All participants were third year psychology students that took part in this experiment in the context of a tutorial series on decision making. Participation in the experiment was voluntary and anonymous. The participation rate was 90%, resulting in a total N = 98. There was a notable gender imbalance with females outnumbering males by 4:1. All students were naïve to the experimental hypothesis and unfamiliar with the theory of unconscious thought.

2.1.2 Materials and procedure

The participants were tested in three groups of about 32 individuals. They were seated about 1.5 metres apart and were not permitted to talk during the experiment. In line with the procedure described in Dijksterhuis, Bos, Nordgren, and van Baaren (2006) participants were presented with 48 sentences describing four fictitious cars (the Hatsdun, Kaiwa, Dasuka and Nabusi) on 12 attributes (e.g., milage, handling, service) in either negative or positive terms (Table 1). The sentences were randomized and each sentence was presented for five seconds

https://doi.org/10.1017/S1930297500000863 Published online by Cambridge University Press

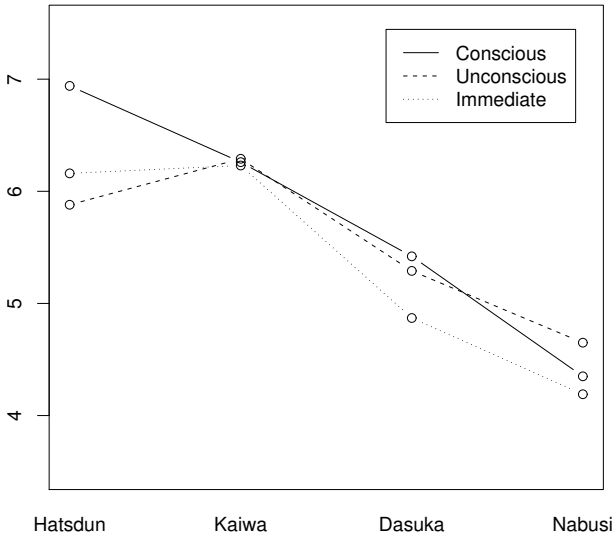

Figure 1: Mean ratings of the four cars by group

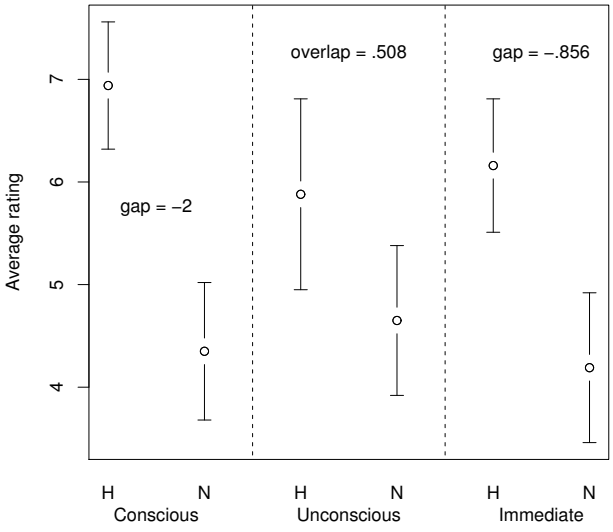

Figure 2: 95% confidence intervals for the difference scores between Hatsdun and Nabusi for groups Conscious, Unconscious, and Immediate, with proportions overlapping or showing a difference. Proportion overlap or gap is expressed in terms of the average of the half-widths of the two confidence intervals.

using *Microsoft Powerpoint*. The sentences were taken from the supplementary material published in the Dijksterhuis et al. (2006) study. All sentences were examined by two individuals to ensure that each sentence was intelligible to Australian subjects, and all were. Before the presentation, students were introduced to the names of the four cars and were told that they would have to pay attention to the material as they would later have to pass judgment on each of the cars. Response sheets were handed face-down to all participants which sequentially assigned participants to groups *c*, *u* or *i*. After the presentation, each individual filled in the response form, following the directions printed on the form.

In the *conscious thought* group *c* participants were instructed to think for four minutes about the different cars and their attributes before rating them. Participants in group *u* (*unconscious thought*) had to work on a word search task for four minutes after which they filled in their response sheet. Participants in group *i* (*immediate*) rated the cars right after the presentation finished. The ratings of the cars for all three groups had to be made on a 10-point rating scale, where 10 indicated the best possible rating. The sequence within which the cars appeared on the response sheet was also randomized separately for each subject. Thus only a few individuals within each group received the same rating sequence. Following their rating of the cars, participants were asked to indicate which two of the 12 presented attributes they considered to be the most important to themselves. The entire experiment, from giving instructions to the collection of the response sheets, took about 25 minutes per group.

2.2 Results

Out of the 96 participants, 32 were assigned to group *c*, 34 to group *u* and 30 to group *i*. The results clearly demonstrated the main effect for cars. In all three groups, the Hatsdun (best car based on number of positive features) was rated higher (m_c = 6.93, sd_c = 1.71; m_u = 5.88, sd_c = 2.66; m_i = 6.16, sd_i = 1.73) than the Nabusi (m_c = 4.35, sd_c = 1.85; m_u = 4.64, sd_c = 2.08; m_i = 4.19, sd_i = 1.95), which was the worst car based on number of positive features. Figure 1 illustrates this point. This difference was very pronounced for group *c*, somewhat less clear for group *i* and least evident for group *u* (see Figure 2).

The biggest surprise in these results was that the differences among the cars was smaller and the within group variance was larger for the unconscious decision makers than for the conscious ones.[1] This was exactly opposite to the predictions of the unconscious thought theory and the findings in the original experiment by Dijksterhuis et al. (2006). In fact, for group *u* the car with the second

[1]The within-group variance is of interest because, according to the theory, we would expect a homogeneous improvement for all participants following unconscious thought, and, especially, participants should be better equipped to differentiate clearly between different choice levels (distinguish best from second best, from third best, etc.). That was precisely not the case as the scores for this group were highly variable. Some rated the Hatsdun very low and the Nabusi very high and vice versa. It is this variability in the ratings for the different cars within each group that determines the width of the confidence interval in Figure 3.

https://doi.org/10.1017/S1930297500000863 Published online by Cambridge University Press

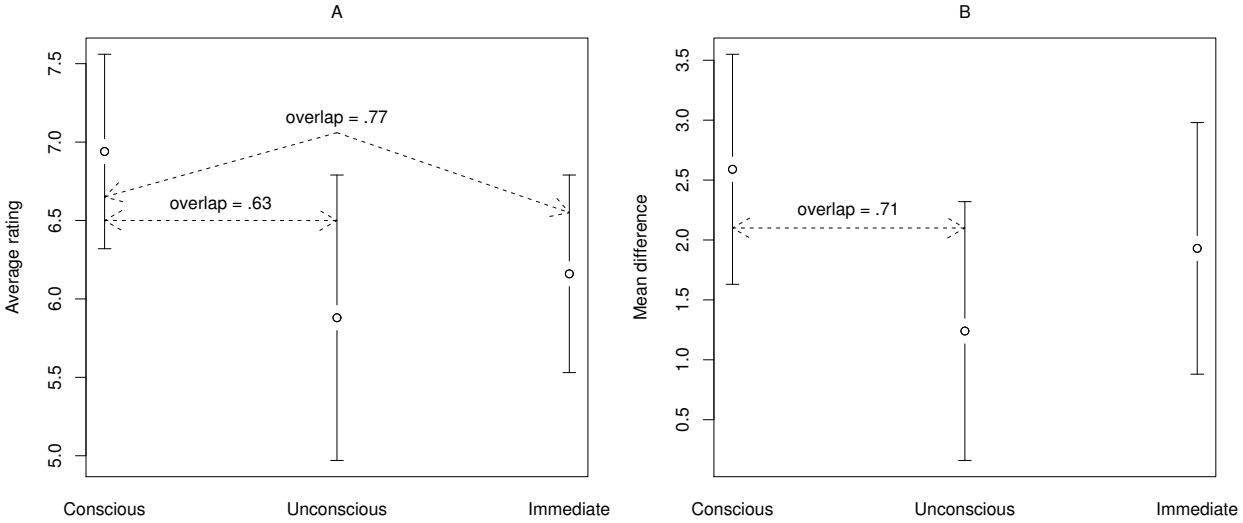

Figure 3: (A) 95% confidence interval showing the mean comparison for Hatsdun ratings between groups; (B) 95% confidence intervals showing the comparison of mean differences between H and N by group. Proportion overlap or gap is expressed in terms of the average of the half-widths of the two confidence intervals.

most positive attributes, the Kaiwa, was rated higher than the supposedly better car, the Hatsdun. Looking at the group differences for the Hatsdun only, it becomes clear that conscious thought was set apart from the other two groups. However, the overlap between *c* and *u* was still substantial, mostly due to the confidence interval width for *u* (see Figure 3a). A similar result ensued when comparing the mean difference scores between the Hatsdun and the Nabusi for each group. Again, the degree of variation was smallest for group *c* and largest for group *u*, but the overlap was still fairly substantial. Figure 3b illustrates this point.

One potential explanation for the surprising effect could be that the values for the Hatsdun were more extreme for group *c*, whereas participants in group *u* generally were more careful in their ratings. For example, participants in group *c* may have been more inclined to give a rating of 10 to their favourite car than participants in group *u*, who might have avoided the scale ends altogether. This hypothetical difference in scale usage would have resulted in a clear distinction of the normatively best car for group *c* and a very marginal distinction for group *u*. If this explanation were valid, then a conversion of scores into ranks should offset the effect, and the Hatsdun should have been rated as the best car most frequently in all three groups. Table 2 shows that this was not the case. Even after the conversion into ranks, group *c* showed much clearer results than group *u*. This suggests that individuals in the unconscious decision group were less sure about which car is best and, accordingly, score preferences fluctuated.

A second explanation for the results is that the importance individuals placed on particular features of cars differed across groups and therefore the Kaiwa was rated higher than the Hatsdun in group *u*. Since the feature valence for the Kaiwa and the Hatsdun were not identical — the Kaiwa, for example, was described as having more leg space than the Hatsdun — this may have influenced the results if leg space was an important feature to many participants. Note that this explanation is somewhat at odds with the unconscious thought theory, which posits that not so much individual features as the overall attractiveness of a choice option would be considered by unconscious thought in the decision making process.

Table 2: Rank averages for the four cars by group.

	Hatsdun	Kaiwa	Dasuka	Nabusi
Conscious	1.68	1.97	2.71	3.29
Unconscious	2.31	1.97	2.54	2.86
Immediate	2.00	1.93	2.60	3.10

The data that were collected on the two most important features for each participant allowed an educated guess about the validity of this explanation. As Table 3 illustrates, participants in groups *u*, *c*, and *i* had fairly similar feature preferences. Differences ensued only for those features that were only seldom selected as one of the two most important. Linking the results of Table 3 to Table 1, which shows the positive and negative attributes for each car, it is clear that differences in attribute valence between the Kaiwa and the Hatsdun, with the exception of *handling,* were restricted to the less important categories.

https://doi.org/10.1017/S1930297500000863 Published online by Cambridge University Press

Table 3: Distribution of feature importance rating across groups. Distributions are fairly similar for all three groups and clearly distinguishes between important and unimportant features.

Group	EN	CU	CO	SOU	SE	HA	MI	LS	TR	SF	GE	A
C	12	1	4	3	9	11	12	2	0	3	3	3
U	10	0	4	2	12	11	9	6	1	3	1	7
I	15	2	3	1	8	11	8	2	0	1	5	3

Note. EN=Environment, CU = Cupholders, CO = Colours, SOU = Sound system, SE = Service, HA = Handling, MI = Milage, LS = Leg space, TR = Trunk space, SF = Sunroof, GE = Gears, A = Age. The numbers show how often each feature was named as one of the two most important ones.

The important ones had the same valence for either car, and, in the case of *handling*, it was the Hatsdun that had a positive valence for this feature, not the Kaiwa. This result implies that, regardless of whether a decision about the better car was made on the basis of counting the number of positive attributes (Hatsdun = 9, Kaiwa = 7) or also by the value that a person assigns to the attribute into consideration, there should have been no difference between the two cars or, if anything, a preference for the Hatsdun.

Another transformation illustrates the discrepancy even further. For each of the attributes, a proportion based on the within group count was calculated that showed its importance relative to the other eleven. These proportions were then multiplied by the valence score for each attribute for each car (1 if the attribute was positive for the car and 0 if it was negative). These scores were then summed and averaged, using the number of positive attributes for the car, which yielded the mean importance score for each positive attribute for each car. Comparing these scores for each of the cars and groups (Table 4), it is clear that the differences across groups were minute and that, in fact, the cars with more positive attributes also were described in positive terms for those categories that were regarded as important by the participants. Again, the data do not provide any hints for the unexpected group differences between *c* and *u*.

A third potential explanation is that the sequence of statements during the presentation may have had a variable effect on the groups. This is unlikely since all groups saw the same sequence. With the data collected, however, it is also not possible to disprove this hypothesis. An inspection of the items showed only one permutation that could have lead to a primacy or recency effect. Out of the last eight items, four were negative statements about the Nabusi, yet only one of these four related to an important attribute.

In summary, the experimental manipulation for the cars was successful, in that the cars with more features were rated as the better ones, and some of these cars' features were also considered the most important, such as their environmental performance. The two cars with less positive features overall additionally had less desirable features, such as cupholders or trunk space. The clearest difference between the cars was obtained for the conscious thinkers, the least distinction was achieved by the unconscious thinkers. This result was not statistically conclusive due to the high variability within group *u*. Artefactual explanations such as primacy and recency effects or different preferences for features and hence cars by the various groups were rejected on the basis of the data.

Table 4: Mean feature importance for each car by groups and overall.

	Overall	C	U	I
Hatsdun	0.096	0.097	0.096	0.096
Kaiwa	0.096	0.095	0.093	0.099
Dasuka	0.063	0.063	0.061	0.065
Nabusi	0.041	0.040	0.049	0.034

3 Meta analysis

3.1 Method

3.1.1 Study selection

The purpose of the meta-analysis was to compile all relevant study data that bear on the benefit of unconscious thought for normative decision making. The studies carried out by Dijksterhuis and colleagues show high methodological homogeneity and largely reflect the method used in the empirical study described here. The focus of the meta-analysis was restricted to those studies that directly compared unconscious and conscious

https://doi.org/10.1017/S1930297500000863 Published online by Cambridge University Press

thought with respect to a decision making task after the presentation of a large amount of stimulus material for different choice alternatives.

Studies on incubation, which arguably also deal with unconscious thought, were not included. Incubation studies usually present the participant with a problem solving task (e.g., Vul & Pashler, 2007) or judge the creative output after a period of incubation (e.g., Dijksterhuis & Meurs, 2006). Both of these approaches are clearly different from the methodological selection criterion outlined and thus do not qualify for the meta-analysis.

Other studies that deal more directly with unconscious thought but also were not included are those that operationalise good decision making as post-choice satisfaction (e.g., Dijksterhuis & van Olden, 2006). Judging the degree of content with an item after a period of time does not appear to be functionally equivalent to making a relatively instantaneous decision based on comparative judgment. Thus, including studies which use post-choice satisfaction as the dependent variable would have introduced further error variance into the analysis beyond the normal sampling variability and ultimately biased the overall effect size estimate.

In order to be included a study had to focus on normative decision making, compared unconscious and conscious thought conditions, operationalise unconscious thought as a distraction period following a standardized encoding period, present each piece of information for the same amount of time or at least make provisions for the participants to do so, make participants choose from or evaluate multiple, specified choice options, and finally operationalise choice options in terms of degree of good decision making rather than choosing the correct out of a number of incorrect options.

Studies were sourced using the *psychinfo* database (1806-present) and *Google Scholar* with the keywords *"unconscious thought"*, *"decision making"*, and *"incubation"* and results were narrowed down with combinations of these. Additionally, references cited in Dijksterhuis (2004b), Dijksterhuis et al. (2006), and Dijksterhuis et al. (in press) were checked for compliance with the selection criterion. Finally, studies citing either Dijksterhuis (2004b) or Dijksterhuis et al. (2006) were checked.

Two sets of data were identified to be of potential usefulness to this analysis but could not be obtained. These were one data set pertaining to decision quality after varying intervals of unconscious thought (Dijksterhuis, 2004a, as cited in Dijksterhuis and Nordgren, 2006) and a study that used the same stimulus material as the present study (Dijksterhuis, Bos, van Baaren, & van der Leij, in prep., as cited in Dijksterhuis et al., in press). However, instead of a contrast between unconscious thought and conscious thought, it focuses on the comparison between immediate decision making and unconscious thought.

Overall, only two published studies comprising six experiments were deemed to sufficiently fulfill the original selection criteria. These experiments were number 1 to 4 in Dijksterhuis (2004b), Experiments 1 and 2 in Dijksterhuis et al. (2006). All other studies cited in this meta-analysis have not yet been published and had either been cited in work by Dijksterhuis or came to my attention through word of mouth. This class of unpublished works included five additional studies incorporating 10 sets of relevant data. These data were extracted from Experiments 1 and 2 of the Ham, Bos and Doorn study (Ham, Bos, & Doorn, submitted), Experiments 1, 2 and 3 of Newell et al.'s (in press) study, Experiments 1 and 2 reported by Payne et al. (2007), aggregate findings reported by Phillips et al. (2007), as well as from a study by Lerouge (submitted).

Experiment 2 of the Dijksterhuis et al. (2006) was the most comparable to the present empirical study. The main points of distinction were the difference in dependent variable scaling (50 point visual analogue versus 10 point rating) and the presentation time per item (8 versus 5 seconds). Experiment 1 (Dijksterhuis, Bos, Nordgren, & van Baaren, 2006) looked at proportional differences between participants in the unconscious and conscious thinking groups with respect to selecting the best choice alternative. In Dijsterhuis' 2004 study, Experiments 1 and 2 were fairly similar to the present experiment, albeit the stimulus material was different. Experiment 3 (Dijksterhuis, 2004b) was functionally similar to Experiment 2 but with yet again other stimulus materials. It also looked at the importance an individual places on the different information attributes that are presented and correlated this with the decision score. The present experiment had a similar index, however, in a more rudimentary form.

Experiment 4 (Dijksterhuis, Bos, Nordgren, & van Baaren, 2006) was not directly concerned with alternative selection but rather with the re-attribution of item content to its source. In comparison to the other experiments, the encoded items did not have to be integrated into global judgments but rather had to be remembered individually. The experiment also fostered response speed as a second dependent variable, but these results were not included here. The decision to include this experiment is debatable, but it seemed sufficiently suitable as the attribution still involved some decision making, although on an item-by-item level, after the same encoding procedure as in the other experiments.

Ham et al.'s experiments (Ham, Bos, & Doorn, submitted) were again functionally quite similar to the present experiment but focused on justice judgments rather than consumer choice. Newell et al. (in press) presented four experiments. The first three experiments were included in the meta-analysis, the fourth one was not. In all experiments Newell et al. collected data on a variety of depen-

Table 5: Overview of key features of the experiments included in the meta-analysis.

Expt.[1]	Groups[2]	Material type	Gender ratio (M/F)	n	Judgment type[3]	ES[4]	Number of attributes	Presentation	Filler task	Presentation time (sec)[5]	Interval (min)
1	c,u,i	Cars	0.250	32	1	1	12	Random	word search	5	4
2	c,u	Cars	?	20	2	2	12	Random	anagrams	8	4
3	c,u	Cars	?	13	1	1	12	Random	anagrams	8	4
4	c,u,i	Apartments	0.312	21	1	1	12	Random	n-back task	4	3
5	c,u,i	Apartments	0.175	31	2	2	15	Fixed	n-back task	12	3
6	c,u,i	Person	0.355	48	1	1	12	Random	anagrams	2	4
7	c,u,i	Person	0.295	38	3	3	N/A	Random	anagrams	2	4
8	c,u,i	Person	0.250	18	3	3	14	Fixed	n-back task	26	3
9	c,u,i	Person	0.500	17	3	3	14	Fixed	n-back task	26	3
10	c,u,i	Notebooks	?	21	1	1	12	Fixed	anagrams	20	4
11	c,u,i	Notebooks	?	21	1	1	12	Fixed	anagrams	20	4
12	c,u,i	Apartments	0.610	23	2	2	10	Random	anagrams	4	4
13	c,u,i,(c+)	Apartments	0.415	23	2	2	10	Fixed	anagrams	180	8
14	c,u,i	Cars	0.607	30	2	2	12	Random	anagrams	4	4
15	c,u,(c+)	Lottery	1.158	20	2	2	12	Random	anagrams	6	4
16	c,u,i,(c+)	Lottery	1.947	28	2	2	12	Random	anagrams	6	4
17	c,u	Cars	0.428	40	1	1	12	Collapsed	anagrams	8	4

[1] The sequence for this table is the same as in Figure 4 and Table 6.

[2] The combination (c+) indicates that there was a special condition involving conscious thought. Data from these conditions were not used here.

[3] Value "1" indicates use of rating scales for all options, value "2" stands for selection of a particular (usually best) alternative, value "3" are other, generic measures.

[4] ES stands for effect size estimator. Value "1" indicates difference between highest and lowest rated option, "2" indicates difference between percentages of correct choice in groups, and "3" indicates other, generic indices.

[5] Values >10 indicate that all information for a given option was presented simultaneously, the value >100 indicates that all information for all options were presented simultaneously.

dent measures, such as a recall test for attributes, but only the data on choice preferences were used for the meta-analysis. Experiment 3 is virtually identical with the empirical study presented here in that it is also a direct replication of Dijksterhuis et al. (2006) using the same materials. In contrast, Experiments 1 and 2 also used the same normative choice methodology but with original stimulus material. In Experiment 2 an additional experimental group is included: conscious thinkers that have access to the relevant information during the decision making period. Data from this group were not included in the meta-analysis but based on the results reported by Newell et al., their inclusion would not have much difference. Experiment 4 investigated the impact of primacy and recency effect on conscious and unconscious choices. It was unsuitable for inclusion to the meta-analysis as it only contained two choice alternatives but mainly because the two cars were both described by 10 positive attributes so that there was no normatively 'rational' choice.

Payne et al. (2007) carried out two experiments that explored the boundary conditions of the unconscious thought effect. In addition to the contrast between usual conscious thought and unconscious thought, they included a further condition where the decision interval for conscious thought was self-paced. Only the conscious thought condition with a fixed time interval between pre-

https://doi.org/10.1017/S1930297500000863 Published online by Cambridge University Press

sentation and decision was included in the meta-analysis. Both of Payne et al.'s experiments used a design whereby different numerical values were assigned to each piece of information and which resulted in different expected utilities for each choice option. Contrary to other studies, this method required successful and precise analytic integration of the values for each option, instead of choices based on the gist of all items.

The data set obtained from Phillips et al.'s study is part of a larger online study. Phillips et al. looked at, among other things, the effect of item presentation order, but the data presented here has been collapsed across all presentation order positions.

Finally Lerouge's study (submitted) dealt explicitly with consumer choice but primarily investigated the impact of configural versus featural processing as a moderator of unconscious thought. Only Experiment 2 was included, since Experiment 1 generated data only for immediate and unconscious thought, but not for conscious thought. All experiments listed here aside from Dijksterhuis et al. (2006) also included the immediate choice condition, which is of subordinate interest in the present study.

The selection of studies was homogeneous in terms of general methodological approach, albeit with much variation in the exact procedure. Table 5 provides an overview of some key aspects that lend themselves to meta-analytic investigation. A variety of decision processes and information conditions is represented. Decision making success is identified as choosing the best option and comparing the proportions across groups or by rating each choice option and then deriving some form of mean difference between groups. Ham et al.'s study presents an exception to these two patterns. Regarding the information presentation, most experiments presented the items one by one, but few presented multiple pieces of information simultaneously. Most studies included three conditions: immediate, conscious and unconscious decision making, but some (Newell et al., in press; Payne et al., 2007) had an additional condition to test specific predictions. The gender ratio also varied, although females predominated in most studies. There were minor variations with respect to the number of attributes for each choice option or the interval length after the presentation of the stimulus material (with the exception of Newell et al., Experiment 2).

3.1.2 Meta analytic procedure

For each of the selected experiments, standardized effect sizes (g) were calculated following the guidelines presented in Grissom and Kim (2005) for mean differences and DeCoster (2004) for proportion differences. In line with recommendations of Schmidt, Oh and Hayes (in press) a random effects model was chosen for the analysis. Weights and the mean effect size were calculated using the procedure described in Borenstein, Hedges, Higgins, & Rothstein (2008). Gender ratio, presentation time per item, and decision (or better distraction) interval span were defined as moderators. Since the moderator variable investigated in Lerouge (submitted) had an appreciable effect on the results, the experiment was treated as two separate data sets for the meta-analysis in order to preserve this noteworthy contrast. This decision did not affect the final estimates of the population effect size and margin of error. The data as presented below overestimate the parameters by about 0.006, in contrast to estimates that treat Lerouge's study as a single data set.

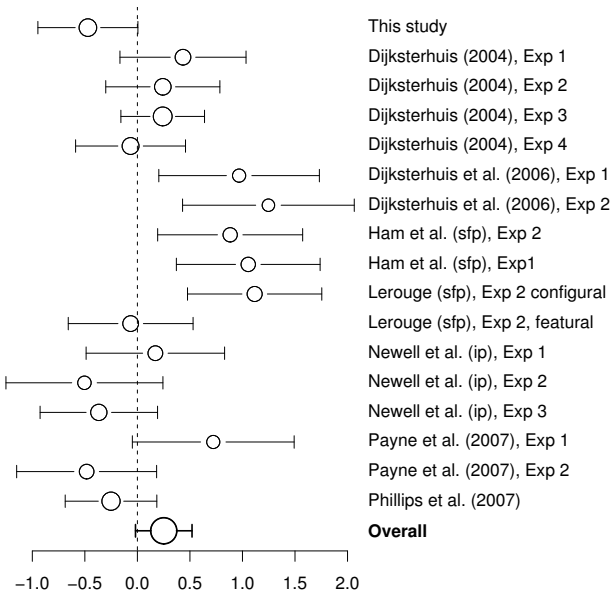

Figure 4: Forest plot of studies displaying effect sizes and 95% confidence intervals.

3.2 Results

3.2.1 Unconscious versus conscious thought

Altogether 17 experiments were included in the analysis with a combined participant number of 888. The mean effect size was g = .251, and the range from g = (-.483) in Payne et al.'s study (Experiment 2) to 1.25 (Dijksterhuis, Bos, Nordgren, & van Baaren, 2006, Experiment 2). Figure 4 shows a forest plot of effect sizes with respective confidence intervals by study and Table 6 provides the numerical effect size values, the standard error and the relative weights.

The amount of variability between the effect sizes was substantial ($Q[df$=16] = 54.994, $p \leq .000$; I^2= 70.906). Only five out of the included 17 experiments returned results that can be described as "statistically significant" in

https://doi.org/10.1017/S1930297500000863 Published online by Cambridge University Press

classical terms. Each of these five provided evidence for the superiority of unconscious thought. They also had the largest effect sizes but at the same time the smallest sample sizes. Newer data that still await publication provided evidence conflicting with the unconscious thought theory. The aggregate estimate shows a modest benefit for unconscious thought, although, from a significance-testing perspective, the confidence interval includes "0" and can thus be interpreted as non-significant support.

3.2.2 Moderator variables

Given the high level of study heterogeneity, several meta-regressions were carried out to investigate the effect of potential moderator variables. The statistical package *Comprehensive Meta Analysis*™ (Borenstein, Hedges, Higgins, & Rothstein, 2008) was used for this purpose.

Gender ratio: Dijksterhuis (2004), in his discussion of Experiment 1, found an interaction between gender and thought condition. Males were choosing particularly well following unconscious thought. Other studies have not specifically investigated this point, but it was worthwhile to follow up with a large set of data here. The vast majority of experiments exhibited a surplus of females, with the exception of Payne et al. (2007). The experiments by Lerouge (submitted) and Dijksterhuis et al. (2006) were not included as no data on gender were available. The regression analysis suggested that the gender ratio of a study is a poor predictor of effect size ($\beta = -0.214$, CI_{95} $[-0.786, 0.357]$, $SE = 0.291$).

Item presentation duration: The analysis of item presentation duration focused only on those studies that showed each piece of information individually. Studies (Dijksterhuis, 2004, Experiment 4; Ham et al., submitted, Experiments 1 and 2, Lerouge, submitted, both conditions; Newell et al., in press, Experiment 2) that showed items list-wise or all simultaneously were excluded. The analysis showed a trend that longer presentation times per item led to less advantage for unconscious thought, but this relationship was slight and did not reach statistical significance ($\beta = -0.095$, CI_{95} $[-0.232, 0.042]$, $SE = 0.07$).

Thought interval: All studies were included in this meta-regression. Similar to the item presentation variable, the results suggested that a longer interval between information presentation and decision is favourable for conscious thought rather than unconscious thought. Again, though, the result did not reach statistical significance ($\beta = -0.199$, CI_{95} $[-0.445, 0.048]$, $SE = 0.126$) and was strongly influence by Newell et al.'s (in press) second experiment, which had a substantially longer interval than all other studies.

Table 6: Effect sizes (g), standard errors (SE) and relative weights (w) for the experiments included in the meta-analysis. The abbreviation sfp means "submitted for publication"; ip, "in press."

Study name	g	SE	w
This study	0.471	0.243	6.743
Dijksterhuis (2004), Exp 1	0.434	0.306	5.998
Dijksterhuis (2004), Exp 2	0.242	0.277	6.340
Dijksterhuis (2004), Exp 3	0.241	0.203	7.205
Dijksterhuis (2004), Exp 4	0.065	0.267	6.459
Dijksterhuis et al. (2006), Exp 1	0.968	0.390	5.054
Dijksterhuis et al. (2006), Exp 2	1.247	0.417	4.774
Ham et al. (sfp), Exp 2	0.883	0.352	5.469
Ham et al. (sfp), Exp 1	1.055	0.349	5.503
Lerouge (sfp), Exp 2 configural	1.116	0.326	5.765
Lerouge (sfp), Exp 2, featural	−0.064	0.303	6.033
Newell et al. (ip), Exp 1	0.171	0.336	5.650
Newell et al. (ip), Exp 2	−0.504	0.381	5.150
Newell et al. (ip), Exp 3	−0.367	0.285	6.245
Payne et al. (2007), Exp 1	0.722	0.393	5.025
Payne et al. (2007), Exp 2	−0.483	0.340	5.604
Phillips et al. (2007)	−0.251	0.222	6.984
Overall	0.251	0.137	

Presentation format: A last moderator variable was the effect of presenting all pieces of information either individually or in clusters (as lists for each choice option or all simultaneously). This information was not available for Phillips et al.'s data, which was therefore excluded from the analysis. The results showed that the aggregate effect size was lower ($g = 0.147$, CI_{95} $[-0.037, 0.331]$, $SE = 0.094$) for experiments that presented the items individually ($n = 9$) than for the overall estimate. On the other hand, the results for studies that presented multiple pieces of information at the same time ($n = 6$) suggested a higher and most likely positive effect ($g = 0.369$, CI_{95} $[0.110, 0.627]$, $SE = 0.132$). This means that unconscious thought may actually be helpful when much information is presented simultaneously, but not when bits of information are presented individually.

https://doi.org/10.1017/S1930297500000863 Published online by Cambridge University Press

3.2.3 Unconscious thought versus immediate decision making

Altogether 13 out of the 17 data sets were included in a meta-analysis comparing decision making after unconscious thought with immediate decision making. The five excluded data sets did not feature the immediate decision conditions. These were Dijksterhuis et al. (2006, Experiments 1 and 2), Payne et al. (2007, Experiment 1) and Phillips et al. (2007). The results were similar to the comparison of unconscious and conscious thought. There was a modest trend in favour of unconscious thought (g = 0.189, CI_{95} [−0.05, 0.428]) but a significant amount of heterogeneity across studies ($Q[df=12] = 26.691$, p = .009; $I^2 = 55.041$). Further analyses to identify true moderator variables were not carried out.

3.2.4 Conscious thought versus immediate decision making

The same 13 data sets were used for the conscious thought versus immediate decision making meta-analysis. The results showed only a very slight advantage for conscious thought over immediate decision making (g = 0.084, CI_{95} [−0.72, 0.24]) with substantially more agreement among experiments than for the other comparisons ($Q[df=12] = 9.77$, p = .636; $I^2 = 0$).

4 Discussion

4.1 Empirical study

The results of the present experiment are noticeably at odds with the theory of unconscious thought. According to the obtained data set, it is a better idea to consciously think about different choice alternatives in order to arrive at the best choice than letting the unconscious do the work. Not only, it seems, does conscious thinking lead to the better identification of the top choice, but it also allows to differentiate between gradual choice options. With unconscious thought the choice is a muddy one; it did not consistently help individuals to differentiate between choice alternatives. Unconscious thought failed to allow a clear distinction of the cars; contrary to the expectations conscious thought did do just that.

While some alternative explanations for the divergence from the expected results were investigated (recency effects, weighing differences, scale usage differences), other factors in which the present study differed from Dijksterhuis et al. (2006) may have been responsible for the reversal of results. An obvious difference between this and the Dijksterhuis study was that this experiment was carried out with English material and Australian students. It is difficult to imagine, however, how cultural or linguistic variation could have had such a radical effect on decision making quality. Similarly, delivering the information to a group rather than on an individual basis is unlikely to have had any impact, especially since participants were engaged in the task and did not distract one another during the task presentation or the thought interval. One other noteworthy difference between this experiment and the original study is the distractor task used. Dijksterhuis et al. mainly used anagrams to keep participants in the unconscious thought condition busy; a word search task was used here. Anagrams have been widely used in the study of incubation (see for example Vul & Pashler, 2007) and seem to be positively related to the kind of processes activated during unconscious thought. Hence, anagram solving might have had a positive mediator function. Word search, on the other hand, does not have much in common with unconscious thought, as most people scan the array of letters systematically for the correct combinations. Hence it is more akin to conscious thought. However, in some studies Dijksterhuis (2004b) and also Ham (Ham, Bos, & Doorn, submitted) used the n-back task (Kane, Conway, Miura, & Colflesh, 2007). This task puts high demand on executive functioning and also can hardly be conceived as supporting unconscious thought. Nevertheless, the possibility that the type of distractor tasks affects decision making efficiency under unconscious thought conditions warrants further study and could be used as a moderator variable in future meta-analyses when more experiments with distractor tasks other than anagrams are available. If all these three alternative explanations can be discounted, then the results of the present study provide strong evidence that the true effect size for unconscious thought is much smaller than assumed so far or that this particular experimental approach is not very suitable to demonstrate the unconscious thought effect reliably. Further support for either of these two conclusions comes from the meta-analytic findings presented here.

4.2 Meta-analysis

The statistical synthesis of all available data provides at best suggestive evidence in favour of unconscious thought, but, on the basis of 888 subjects tested under similar conditions, there is no convincing statistical evidence. The true effect in the population may be anything between a moderate benefit after unconscious thought to a slight advantage following conscious thought. One sign for caution is that the experiments with fewer participants consistently generated substantially larger effect sizes than the larger studies.

https://doi.org/10.1017/S1930297500000863 Published online by Cambridge University Press

4.3 Moderator variables

Four moderator variables were investigated in the present meta-analysis. Of these, only the presentation format as either single item or list-wise, did help to explain the variance between studies, which is substantial. On the other hand, gender-ratio, presentation time per item and thought interval length were very weak predictors of effect. This weakness, however, may have partially been due to methodological constraints. For example, the moderator effect for gender ratio looked at the total number of males and females in the experiment. A better estimator would be the ratio of males to females within the specific conditions, since there may be interactions that cannot be uncovered by looking at the total number for each gender alone. Secondly, there may be higher order interactions between different moderators that obscure any simple effects. This meta-analysis carried out only simple meta-regressions, but there is reason to believe that multiple meta-regression may find relationships, given that I^2 turned out to be large. Before such analysis is feasible, though, more studies are required to ensure the statistical soundness of the analysis (Borenstein et al., 2008).

Beside the ones investigated here, further potential moderators have been identified. Configural versus featural processing (Lerouge, submitted) or self-paced decision making (Payne et al., 2007) and perhaps the type of distractor task are promising candidates. These will require much more data do make them suitable for meta-analysis. Further, the present selection of experiments was restricted to normative choices, but this in itself may be a factor affecting the efficiency of unconscious or conscious thought. Future approaches should thus include studies that operationalise good decision making with subjective degrees of decision satisfaction. Finally, the meta-analytic confirmed effect of presentation format should be investigated in a specifically designed study. At this stage, with few experiments, the result may still be spurious, especially since higher order interaction of moderators and task conditions are unclear.

4.4 Other comparisons

The primary focus of the present study was to illuminate the current state of affairs regarding unconscious versus conscious thought. The comparison between unconscious or conscious thought and immediate decision can further help to draw some useful conclusions. On the basis of 13 sets of data, there is no statistical evidence to suggest that either form of thought following stimulus presentation leads to better choices than immediate decision making. This, again, is quite surprising given prior evidence for conscious thought processing as represented in, for example, the weighted-additive model of decision making (Keeney & Raiffa, 1976). The fact that the meta-analytic review is not able to distinguish between the three alternative ways of decision making at all reinforces two possibilities: 1., other variables such as presentation format or even gender significantly interact with decision making condition; or, 2., the general approach is not suitable to differentiate modes of thought and might instead reflect other influences (see, for example, Newell et al., in press, Experiment 4).

As shown by the present study, the early finding that unconscious thought leads to normatively better decisions than conscious thought (Dijksterhuis, 2004, Experiment 1), which provided the starting point for further developments of the theory of unconscious thought, is unsupported when looked at from a meta-analytic perspective. However, this insight became possible only through the inclusion of many data sets that are yet to be published. All currently published data, with the exception of the present empirical study, showed uniform support for the theory of unconscious thought. The actual variability in results that was demonstrated by the inclusion of additional, unpublished data thus reinforces the need to get more results out in the open.

4.5 Conclusion

In summary, the findings reported here in conjunction with the ones found in Newell et al. (in press), Payne et al. (2007), Phillips et al. (2007) and Lerouge (submitted) cast doubt on the unqualified claim that unconscious thought is the superior way of processing information for important and complex choices. So far the bulk of studies had provided confirmatory evidence for unconscious thought in different contexts such as justice and consumer choice, in the laboratory as well as in more naturalistic studies, and with different dependent variables. Taken together, these studies were certainly bolstering the divergent validity of the theory of unconscious thought. However, with the inclusion of new results, the theory will have to explain the effects of moderator variables as shown in Lerouge (submitted) or Payne et al. (2007) and, at least for normative choices, will have to step away from choice model that identifies the rational choice as the one with most positive attributes (see Newell et al.) if it is to retain internal validity and produce correct predictions. For the time being there is little guarantee that unconscious thought will truly be a good aid when faced with complex and important decisions.

https://doi.org/10.1017/S1930297500000863 Published online by Cambridge University Press

References

Bekker, H. L. (2006). Making choices without deliberating. *Science, 312,* 1472.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2008). Comprehensive Meta-Analysis (Version 2.2.046) [Computer software]. Englewood, NJ: Biostat.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2008). Introduction to meta-analysis. *Manuscript in preparation*.

DeCoster, J. (2004). Meta-analysis notes. Retrieved 01, 11, 2007, from http://www.stat-help.com/notes.html

Dijksterhuis, A. (2004a). [decision quality after different amounts of unconscious thought]. *Unpublished raw data*.

Dijksterhuis, A. (2004b). Think different: The merits of unconscious though in preference development and decision making. *Journal of Personality & 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., Bos, M. W., van Baaren, R. B., & van der Leij, A. (2007). Unconscious thought and the weighting problem in decision making. *Manuscript in preparation*.

Dijksterhuis, A., & Meurs, T. (2006). Where creativity resides: The generative power of unconscious thought. *Consciousness and Cognition, 15,* 135–146.

Dijksterhuis, A., & Nordgren, L. F. (2006). A theory of unconscious thought. *Perspectives on Psychological Science, 1,* 95–109.

Dijksterhuis, A., van Baaren, R. B., Bongers, K. C. A., Bos, M. W., van Leeuwen, M. L., & van der Leij, A. (in press). The rational unconscious: Conscious versus unconscious thought in complex consumer choice.

Dijksterhuis, A., & van Olden, Z. (2006). On the benefits of thinking unconsciously: Unconscious thought increases post-choice satisfaction. *Journal of Experimental Social Psychology, 42,* 627–631.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (in press). G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*.

Grissom, R. J., & Kim, J. J. (2005). *Effect sizes for research: A broad practical approach*. London: Lawrence Erlbaum.

Ham, J., Bos, K. v. d., & Doorn, E. A. V. (submitted for publication). Lady justice thinks unconsciously: Evidence for more accurate justice judgments without conscious thought.

Kane, M. J., Conway, A. R., Miura, T. K., & Colflesh, G. J. (2007). Working memory, attention control, and the n-back task: A question of construct validity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 615–622.

Keeney, R. L., & Raiffa, H. (1976). *Decisions with multiple objectives: Preference and value tradeoffs*. New York: Wiley.

Lerouge, D. (submitted for publication). Evaluating the benefits of distraction on product evaluations: The mindset effect. *Journal of Consumer Research*.

Newell, B. R., Wong, K. Y., Cheung, C. H. J., & Rakow, T. (in press). Think, blink or sleep on it? The impact of modes of thought on complex decision making. *Quarterly Journal of Experimental Psychology*.

Payne, J. W., Samper, A., Bettman, J. R., & Luce, M. F. (2007). Boundary conditions on unconscious thought in complex decision making. *Manuscript in preparation.*

Phillips, Nathaniel D., González-Vallejo, C., Bellezza, F. S., Chimeli, Janna., Harman, J., Lassiter, G. D., & Lindberg, Matthew J. (2007). [Comparing the effects of distraction and deliberation in an online attitude updating task]. *Unpublished raw data*.

Schmidt, F. L., Oh, I.-S., & Hayes, T. L. (in press). Fixed versus random effects models in meta-analysis: Model properties and an empirical comparison of differences in results. *British Journal of Mathematical and Statistical Psychology*.

Shanks, D. R. (2006). Making choices without deliberating. *Science, 313,* 760.

Vul, E., & Pashler, H. (2007). Incubation benefits only after people have been misdirected. *Memory and Cognition, 34,* 701–710.