On making the right choice: A meta-analysis and large-scale replication attempt of the unconscious thought advantage

Mark R. Nieuwenstein[*]   Tjardie Wierenga[†]   Richard D. Morey[†]   Jelte M. Wicherts[‡]

Tesse N. Blom[§]   Eric-Jan Wagenmakers[§]   Hedderik van Rijn[†]

Abstract

Are difficult decisions best made after a momentary diversion of thought? Previous research addressing this important question has yielded dozens of experiments in which participants were asked to choose the best of several options (e.g., cars or apartments) either after conscious deliberation, or after a momentary diversion of thought induced by an unrelated task. The results of these studies were mixed. Some found that participants who had first performed the unrelated task were more likely to choose the best option, whereas others found no evidence for this so-called unconscious thought advantage (UTA). The current study examined two accounts of this inconsistency in previous findings. According to the reliability account, the UTA does not exist and previous reports of this effect concern nothing but spurious effects obtained with an unreliable paradigm. In contrast, the moderator account proposes that the UTA is a real effect that occurs only when certain conditions are met in the choice task. To test these accounts, we conducted a meta-analysis and a large-scale replication study (N = 399) that met the conditions deemed optimal for replicating the UTA. Consistent with the reliability account, the large-scale replication study yielded no evidence for the UTA, and the meta-analysis showed that previous reports of the UTA were confined to underpowered studies that used relatively small sample sizes. Furthermore, the results of the large-scale study also dispelled the recent suggestion that the UTA might be gender-specific. Accordingly, we conclude that there exists no reliable support for the claim that a momentary diversion of thought leads to better decision making than a period of deliberation.

Keywords: unconscious thought, deliberation without attention, decision making, meta-analysis, publication bias, funnel plot, large-scale replication study, Bayes factor.

1 Introduction

While research on human judgment and decision making has yielded many findings that suggest that the best way to make a difficult choice is to think carefully about the options and their consequences (e.g., Baron, 2008; Kahneman, 2011), the theory of unconscious thought (Dijksterhuis & Nordgren, 2006) proposes that this is not necessarily the best way to make a difficult choice. Rather, this theory proposes that the best way to make a difficult decision is to refrain from painstaking conscious deliberation and to let one's unconscious mind solve the problem while one engages in more enjoyable activities such as solving a cross-word puzzle. More specifically, this theory claims the existence of an unconscious form of thought that has

a much greater information-processing capacity than conscious thought. As a result, a momentary diversion of attention would benefit making a difficult decision because it allows the clever unconscious mind to take charge and solve the problem at hand.

So should decision makers really be told—as they have been (BBC News, 2006; Hoare, 2012)—to refrain from conscious deliberation and to rely on their unconscious minds in making difficult decisions? Research examining this matter began with seminal studies in which Dijksterhuis and colleagues (Dijksterhuis, 2004; Dijksterhuis, Bos, Nordgren, & Van Baaren, 2006) presented participants a large number of properties of different choice options (e.g., cars, candidate roommates, apartments), and then asked the participants to select the best option either after a period of conscious deliberation or after performing an unrelated task. (See Figure 1 for a graphical depiction of the paradigm.) Although the statistical analyses reported by Dijksterhuis and colleagues were suboptimal (e.g., Hasselman, Crielaard, & Bosman, *submitted*; Nieuwenstein & Van Rijn, 2012), the results of some of these experiments did show that participants who had first performed the unrelated task were more likely to select the
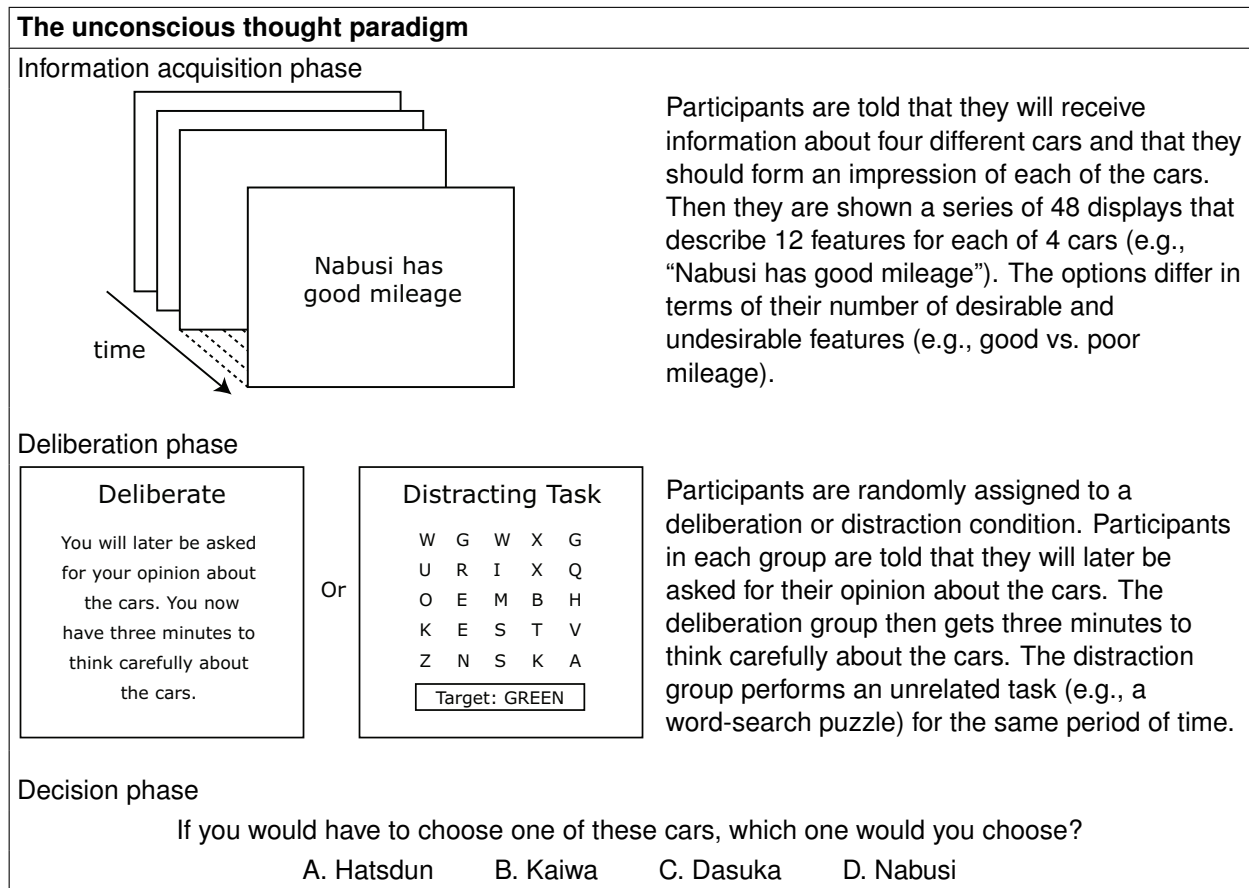
[*]University of Groningen, The Netherlands, Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands. Email: m.r.nieuwenstein@rug.nl.

[†]University of Groningen, The Netherlands.

[‡]Tilburg University, The Netherlands.

[§]University of Amsterdam, The Netherlands.

Figure 1: The paradigm that was introduced by Dijksterhuis (2004) to examine the potential benefits of distraction in complex decision making.

The unconscious thought paradigm

Information acquisition phase

Nabusi has good mileage

time

Participants are told that they will receive information about four different cars and that they should form an impression of each of the cars. Then they are shown a series of 48 displays that describe 12 features for each of 4 cars (e.g., "Nabusi has good mileage"). The options differ in terms of their number of desirable and undesirable features (e.g., good vs. poor mileage).

Deliberation phase

Deliberate

You will later be asked for your opinion about the cars. You now have three minutes to think carefully about the cars.

Or

Distracting Task

W	G	W	X	G
U	R	I	X	Q
O	E	M	B	H
K	E	S	T	V
Z	N	S	K	A

Target: GREEN

Participants are randomly assigned to a deliberation or distraction condition. Participants in each group are told that they will later be asked for their opinion about the cars. The deliberation group then gets three minutes to think carefully about the cars. The distraction group performs an unrelated task (e.g., a word-search puzzle) for the same period of time.

Decision phase

If you would have to choose one of these cars, which one would you choose?

A. Hatsdun   B. Kaiwa   C. Dasuka   D. Nabusi

best option than participants who were given the opportunity to deliberate—a phenomenon termed the unconscious thought advantage (UTA; Dijksterhuis, 2004; Dijksterhuis et al., 2006; see also Dijksterhuis & Nordgren, 2006). Following these reports, many other researchers attempted to replicate the finding of an UTA (e.g., Acker, 2008, who reviewed results available in 2008) and the results of these replication attempts were mixed, as they were split almost evenly between studies that did and did not find evidence for the UTA. (For a recent overview, see Nieuwenstein & van Rijn, 2012.)

2   The current study

In the current study, we contrast two explanations for the inconsistent results of previous studies examining the UTA. According to the reliability account, the UTA does not exist and previous reports of this effect concern nothing but spurious differences obtained from an unreliable paradigm. In contrast, the moderator account proposes that the UTA is real but observed only when specific con-

ditions are met in the choice task. In the following sections, we first elaborate on the argumentation underlying these accounts before turning to the approach we took to adjudicate between them.

2.1   The reliability account

The reliability account was already hinted at in one of the early studies that failed to replicate the UTA. In this study, Acker (2008) conducted a meta-analysis on 17 experiments that were available at that time. The analysis showed that only five of these experiments reported a statistically significant UTA effect. Furthermore, Acker found that these experiments had "the largest effect sizes but at the same time the smallest sample sizes" (p. 299; Acker, 2008), thus raising the possibility that the results found in these studies concerned spurious effects (see also Bakker, Van Dijk, & Wicherts, 2012; Newell & Rakow, 2011; Rothstein, Sutton, & Borenstein, 2005).

Indeed, the unconscious thought paradigm illustrated in Figure 1 has three properties that together seem to make a potent recipe for spurious results, especially with small

sample sizes. To start, the paradigm involves a complex task for which performance is likely to depend on a host of factors that can differ across time and participants, including concentration, mindset, gender, motivation, expertise about the choice at hand, attention and memory. Secondly, the paradigm uses a between-subjects manipulation of mode of thought with random assignment, meaning that the effect of the distraction vs. deliberation manipulation is assessed by comparing the performance of different participants. Thirdly, the performance measure for the task stems from only a single observation for each participant, meaning that each participant carries out the task only once, without practice. Arguably, this combination of properties makes a potent recipe for spurious results because the use of random assignment does not necessarily guarantee an equal distribution of task-relevant factors across two groups of participants, especially when the number of such factors is large (Hsu, 1989; Krause & Howard, 2003), as would seem to be the case in the unconscious thought paradigm. Moreover, the use of a single-trial design entails that the performance measure derived for each participant is bound to be an unreliable index of true, mean performance of that participant. Accordingly, it seems clear that the reliability and validity of results of studies examining the UTA hinges critically on whether these studies used a sample size that was sufficiently large to balance out the many potential confounding factors in the comparison of performance in the deliberation and distraction conditions. By implication, it stands to reason that the small-sample studies that found a statistically significant difference in performance in the deliberation and distraction conditions concerned a spurious difference.

2.2 The moderator account

In contrast to the reliability account, the moderator account proposes that the UTA is a real effect that is observed only when certain conditions are met with regard to the choice task. This account was proposed in a recent meta-analysis that was conducted by proponents of the theory of unconscious thought (Strick, Dijksterhuis, Bos, Sjoerdsma, & Van Baaren, 2011). The analysis included a large collection of published and unpublished data sets and it examined a large number of potential moderators of the UTA, including seemingly trivial methodological details such as whether the distracting task involved a word-search puzzle or an anagram task. The results yielded a pooled effect size of .218 (CI: .130-.307, $p < .01$), suggesting that, overall, a benefit of distraction in making complex choices does exist. Furthermore, many of the moderator variables included in the analysis indeed had a significant effect on the magnitude of this benefit (see Table 1). Specifically, the effect size of the UTA was found to depend on the complexity of the choice problem, the type of

goal participants were led to adopt during the information acquisition phase of the task, the manner in which the information about the choice alternatives was presented, the duration of the deliberation or distraction phase, and the nature of the task that was used to divert attention in the distraction condition. Accordingly, Strick et al. concluded that the UTA is real but the occurrence of this effect requires that certain conditions be met, as indicated by the results of the moderator analyses.

3 Outline of the current study

In the current study, we set out to adjudicate between the reliability and moderator accounts. To this end, we conducted a large-scale replication study that met each of the conditions found to yield a strong effect in the meta-analysis by Strick et al. (2011; see Table 1), and we conducted a meta-analysis that moved beyond the analysis by Strick et al. by examining the relationship between sample and effect sizes using a funnel-plot (i.e., a plot that depicts effect sizes against a measure of study precision that is directly related to sample size, such as the inverse of the standard error; e.g., Egger, Smith, Schneider, & Meyer, 1997; Light & Pillemer, 1984). According to the reliability account, previous findings of a significant benefit of distraction concern nothing but a spurious result, and, therefore, these findings would be expected to be confined to studies that used relatively small sample sizes because the probability of a spurious effect should decrease with increasing sample size. Furthermore, the reliability account also predicts that our large-scale replication study should show no significant UTA, in spite of the fact that the design of this study adhered to the recommendations provided by Strick et al.'s (2011) meta-analysis. In contrast, the moderator account would predict that the UTA should also be observed in studies that used a relatively large sample size, provided that they met the conditions under which the UTA is expected to occur (Strick et al., 2011). Thus, according to the moderator account, our large-scale replication study would also be predicted to reveal the UTA.

4 The large-scale replication study[1]

The starting point for the large-scale replication study was a recent study in which Nieuwenstein and Van Rijn (2012) conducted a first test of the moderator account and found a number of results that warranted further empirical confirmation. In this earlier study, Nieuwenstein and Van Rijn

---

[1]In this article, we report all measures, conditions, data exclusions, and the factors underlying the determination of sample size for new results.

Table 1: Moderators of the UTA identified in the meta-analysis by Strick, Dijksterhuis, Bos, Sjoerdsma, & Van Baaren (2011), and the manner in which these conditions were incorporated in the current large-scale replication attempt (see also Nieuwenstein & Van Rijn, 2012).

Factor	Description	Current study
Mindset	The UTA is larger when participants are led to adopt a configural mindset during the information acquisition phase. This entails that they should be instructed to form a global impression of the options.	☑
Pictorial information	The UTA is larger when verbal and pictorial information are combined in presenting the options during the information acquisition phase.	☑
Presentation format	The UTA is larger when the information about the choice options is presented grouped per option, as opposed to in a random order.	☑
Complexity	The UTA is larger for more complex decision problems. Complexity was defined by Dijksterhuis and Nordgren (2006) as the total number of attributes involved in a choice. Choices involving 4 options with 4 attributes are considered to be simple while choices involving 3 or more options with 10 or more attributes are considered to be complex.	☑ (4x12)
Presentation time	The UTA is larger when the attributes of the options are presented for a relatively short duration. The range of presentation times used in previously published studies is 2–14 seconds.	☑ (2.5 sec)
Goal	The UTA is larger when participants are told that they will later need to make a decision or judgment about the options at hand.	☑
Distracting task	The UTA is larger in studies that used a word-search puzzle (as opposed to an anagram or n-back task) as the distracting task during the UT period.	☑
Duration deliberation phase	The UTA is larger when the duration of the deliberation phase is relatively short. The range of durations used in previous studies is 3–8 minutes.	☑ (3 min. or self-paced)

used a task that met the conditions under which the UTA should be strong according to Strick et al. (2011), with the contrast between deliberation and distraction implemented as a within-subjects design so as to preclude the possibility that any observed UTA could be due to a spurious between-group difference. The results of four such experiments did not yield a statistically significant UTA effect, suggesting that even when all the moderator conditions identified by Strick et al. are met, the UTA is either small or does not occur at all. Importantly, however, these experiments used a relatively small sample size (24-48 participants), and the experiment that used the largest sample size (N = 48) did show a non-significant difference in the direction of the UTA. Furthermore, the results also suggested that perhaps the UTA is gender-specific, as a post-hoc exploratory analysis across all four experiments yielded a significant interaction of mode of thought and gender, with male participants showing a statistically significant conscious thought advantage while female par-

ticipants showed a non-significant trend towards an UTA. Lastly, the results of these experiments also suggested that insofar as the UTA indeed exists, it might occur only when the duration of the deliberation phase in the conscious deliberation condition is fixed at several minutes. Specifically, the results showed that participants needed only 30 seconds to deliberate about their choice, and they also provided evidence to suggest that performance in the conscious deliberation condition is better when the deliberation phase is self-paced, as opposed to fixed and unnecessarily long (see also, Payne, Samper, Bettman, & Luce, 2008).

Given the concerns about the reliability of results obtained in the unconscious thought paradigm, and given the post-hoc nature of the exploratory analyses that suggested that the UTA might be gender-specific, it is clear that the results reported by Nieuwenstein and Van Rijn (2012) warrant a more powerful test with a larger group of participants. To this end, the current study replicated the

first experiment in Nieuwenstein and Van Rijn—i.e., the one that showed a non-significant difference in the direction of the UTA—with a sample of participants that was nearly an order of magnitude larger (N = 399) than the sample used by Nieuwenstein and Van Rijn, thus offering a much more powerful test of the UTA[2] and the potential moderating role of gender. Furthermore, this large-scale replication attempt also used a within-subjects design for the comparison of the deliberation and distraction conditions, with the order of these conditions counterbalanced across participants. In addition, the experiment included two versions of the deliberation condition that differed in whether the duration of the deliberation phase was fixed or self-paced, thus allowing us to verify if performance in the deliberation condition—and perhaps the occurrence of the UTA—indeed depends on the duration of the deliberation phase. The duration of the deliberation phase was varied between subjects, and we used two different choice sets for the two choices that were to be made by each participant (i.e., a choice between four cars or four apartments), with a random distribution of these choice sets across the two choice conditions.

4.1 Methods

4.1.1 Participants

The study was conducted as part of a test session at the University of Amsterdam[3] in which all first-year undergraduates in Psychology could participate on a voluntary basis to obtain course credit. The number of students who took part in the study was 423 and this sample included 24 non-native speakers of Dutch, whose data were excluded from analysis. Exclusion of these participants did not change the results. The remaining 399 participants were 19.7 years old on average (SD = 1.86 years), and they included 130 males.

4.1.2 Materials

The experiment was conducted on a computer, using a program written in Adobe Authorware. The experiment comprised two choice tasks and a word-search task. The word-search puzzle task was used to distract participants during the unconscious deliberation phase.

For each of the choice tasks, participants received information about four options—cars or apartments—that were described in terms of twelve properties that could be desirable or undesirable.[4] The quality of the options was defined in terms of their number of desirable properties, such that the best option had 9 desirable properties whereas two intermediate options each had 6 desirable properties, and the worst option had only 3 desirable properties. During the information acquisition phase, these properties were presented one after the other in a series of timed displays that each included the fictitious name of the option, a sentence describing a property of the option, and a picture of the choice option. The pictures depicted real cars and apartment buildings (see also Nieuwenstein & Van Rijn, 2012). The word-search puzzle task comprised a 10x10 array of letters that was shown together with a target word. The letters were indexed by the numbers 1–100 and the task for the participants was to find the target word and type in the numbers that corresponded to the first and last letter of the word. The target words denoted countries, vegetables, or fruits, and could be written in the array in any direction.

4.1.3 Procedure

At the start of the study, the participants practiced the word-search puzzle task they would later be asked to do again during the unconscious deliberation phase. After practicing this task for one minute, the participants were informed that they would now see a presentation about four [cars/apartments] that would each be described in terms of different properties. In accordance with the recommendations by Strick et al. (2011), participants were instructed that they should form a good impression of each of these options. They were then shown a sequence of 48 displays of the options and their properties. The properties were presented grouped by option and the twelve properties were presented in the same order for each of the four options. The duration of each display was set at 2.5 seconds. In the distraction condition, this information acquisition phase was followed by an instruction telling the participants that they would later be asked for their opinion about the options and that they would first have to do the word-search puzzle task for a period of three minutes. In the deliberation conditions, participants were also told that they would later be asked for their opinion about the options, and they were instructed that they would first get three minutes (fixed deliberation phase) or as long as they needed (self-paced deliberation phase) to think carefully about the options. During this period, the pictures and names of the options remained in view, together with

---

[2]It is worth noting that, if the effect size of the UTA is 0.218, as suggested by the meta-analysis by Strick et al. (2011), then one needs a sample size of 175 participants to acquire a power of .8 in a within-subjects comparison, or a sample size of 548 for a between-subjects comparison. These estimates are based on a power computation for a one-tailed Wilcoxon signed ranks test for two proportions. Computation was done using G-power (Faul, Erdfelder, Buchner, & Lang, 2009), retrieved from http://www.psycho.uni-duesseldorf.de/abteilungen/aap/gpower3/.

[3]The seminal studies by Dijksterhuis and colleagues (Dijksterhuis, 2004; Dijksterhuis et al., 2006) were also conducted with undergraduates of the University of Amsterdam.

[4]These stimuli were also used in the first two studies that found evidence for the UTA (Dijksterhuis, 2004; Dijksterhuis et al., 2006; for a detailed description of the stimuli, see Nieuwenstein & Van Rijn, 2012).

Table 2: Number of participants (N) included in each of the four versions of the task.

Order of choice conditions	Duration conscious deliberation phase	N
Deliberation—Distraction	Fixed	99
Distraction—Deliberation	Fixed	103
Deliberation—Distraction	Self-paced	97
Distraction—Deliberation	Self-paced	100

a counter that indicated the passage of time in seconds. In the self-paced deliberation condition, the same display was shown but now participants could press a designated key once they had made up their mind. At this point, participants received the instruction to select the best option by pressing a corresponding key on the keyboard. In the fixed deliberation condition, this instruction appeared automatically after three minutes had passed. After selecting the best option, participants were asked to indicate on a 10-pt. scale how confident they were about their choice. In addition, participants in the deliberation condition with a fixed 3-minute deliberation phase were asked to estimate how long they had needed to arrive at a decision. For participants in the self-paced deliberation condition, the program registered how long it took before they indicated they had made up their mind.

4.1.4   Design

Each participant made one choice after conscious deliberation and one choice after doing the word-search task, and the order of these conditions was counterbalanced across participants. For half the participants, the duration of the deliberation phase in the conscious deliberation condition was fixed at 3 minutes and it was self-paced for the other participants. The duration of the word-search task that was used to induce a diversion of thought in the distraction condition was three minutes for all participants. The two orders of the deliberation and distraction conditions and the two durations of the deliberation phase were crossed to create four different versions of the task, and participants were randomly assigned to one of these four versions (see Table 2). The two choice sets (cars and apartments) were randomly assigned to the deliberation and distraction conditions, yielding a balanced design of within and between-subject factors.

4.1.5   Data-analysis

The plan for data-analysis was to examine accuracy on the choice task for main effects and interactions of mode of

thought (deliberation vs. distraction), gender (male vs. female), and the duration of the deliberation phase in the deliberation condition (fixed vs. self-paced). Choice accuracy was defined in terms of whether a participant selected the option with the greatest number of desirable properties, as is typically done in this paradigm. Since this outcome has a binomial distribution, the data were modelled using a logit function and analyzed using a generalized linear model (GLM). The effects that were tested using the GLM were estimated using generalized estimating equations so as to allow for the possibility that the observations could be correlated across the within-subjects factor of mode of thought. The confidence ratings were treated as an ordinal variable and analyzed for the same effects using a GLM.

4.2   Results

As a first step in analyzing the data, we examined how long participants needed to deliberate about their choice in the fixed and self-paced conscious thought conditions, and we examined if choice accuracy in this condition depended on whether the duration of the deliberation phase was self-paced or fixed at three minutes. The analysis of deliberation time showed that on average, participants in the self-paced condition took only 23 seconds to deliberate ($SD = 19.4$, *95% CI* = [20.5; 26.1]). In addition, this analysis showed that there was no significant relationship between choice accuracy and deliberation time, with the mean deliberation times being 25.0 ($SD = 25.4$, *95% CI* = [20.0; 30.7]) and 21.7 seconds ($SD = 13.1$, *95% CI* = [13.5; 24.4]), respectively, for participants who made an incorrect or correct choice ($t[195] = 1.17$, $p = .24$, Cohen's $d = .17$). A similar result was found for participants for whom the duration of the deliberation phase was fixed at three minutes. To be precise, these participants reported that they had needed 37 seconds ($SD = 31.0$, *95% CI* = [32.7; 41.4]) on average to deliberate, and for these participants too, self-reported deliberation time did not differ between participants who made a correct or incorrect choice, $M = 37.7$ ($SD = 30.1$, *95% CI* = [31.3; 44.3]) vs. $M = 37.1$ ($SD = 31.8$, *95% CI* = [32.1; 42.7]) seconds respectively, $t(200) = .15$, $p = .88$, Cohen's $d = 0.02$. Lastly, a comparison of choice accuracy in the deliberation conditions with a self-paced and fixed deliberation phase showed no significant effect of the duration of the deliberation phase, with the percentage of correct choices being 59.4 and 56.9%, respectively, for the fixed and self-paced conditions, $Z = .52$, $p = .61$.

The main analysis of interest examined choice accuracy for effects of mode of thought and gender. As can be seen in Tables 3A and 3B, there were no significant effects involving mode of thought, with the percentage of correct choices being 58.2% and 61.9%, respectively, in the de-

Table 3: A. Percentage of participants who chose the option with the largest number of desirable properties in in the deliberation and distraction conditions, shown separately for male and female participants.

Condition	Gender	N	Choice Accuracy (% correct)
Deliberation	Male	130	50.8
	Female	269	61.7
Distraction	Male	130	55.4
	Female	269	65.1

B. Outcomes of general linear model examining effects of mode of thought (deliberation vs. distraction) and gender on choice accuracy.

Source	Wald χ^2 (df=1)	p-value
Intercept	20.56	<.001
Mode of thought	1.09	.30
Gender	8.24	<.01
Mode of thought * Gender	.02	.90

liberation and distraction conditions.[5] The sole effect to reach significance was the main effect of gender, with female participants being significantly more likely to select the best option than male participants (63% vs. 53%, respectively). Crucially, however, gender did not interact with mode of thought, thus failing to replicate the interaction effect that was found in an exploratory analysis by Nieuwenstein and Van Rijn (2012). Lastly, the analysis of the confidence ratings did not show significant effects of mode of thought or of the duration of the deliberation phase, whereas it did yield a significant effect of gender, $\chi^2(1) = 13.27$, $p < .001$, with female participants being less confident about their choice than male participants (M = 6.9 vs. M = 7.4, respectively).

4.3 Bayes factor analysis

Though the results of the GLM analysis are clear in demonstrating a lack of a statistically significant UTA, this type of analysis does not allow for a quantification of the extent to which the results support the null hypothesis over an alternative hypothesis that stipulates that the effect does exist. One approach that offers an elegant means to do so is the computation of a Bayes factor (e.g., Dienes, 2008; Dienes, 2011; Jeffreys, 1961; Morey & Rouder,

[5]Assuming an effect size d = .218, as found by Strick et al. (2011) in their meta-analysis, the power for the statistical test of this within-subjects difference between the deliberation and distraction conditions was .997.

2011; Newell & Rakow, 2011; Rouder, Speckman, Sun, Morey, & Iverson, 2009; Wagenmakers, 2007). To be precise, a Bayes factor can be used to competitively contrast two models of the data, which in this case represent the null hypothesis (H_0) that there exists no UTA effect and an alternative hypothesis (H_1), which assumes that this effect does exist. The Bayes factor is the relative likelihood of the data under these two hypotheses, and the outcome of this computation indicates the extent to which rational observers should adjust their relative beliefs in response to the data. Specifically, if the Bayes factor is greater than one, it indicates that belief should be adjusted in favor of the null hypothesis, and if it is less than one, it indicates that belief should be adjusted in favor of the alternative hypothesis.

To competitively contrast the H_0 and H_1 models, we first had to construct a model for H_1 which was intended to fairly represent the outcome a proponent of the UTA would predict for the current study. To construct the model, we used the outcomes of six experiments that were conducted by proponents of the UTA, and that were reported to show a significant UTA (Experiment 2 in Dijksterhuis [2004], Experiment 1 in Dijksterhuis et al. [2006], Experiments 1 and 2 in Nordgren, Bos, & Dijksterhuis [2011], and Experiments 1 and 2 in Strick, Dijksterhuis, & Van Baaren [2010]). The reasons for using these experiments as the basis for the H_1 model were that they were all reported to show evidence in favor of the UTA (even though not all these effects were statistically significant, see the Supplement), and because they were similar to the current study in terms of their outcome measure (proportion of correct choices). The reason why we chose to use only studies that reported proportions correct—as opposed to using all studies done by proponents of the UTA—was that this enabled us to use the same scale to model the data from our own study and from the studies we used to construct the H_1 prior.

Taken together, the six experiments used as a basis for the H_1 prior included 150 participants in the distraction condition and 172 participants in the deliberation condition and the proportions of correct choices in these conditions were .62 and .31, respectively. On the basis of these data, we developed a model for H_1 (see Supplement for details) which can be argued to reflect the prediction a proponent of the UTA would make for the current experiment, according to their own observations. Indeed, it could even be argued that our estimate of this prediction underestimates the effect a proponent of the UTA would predict for the current study, as the 6 studies used for deriving this predicted outcome did not all meet all of the requirements for a strong UTA effect, as suggested by Strick et al. (2011) in their meta-analysis. Thus, to the extent that one believes the UTA is stronger if the recommendations of Strick et al. are followed, one should also believe that

Table 4: Results for a between-subject comparison of performance in the condition that was done first by each participant in the current study.

Condition	N	Choice accuracy (% correct)
Deliberation	196	55.61
Distraction	203	55.17

Figure 2: Graphical depiction of the probability density functions for the effect size of the UTA predicted under H_1 (the prior, depicted as a dashed line), and the posterior probability density function after inclusion of the outcome of the current study (the solid line). Effect size is defined in probit units.
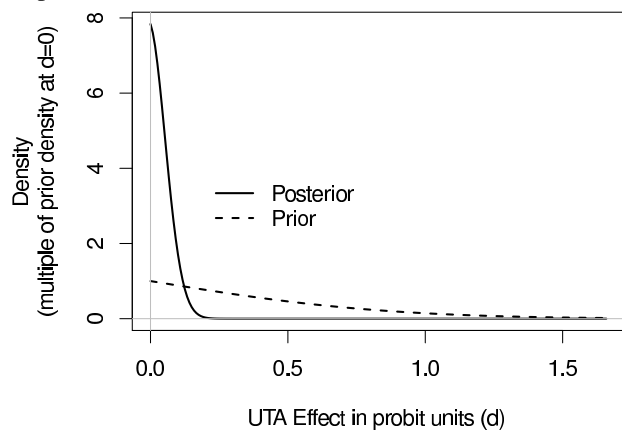
the outcome we derived as a prediction for the H_1 prior *underestimates* the magnitude of the UTA that proponents would predict for our experiment, which met all recommendations of Strick et al.

In computing the Bayes factor, we assumed that the proportions of correct choices in the deliberation and distraction conditions were binomially distributed, and the parameters of these distributions were derived from a standard probit model. By applying this probit model to the 6 previous studies showing the UTA, we derived a distribution of *a priori* expectations for the true effect size under H_1 (depicted by the dashed line in Figure 2). For the current study, we followed a similar procedure to model the results for the between-subjects comparison of the deliberation and distraction conditions, using only the outcomes for the condition that was done first by each participant[6]

(see Table 4). The Bayes factor was then computed as the extent by which the density around the null hypothesis $d = 0$ grew from the prior for H_1 to the posterior after including the data from our large-scale study. As can be seen in Figure 2, the null effect of our study caused the posterior distribution to gather around the null value $d = 0$. Specifically, the density at $d = 0$ grew by a factor of 7.83, meaning that a rational observer who considers H_1 against H_0 should adjust his belief in favor of H_0 by a factor of 7.83.[7]

5   Meta-analysis

Taken together, the results of the large-scale replication study provide compelling evidence against the moderator account, as they make clear that a high-powered study that is optimized in accordance with the purported moderators of the UTA yields no evidence for this effect. By implication, the results of the large-scale replication study may also be considered as support for the reliability account. As described in the introduction, this account not only predicts that the UTA will not be found in a large-scale study but it also predicts that previous studies that did show this effect should be confined to studies that were unreliable due to the use of small sample sizes.

To test this prediction, we examined the relationship between effect and sample sizes for a data set that included both our large-scale study and all previously published experiments that compared the accuracy of difficult choices made after distraction or deliberation. Specifically, we collected data from all published studies that used the same type of multi-attribute choice task, and the same types of deliberation and distraction conditions as Dijksterhuis and colleagues used in their seminal studies from 2004 and 2006 (see Figure 1 for a depiction of the task), and which have since then been used in dozens of replication attempts. (See Table 6 for a list of these studies and their effect and sample sizes.) Based on these data, we constructed a so-called funnel plot in which the effect sizes were plotted against a measure of study precision directly related to sample size, namely the inverse of the standard error (Egger et al., 1997; see also, Bakker et al., 2012; Light & Pillemer, 1984). Of particular relevance to the present study, this type of plot allows one to mark regions of statistical (non)significance, as the significance of a standardized mean difference score is a function of the

---

[6]The reason for using this between-subject comparison was that it was equivalent to the between-subject comparisons reported in the six experiments that formed the basis for the H_1-model. To compute the statistical power of this comparison, we used the meta-analytic effect size computed for the six studies that were used for constructing the H_1-model. Given this effect size of $d = .69$, the power for our between-subjects comparison was .999.

[7]The value of the Bayes factor would decrease if one were to assume that the effect is smaller than our estimate of the effect that would be predicted by proponents of the UTA, and it would increase if one assumes that the effect is larger. To see how the Bayes factor varies across different values of the H_1-prior, we devised an interactive applet which allows for the computation of the Bayes factor for a comparison of proportions, with different values for the H_1-prior: http://glimmer.rstudio.com/rdmorey/bfProportions.

score and its standard error. Thus, a funnel plot allows the viewer to gauge in a single glance both the distribution of significant and non-significant effects, as well as the relationship between these effects and their reliability, defined in terms of standard error. Accordingly, by inspection of the funnel plot, one can determine if previous reports of a significant UTA are indeed confined to studies that were relatively unreliable due to the use of small sample sizes, as predicted by the reliability account.

Aside from using a funnel plot to examine the relationship between effect and sample sizes, we subjected the data set to a quantitative meta-analysis in which we computed the overall effect size, and analyzed and corrected the data set for the existence of publication bias, using procedures described in detail in the following sections.

5.1 Data collection and study inclusion criteria

Studies comparing the effects of distraction and deliberation on human judgment and decision making were identified through searching the Web of Science database with "unconscious thought" and "deliberation without attention" as keywords. In addition, we checked all citations of the two seminal studies by Dijksterhuis and colleagues (Dijksterhuis, 2004; Dijksterhuis et al., 2006), and we cross-checked the studies we found against the set of studies included in the meta-analysis by Strick et al. (2011). All together, this search yielded a set of 54 published research articles that reported a total of 129 unique comparisons of the effects of distraction and deliberation on some measure of judgment or choice accuracy (see Table 5 for a general description of these studies; see the Supplement for a table listing all studies found).

As can be seen in Table 5, the majority of published studies that have compared the effects of distraction and deliberation on judgment and decision making have used a multi-attribute choice task similar to that used in the current large-scale replication attempt. Specifically, of the 54 research articles we found, 33 included one or more studies comparing the effects of distraction and deliberation on a multi-attribute choice task, and these articles together reported a total of 81 such studies (63% of all studies). In comparison, the next largest set of studies—those examining the effects of deliberation and distraction on creativity—included only 13 studies that were reported in 5 research articles. Since our main goal for the meta-analysis was to investigate the relationship between sample and effect sizes, we chose to restrict our analysis to studies using a multi-attribute choice task as these studies constituted the large majority of all studies, and because the use of the same type of task entailed that they could all be assumed to measure the same effect. Studies examining the effects of deliberation and distraction on multi-

attribute choice tasks were included in the meta-analysis if they met the following three inclusion criteria:

1. The study should include sufficient information to compute Hedges *g,* a measure of the standardized mean difference between conditions. This criterion led to the exclusion of one study.

2. The instructions given to the participants had to be similar to the instructions used by Dijksterhuis and colleagues in their seminal studies from 2004 and 2006. This meant that participants should have been instructed to form an impression of the options during the information acquisition phase (as opposed to being instructed to memorize the information about the options) and that they should have been informed prior to the distraction task that they would later be asked to judge or choose amongst the options. This criterion led to the exclusion of 7 studies that each used an instruction to memorize the information during the information acquisition phase.

3. The choice problem used in the multi-attribute choice task should have been complex, as the UTA is only predicted to occur for complex choices. The complexity of a multi-attribute choice task can be defined in terms of the number of options multiplied by the number of attributes used to describe these options. Dijksterhuis and Nordgren (2006) did not propose a criterion for when a multi-attribute choice should be considered to be complex, but the studies by Dijksterhuis and colleagues make clear that choices involving a total of 16 attributes are considered as simple, and therefore unlikely to produce the UTA, whereas choices involving a total of 30 or more attributes were predicted to yield an UTA, and may thus be considered to be complex. Accordingly, we included only studies with a total of at least 30 attributes, resulting in the exclusion of 4 studies that each used a multi-attribute choice task with four options defined by only four attributes.

5.2 Data set and effect size computation

After exclusion of the twelve multi-attribute choice studies that did not meet our inclusion criteria, we had a total 69 studies remaining in our data set. As a subsequent step, we computed composite effect sizes for 7 studies that reported two separate comparisons for two groups of participants. To be precise, we computed composite effect sizes for studies that compared a distraction and deliberation condition separately for two groups of participants that differed in having been primed to obtain a feeling of high or low power (Experiments 1 and 2 in Smith, Dijkster-

Table 5: Brief description of the types of studies found in search for studies comparing the effects of deliberation and distraction on judgment and decision making. N = number of research articles that reported studies in one or more domains, K = total number of studies within a particular domain. References and further details for all studies are provided in the Supplement.

Domain	N	K	Task description
Multi-attribute choice	33	81	Presentation of attributes of several choice options, followed by deliberation or distraction, followed by a rating or choice of the options.
Creativity	5	13	Probe for remote associates test (K = 4) or idea generation task (K = 9), followed by deliberation or distraction, followed by providing the answers to the task.
Post-choice satisfaction	5	8	Product chosen after deliberation or distraction, measurement of post-choice satisfaction 1-5 weeks after choice was made.
Moral judgment	4	7	Presentation of moral dilemma (K = 3) / a description of a job application procedure that varied in terms of fairness (K = 4), followed by deliberation or distraction, followed by judgment of what do to in the dilemma / a judgment of whether the application procedure was fair.
Lie detection	1	5	Presentation of a movie clip in which someone could be lying or telling the truth, followed by deliberation or distraction, followed by judgment of whether the person was lying or telling the truth.
Legal judgment	1	4	Presentation of legal case, followed by deliberation or distraction, followed by a judgment of whether the defendant is guilty.
Clinical diagnosis	3	3	Presentation of complex medical case followed by deliberation or distraction, followed by judgment of life expectancy or diagnosis.
Prediction	1	2	Presentation of forthcoming soccer games, followed by deliberation or distraction, followed by prediction of outcomes of the games.
Thought intrusions	2	2	Presentation of negative movie, followed by deliberation or distraction, followed by measurement of thought intrusions.
Stereotyping	1	2	Activation of stereotype, followed by presentation of behavioral descriptions of a person, followed by deliberation or distraction, followed by judgment of the person in terms of traits related or unrelated to the stereotype.
Persuasion	1	1	Presentation of persuasive message, followed by deliberation or distraction, followed by measurement of attitude towards the topic of the presentation.
Artificial grammar	1	1	Presentation of rules of artificial grammar, followed by deliberation or distraction, followed by evaluation of artificial grammar in new items.

huis, & Wigboldus, 2008), the consumption of a can of 7-Up (Bos, Dijksterhuis, & Van Baaren, 2012), low vs. high need for cognition (Experiment 2 in Lassiter, Lindberg, Gonzalez-Vallejo, Belleza, & Phillips, 2009), or featural vs. configural mindset (Experiments 2 and 3 in Lerouge, 2009). The reason for aggregating the results across these between-subjects factors was that these factors could be expected to vary naturally across participants in the other studies. Lastly, we also computed composite effect sizes for two studies in which the information about the options was presented in two different formats (numerical scores vs. colour-defined scores and numerical scores vs. star-count scores; Abadie, Villejoubert, Waroquier, & Vallée-Tourangeau, 2013a). As a result of computing these composite effect sizes, our data set was reduced to a total of 61 unique effect sizes (see Table 6 for the studies and their effect sizes). The computation of effect sizes was done using the compute.es function in R, and the meta-analysis was done using Viechtbauer's (2010) metafor package.

Table 6: Effect and sample sizes of the studies included in the meta-analysis. Note that the effect sizes derived from the study by Nieuwenstein and Van Rijn (2012) were based on the outcome of between-subjects comparisons of the condition done first in experiments that used a within-subjects design in which each participant made one or more choices after deliberation or distraction.

| Study (experiment, year)	N CT	N UT	Total N	Hedges' g	SE Hedges' g
Abadie et al. (E1, 2013a)	72	72	144	−0.37	0.19
Abadie et al. (E2, 2013a)	79	79	158	−0.62	0.20
Abadie et al. (E2, 2013b)	20	40	60	0.22	0.30
Acker (E1, 2008)	32	34	66	−0.47	0.25
Aczel et al. (E1, 2011)	24	24	48	−0.35	0.29
Ashby et al. (E1, 2011)	20	21	41	0.93	0.33
Ashby et al. (E2, 2011)	26	27	53	1.00	0.29
Ashby et al. (E3, 2011)	18	18	36	−0.21	0.34
Bos et al. (E1a, 2008)	16	16	32	1.48	0.41
Bos et al. (E1, 2012)	82	74	156	−0.10	0.16
Calvillo & Penaloza (E1, 2009)	20	20	40	−0.28	0.32
Calvillo & Penaloza (E2a, 2009)	20	20	40	−0.09	0.32
Calvillo & Penaloza (E2b, 2009)	20	20	40	−0.09	0.32
Dijksterhuis (E1, 2004)	17	22	39	0.42	0.33
Dijksterhuis (E2, 2004)	30	30	60	0.46	0.26
Dijksterhuis (E3, 2004)	46	51	97	0.24	0.20
Dijksterhuis et al. (E1, 2006)	20	20	40	0.86	0.33
Dijksterhuis et al. (E2, 2006)	15	15	30	0.70	0.38
González Vallejo et al. (E2, 2013)	42	42	84	0.00	0.25
Hasford (2014)	27	25	52	0.43	0.32
Hess et al. (E1, 2012)	81	81	162	−0.14	0.16
Huizenga et al. (E1, 2011)	30	90	120	−0.26	0.21
Huizenga et al. (E2, 2011)	37	41	78	−0.50	0.23
Huizenga et al. (E4, 2011)	25	50	75	−0.33	0.25
Lassiter et al. (E1, 2009)	21	21	42	0.51	0.32
Lassiter et al. (E2, 2009)	44	44	88	0.27	0.21
Lerouge (E1, 2009)	42	42	84	0.47	0.22
Lerouge (E2, 2009)	36	36	72	0.38	0.24
McMahon et al. (E1, 2011)	15	44	59	0.62	0.31
McMahon et al. (E2, 2011)	24	48	72	0.67	0.26
Messner et al. (E1, 2011)	20	20	40	0.63	0.33
Newell et al. (E1, 2009)	24	23	47	0.17	0.29
Newell et al. (E2, 2009)	23	23	46	−0.50	0.30
Newell et al. (E3, 2009)	30	30	60	−0.37	0.26
Newell and Rakow (E7, 2011)	20	20	40	−0.32	0.23
Newell and Rakow (E8, 2011)	32	32	64	0.09	0.25
Newell and Rakow (E9, 2011)	32	32	64	0.31	0.25
Newell and Rakow (E10, 2011)	25	25	50	−0.37	0.28
Newell and Rakow (E11, 2011)	30	15	45	−0.05	0.36
Nieuwenstein and Van Rijn (E1, 2012)	24	24	48	0.10	0.32
Nieuwenstein and Van Rijn (E2, 2012)	12	12	24	−0.55	0.45
Nieuwenstein and Van Rijn (E3, 2012)	16	16	32	0.87	0.64
Nieuwenstein and Van Rijn (E4, 2012)	12	12	24	−0.74	0.48
Nieuwenstein et al. (current study)	196	203	399	−0.01	0.10
Nordgren et al. (E1, 2011)	24	27	51	0.27	0.27
Nordgren et al. (E2, 2011)	28	27	55	0.36	0.27
Payne et al. (E1, 2008)	84	83	167	−0.10	0.16
Queen & Hess (E1, 2010)	69	68	137	−0.21	0.17
Rey et al. (E1, 2009)	36	30	66	0.27	0.25
Smith et al. (E1, 2008)	42	39	81	0.32	0.22
Smith et al. (E2, 2008)	85	80	165	0.25	0.16
Strick et al. (E1, 2010)	47	49	96	1.21	0.27
Strick et al. (E2, 2010)	31	31	62	0.58	0.29
Thorsteinson & Withrow (E1, 2009)	19	19	38	0.34	0.33
Thorsteinson & Withrow (E2, 2009)	37	37	74	0.18	0.23
Usher et al. (E1, 2011)	27	25	52	0.78	0.29
Usher et al. (E4, 2011)	14	15	29	1.04	0.40
Waroquier et al. (E1, 2009)	49	49	98	−0.56	0.21
Waroquier et al. (E2, 2009)	16	16	32	−0.09	0.36
Waroquier et al. (E3, 2009)	50	50	100	0.07	0.20
Waroquier et al. (E1, 2010)	49	49	98	0.35	0.20

5.3   Results

The 61 studies included in our data set had a sample size that ranged between 40 and 399, and their effect sizes ranged between $-.74$ and 1.48 (see Table 6). Based on these data, we constructed a funnel-plot to visualize the distribution of significant and non-significant effects, and their relationship to study precision, defined in terms of the inverse of the standard error (see Figure 3a). The white area in the plot marks the region in which effect sizes were non-significant whereas the grey areas mark the regions in which effect sizes were significant either in the direction of a conscious thought advantage (CTA; area on the left, with Hedges' $g < 0$) or an unconscious thought advantage (UTA; area on the right, with Hedges' $g > 0$). As this figure illustrates within a single glance, the published literature on the unconscious thought effect in multi-attribute choice tasks includes predominantly non-significant effects (N = 45), and only 16 statistically significant effects of which 12 were in the direction of the UTA whereas 4 were in opposite direction, that is, in the direction of an advantage for deliberation over distraction. Moreover, the plot shows a clear relationship between study precision and the finding of a significant UTA, such that the finding of a significant UTA appears to be confined to studies that had lower precision. Indeed, the studies with a relatively high precision show either a non-significant difference or an advantage for deliberation. Accordingly, it may be concluded that the observation of a statistically significant UTA appears to be confined to studies that were unreliable due to the use of small sample sizes.

As a subsequent step in our analysis we submitted the data set to a quantitative meta-analysis to compute the overall effect size. The analysis used a random effects model and yielded a pooled effect size of 0.15, with a confidence interval of [0.03; 0.26], a Z-score of 2.54, and $p = 0.01$, thus suggesting the existence of a small but statistically significant UTA. Importantly, however, the distribution of effect sizes shown in Figure 3a suggests that this effect may need correction for publication bias, as the distribution appears to be asymmetrical, with a relatively large number of low-precision UTA effects, and only few low-precision effects of equal magnitude in opposite direction. The reason why such asymmetry may hint at a publication bias is that a theoretical, completely filled-in funnel would be expected to show a symmetrical distribution of studies around the estimated true, mean effect size, such that studies of the same level of precision would be expected to be distributed symmetrically around this mean. An asymmetrical funnel lacking effects of a particular magnitude, direction, and precision is therefore often interpreted to reflect a publication bias against this type of finding (e.g., Egger et al., 1997).

Figure 3: A. A funnel-plot showing the effect sizes of studies comparing choice made after distraction and deliberation plotted as a function of the inverse of their standard error. The grey area marks the area wherein effect sizes are statistically significant at $p < .05$ and the dashed line indicates the pooled effect size, Hedges' $g = .15$.
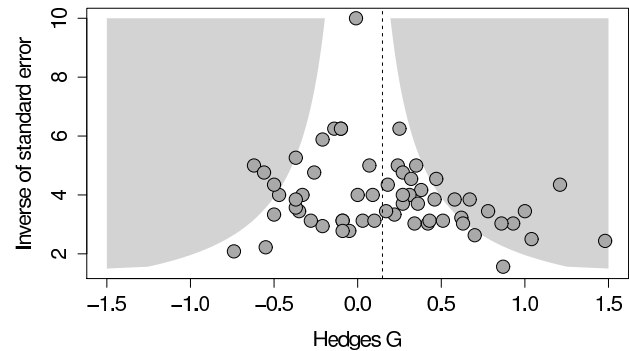
B. A funnel-plot with the same effect sizes as those shown in Figure 3a (grey symbols), with the addition of the effect sizes that were filled in using the trim and fill procedure (open symbols). The dashed line indicates the pooled effect size after inclusion of the filled-in effect sizes.
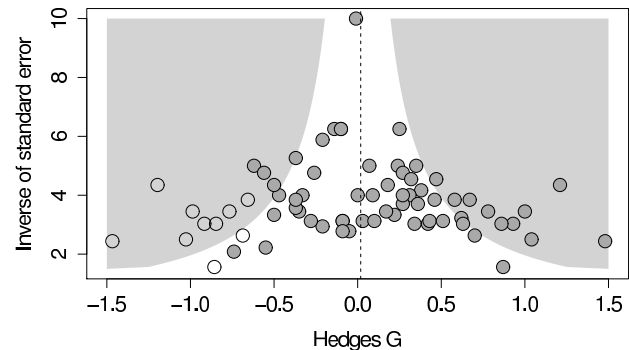
Since publication bias constitutes a common problem in meta-analyses, several procedures have been developed to deal with it. Some of these procedures focus solely on statistically significant effects, for instance by using the distribution of the p-values of these effects as a means to determine whether the distribution matches what could be expected if an effect truly existed (Simonsohn, Nelson, & Simmons, 2014; see also, Van Assen, Van Aert, & Wicherts, *in press*). Other procedures use the distribution of all effect sizes, thus offering methods compatible with the current data set, which featured predominantly non-significant effects (e.g., Duval & Tweedie, 2000; Sterne & Egger, 2005). A first such procedure that is of relevance for the current purposes regards the possibility to test whether the asymmetry in a funnel plot is statistically significant. This can be done by means of a regression analysis in which study precision is used as a predictor of effect sizes (Sterne & Egger, 2005). Using such a test, we

indeed found evidence for significant asymmetry, with Z = 2.11, and p = .04.[8]

Aside from methods to compute the statistical significance of funnel plot asymmetry, researchers have also developed methods to correct for this asymmetry. One such method is the so-called trim-and-fill procedure, which allows one to impute missing effect sizes based on the assumption that effect sizes of equal precision should be distributed symmetrically around the mean effect size[9] (Duval & Tweedie, 2000). The results of applying this procedure to the current data set are shown in Figure 3b, wherein the open symbols denote the 10 effect sizes that were filled in to correct for the asymmetry. After this correction, the overall effect size of the UTA turned nonsignificant, with a pooled Hedges' g = 0.018, a confidence interval of $[-0.10; 0.14]$, a Z-score of 0.30, and p = 0.77.[10]

6   Discussion and conclusions

With several dozen published experiments presenting conflicting results, the unconscious thought advantage (UTA) may be considered one of the most controversial phenomena in psychological science today. While proponents of the UTA have argued that the studies that failed to replicate this effect did not meet certain methodological requirements (Strick et al., 2011), critics have argued that the effect does not exist and that previous reports of the UTA concerned nothing but spurious, unreliable findings (e.g., Acker, 2008; Newell & Rakow, 2011; Nieuwenstein & Van Rijn, 2012). To adjudicate between these opposing views, we conducted a large-scale study that adhered to the conditions deemed optimal for replicating this effect

(Strick et al., 2011), and we conducted a meta-analysis that examined the relationship between the effect and sample sizes of previous studies. The results of the large-scale replication study yielded no evidence for the UTA, and it also dispelled the recent suggestion from Nieuwenstein and Van Rijn (2012) that the UTA might be gender-specific. Furthermore, the meta-analysis showed that previous reports of a statistically significant UTA were confined to studies that were relatively unreliable due to the use of small samples of participants. Accordingly, the results of the current study lead us to conclude that the claim that distraction leads to better decision making than deliberation in a multi-attribute choice task has no reliable support.

What is left to be explained then is why the paradigm shown in Figure 1 yields no difference in the quality of decisions made after distraction or deliberation. Does that mean that decision makers are just as well off if they do not think consciously about their choices (Bargh, 2011)? The answer to this question depends on whether one believes that the choices made in the unconscious thought paradigm truly reflect the outcome of two different modes of thought. On this matter, the literature on human judgment and decision-making offers a sobering perspective. Specifically, this literature includes many findings that show that people rapidly form their opinion when asked to make a judgment (e.g., Baron, 2008; Gigerenzer & Gassmaier, 2011; Kahneman, 2011). Furthermore, an abundance of findings show that once people have formed an opinion, they are unlikely to change that opinion, as they will only tend to seek further evidence to support that opinion (e.g., Bruner & Potter, 1964; Edwards & Smith, 1996; Lord, Ross, & Lepper, 1979). Accordingly, the fact that there is no difference in the accuracy of difficult choices made after distraction or deliberation is naturally explained by assuming that participants have already made up their minds during the information acquisition phase of the task and that the ensuing deliberation or distraction phase does not lead them to change their opinion (see also, Lassiter et al., 2009; Newell & Rakow, 2011). Rather, participants in the distraction condition may simply recall their earlier judgment, whereas participants in the conscious deliberation condition may only search their memory for confirmatory evidence for their earlier established preference.

A last aspect of the data that needs to be explained is why the published literature includes more studies reporting a significant UTA than studies reporting a significant benefit for conscious deliberation. In keeping with the results of our meta-analysis, this asymmetry appears to be due to a publication bias against small sample studies that found evidence for a conscious thought advantage. We can conceive of two reasons for this publication bias. The

---

[8]Some have raised concern about the use of this type of regression analysis to diagnose publication bias (e.g., Ioannidis & Trikalinos, 2007; Terrin, Schmid, Lau, & Olkin, 2003). Importantly, however, the main reasons for concern, namely that the test has low power for small data sets and that the asymmetry of the funnel might reflect true heterogeneity of effects, do not appear to apply to the current meta-analysis, as this analysis included a large set of studies that all used the same paradigm and that should therefore be expected to measure the same, or at least a very similar effect.

[9]Simonsohn et al. (2014) show that the trim-and-fill performs poorly in correcting for publication bias when this bias is based on selective reporting of significant effects, that is, when there is a publication bias against non-significant effects. Since the data set in our meta-analysis comprised predominantly non-significant effects, this concern does not apply to our analysis.

[10]The results of the meta-analysis were the same when we did not include composite effect sizes for the studies by Abadie et al. (2013), Bos et al. (2012), Lassiter et al. (2009), Smith et al. (2008), and Lerouge (2009), but instead included the effect sizes for both groups of participants compared separately in these studies. Specifically, an analysis that included these effect sizes produced a pooled effect size of 0.14, with a confidence interval of [0.02; 0.26], a Z-score of 2.32, and p = 0.02. This analysis also showed evidence for significant funnel plot assymetry, Z = 2.24, p = .02, and the effect size of .14 was reduced to a non-significant effect size of $-.01$ (*95% CI* = $[-.12; .12]$, p = .92) after application of the trim-and-fill procedure.

first is that the UTA concerns a more newsworthy finding than the finding of a conscious thought advantage, as distraction is generally thought to have a detrimental effect on task performance, and, therefore, studies reporting a beneficial effect of distraction will be considered more interesting and newsworthy than studies reporting a detrimental effect of distraction. A second reason could be that any small-sample studies—modeled after the original, small-sample studies by Dijksterhuis and colleagues (2004; 2006)—that produced an effect opposite to that of Dijksterhuis and colleagues are likely to be rejected due the use of a small sample size. This may be considered the catch-22 of the publication of a small sample study that shows a remarkable, but spurious novel effect: Once such a report is published, researchers will generally adhere to the methods of the original study in their replication attempts, and this may either lead to a coincidental replication of the same spurious effect, or to a non-replication that is much more difficult to publish because it is difficult to argue against the existence of a published effect on the basis of a small-sample study (e.g., Frick, 1995).

Aside from a publication bias, another reason for the asymmetry in available findings could be a confirmation bias on part of the researchers who believe in the existence of the UTA. This bias could take different forms as researchers who believe in a certain theory or phenomenon might engage in various questionable research practices, such as p-hacking (e.g., collecting data until the results look the way they should according to one's favorite hypothesis; Bakker et al., 2012; Ioannidis, 2005; Wagenmakers, 2007), selectively reporting one of several indices of performance (Simmons, Nelson, & Simonsohn, 2011), or running several studies to test the same hypothesis, each time under slightly different conditions, until a theory-predicted result is found (e.g., Greenwald, Pratkanis, Leippe, & Baumgardner, 1986). Of course, the risk of these practices is that they are bound to produce a predicted outcome at some point, if only by mere coincidence.

To conclude, the current study shows that previous findings suggesting the existence of an unconscious thought advantage in complex decision making concern spurious effects that were obtained with unreliable methods. Accordingly, our findings make clear that future research on the UTA should use more reliable methods, and they also make clear that the results of previous studies on this effect should be interpreted with great caution until they have been replicated in a properly powered study. Until that day, the idea that a momentary diversion of thought leads to better decision making than a period of deliberation remains an intriguing but speculative hypothesis that lacks empirical support.

References[11]

Abadie, M., Villejoubert, G., Waroquier, L., & Vallée-Tourangeau, F. (2013a). The interplay between presentation material and decision mode for complex choice preferences. *Journal of Cognitive Psychology*, *25,* 682-691.

Abadie, M., Waroquier, L., & Terrier, P. (2013b). Gist memory in the unconscious-thought effect. *Psychological Science*, *24*, 1253–1259.

Acker, F. (2008). New findings on unconscious versus conscious thought in decision making: Additional empirical data and meta-analysis. *Judgment and Decision Making*, *3*, 292–303.

Aczel, B., Lukacs, B., Komlos, J., & Aitken, M. R. F. (2011). Unconscious intuition or conscious analysis? Critical questions for the deliberation without attention paradigm. *Judgment and Decision Making, 6,* 351-358.

Ashby, N. J. S., Glöckner, A., & Dickert, S. (2011). Conscious and unconscious thought in risky choice: Testing the capacity principle and the appropriate weighting principle of unconscious thought theory. *Frontiers in Psychology*, *2*, Article 261.

Bakker, M., Van Dijk, A., & Wicherts, J. M. (2012). The rules of the game called psychological science. *Perspectives on Psychological Science*, *7*, 543–554.

Bargh, J. (2011). Unconscious thought theory and its discontents: A critique of the critiques. *Social Cognition*, *29*, 629–647.

Baron, J. (2008). *Thinking and Deciding*. Cambridge University Press, NY.

BBC News (2006). *Sleep on it, decision-makers told.* Retrieved from http://news.bbc.co.uk/go/pr/fr/-/2/hi/health/4723216.stm.

Bonke, B., Zietse, R., Norman, G., Schmidt, H. G., Bindels, R., Mamede, S., & Rikers, R. (2014). Conscious versus unconscious thinking in the medical domain: The deliberation-without-attention effect examined. *Perspectives on Medical Education*, *3*, 179–189.

Bos, M. W., & Dijksterhuis, A. (2011). Unconscious thought works bottom-up and conscious thought works top-down when forming an impression. *Social Cognition*, *29*, 727–737.

Bos, M. W., Dijksterhuis, A., & Van Baaren, R. B. (2008). On the goal-dependency of unconscious thought. *Journal of Experimental Social Psychology*, *44*, 1114–1120.

Bos, M. W., Dijksterhuis, A., & Van Baaren, R. (2012). Food for thought? Trust your unconscious when energy is low. *Journal of Neuroscience, Psychology, & Economics*, *5*, 124–130.

---

[11]This section includes references from both the text and the meta-analysis.

Bruner, J. S., & Potter, M. C. (1964). Interference in visual recognition. *Science, 144,* 424–425.

Calvillo, D. P., & Penaloza, A. (2009). Are complex decisions better left to the unconscious? Further failed replications of the deliberation-without-attention effect. *Judgment and Decision Making*, *4*, 509–517.

Creswell, J. D., Bursley, J. K., & Satpute, A. B. (2013). Neural reactivation links unconscious thought to decision-making performance. *Social Cognitive and Affective Neuroscience*, *8*, 863–869.

De Vries, M., Witteman, C. L. M., Holland, R. W., & Dijksterhuis, A. (2010). The unconscious thought effect in clinical decision making: An example in diagnosis. *Medical Decision Making*, *30*, 578–581.

Dienes, Z. (2008). *Understanding Psychology as a Science: An Introduction to Scientific and Statistical Inference.* Palgrave Mac Millan, NY.

Dienes, Z. (2011). Bayesian versus orthodox statistics: Which side are you on? *Perspectives on Psychological Science*, *6*, 274–290.

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-withoutattention effect. *Science*, *311*, 1005–1007.

Dijksterhuis, A., Bos, M. W., Van der Leij, A., & Van Baaren, R. B. (2009). Predicting soccer matches after unconscious and conscious thought as a function of expertise. *Psychological Science*, *20*, 1381–1387.

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*, 95–180.

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

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*, 455–463.

Edwards, K., & Smith, E. E. (1996). A disconfirmation bias in the evaluation of arguments. *Journal of Personality and Social Psychology*, *71*, 5–24.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple graphical test. *British Medical Journal*, *315*, 629–634.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149–1160.

Frick, R. W. (1995). Accepting the null hypothesis. *Memory & Cognition*, *23*, 132–138.

Gigerenzer, G., & Gassmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology, 62*, 451-482.

Goa, J., Zhang, C., Wang, K., & Ba, S. (2012). Understanding online purchase decision making: The effects of unconscious thought, information quality, and information quantity. *Decision Support Systems*, *53*, 772-781.

González-Vallejo, C., & Cheng, J., Phillips, N., Chimeli, J., Bellezza, F., Harman, J., Lassiter, G. D., & Lindberg, M. J. (2013). Early positive information impacts final evaluations: No deliberation-without-attention effect and a test of a dynamic judgment model. *Journal of Behavioral Decision Making*. DOI: 10.1002/bdm.1796

Greenwald, A. G., Leippe, M. R., Pratkanis, A. R., & Baumgardner, M. H. (1986). Under what conditions does theory obstruct research progress. *Psychological Review*, *93*, 216–229.

Ham, J., & Van den Bos, K. (2010a). On unconscious morality: The effects of unconscious thinking on moral decision making. *Social Cognition*, *28*, 74–83.

Ham, J., & Van den Bos, K. (2010b). The merits of unconscious processing of directly and indirectly obtained information about social justice. *Social Cognition*, *28*, 180–190.

Ham, J., & Van den Bos, K. (2011). On unconscious and conscious thought and accuracy of implicit and explicit judgments. *Social Cognition*, *29*, 648–667.

Ham, J., Van den Bos, K., & Van Doorn, E. (2009). Lady Justice thinks unconsciously: Unconscious thought can lead to more accurate justice judgments. *Social Cognition*, *27*, 509–521.

Handley, I. M., & Runnion, B. M. (2011). Evidence that unconscious thinking influences persuasion based on argument quality. *Social Cognition*, *39*, 668–682.

Hasford, J. (2014). Should I think carefully or sleep on it? Investigating the moderating role of attribute learning. *Journal of Experimental Social Psychology*, *51*, 51–55.

Hasselman, F., Crielaard, S. V. & Bosman, A. M. T. (*submitted*). Think indifferent: On the perils of scientific deliberation, without attention for critical evaluation.

Hess, T. M., Queen, T. L., & Patterson, T. R. (2012). To deliberate or not to deliberate: Interactions between age, task characteristics, and cognitive activity on decision making. *Journal of Behavioral Decision Making*, *25*, 29–40.

Hoare, R. (2012). *Got a big decision to make? Sleep on it*. Retrieved from http://edition.cnn.com/2012/08/27/business/unconscious-mind-sleep-decision

Hsu, L. M. (1989). Random sampling, randomization, and equivalence of contrasted groups in psychotherapy outcome research. *Journal of Consulting and Clinical Psychology*, *57*, 131–137.

Huizenga, H. M., Wetzels, R., Van Ravenzwaaij, D., & Wagenmakers, E. J. (2011). Four empirical tests of unconscious thought theory. *Organizational Behavior and Human Decision Processes*, *117*, 332–340.

Ioannidis, J. P. A. (2005). Why most published research findings are false. *Plos Medicine*, *2*, e124. http://dx.doi.org/10.1371/journal.pmed.0020124.

Ioannidis, J. P. A., & Trikalinos, T. A. (2007). The appropriateness of asymmetry tests for publication bias in meta-analysis: A large survey. *Canadian Medical Association Journal*, *176*, 1091–1096.

Jeffreys, H. (1961). *Theory of Probability*. Oxford University Press, Oxford, UK. Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist, 58,* 697-720.

Kahneman, D. (2011). *Thinking fast and slow.* New York: Farrar, Straus, and Giroux.

Krans, J., & Bos, M. W. (2012). To think or not to think about trauma? An experimental investigation into unconscious thought and intrusion development. *Journal of Experimental Psychopathology, 3*, 310–321.

Krans, J., Janecko, D., & Bos, M. W. (2013). Unconscious thought reduces intrusion development: A replication and extension. *Journal of Behavior Therapy and Experimental Psychiatry*, *44*, 179–185.

Krause, M. S., & Howard, K. I. (2003). What random assignment does and does not do. *Journal of Clinical Psychology*, *59*, 751–766.

Lassiter, G. D., Lindberg, M. J., Gonzalez-Vallejo, C., Belleza, 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.

Light, R. J., & Pillemer, D. B. (1984). *Summing Up: The Science of Reviewing Research.* Harvard University Press: Cambridge, MA.

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: Effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, 37, 2098-2109.

Mamede, S., Schmidt, H. G., Rikers, R. M. J. P., Custers, E. J. F. M.,Splinter, T. A. W., & Van Saase, J. L. C. M. (2010). Conscious thought beats deliberation without attention in diagnostic decision-making: At least when you are an expert. *Psychological Research-Psychologische Forschung*, *74*, 586–592.

McMahon, K., Sparrow, B., Chatman, L., & Riddle, T. (2011). Driven to distraction: Impacts of distractor type and heuristic use in unconscious and conscious decision making. *Social Cognition*, *29*, 683-698.

Mealor, A. D., & Dienes, Z. (2012). Conscious and unconscious thought in artificial grammar processing. *Consciousness and Cognition*, *21*, 865-874.

Messner, C., Wänke, M., & Weibel, C. (2011). Unconscious personnel selection. *Social Cognition*, *29*, 699–710.

Messner, C., & Wänke, M. (2011). Unconscious information processing reduces information overload and increases product satisfaction. *Journal of Consumer Psychology*, *21*, 9–13.

Morey, R. D., & Rouder, J. N. (2011). Bayes factor approaches for testing interval null hypotheses. *Psychological Methods*, *16*, 406–419.

Newell, B. R., & Rakow, T. (2011). On the morality of unconscious thought (research): Can we accept the null hypothesis? *Social Cognition*, *29*, 711–726.

Newell, B. R., Wong, K. Y., Cheung, J. C. H., & Rakow, T. (2009). Think, blink or sleep on it? The impact of modes of thought on complex decision making. *Quarterly Journal of Experimental Psychology*, *62*, 707–732.

Nieuwenstein, M. R., & Van Rijn, H. (2012). The unconscious thought advantage: Further replication failures from a search for confirmatory evidence. *Judgment and Decision Making, 7*, 779–798.

Nordgren, L. F., Bos, M. W., & Dijksterhuis, A. (2011). The best of both worlds: Integrating conscious and unconscious thought best solves complex decisions. *Journal of Experimental Social Psychology*, *47*, 509–511.

Payne, J., Samper, A., Bettman, J. R., & Luce, M. F. (2008). Boundary conditions on unconscious thought in complex decision making. *Psychological Science*, *19*, 1118–1123.

Queen, T. L., & Hess, T. M. (2010). Age differences in the effects of conscious and unconscious thought in decision making. *Psychology and Aging*, *25*, 251–261.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Reinhard, M. A., Greifeneder, R., Scharmach, M. (2013). Unconscious processes improve lie detection. *Journal of Personality and Social Psychology*, *105*, 721–739.

Rey, A., Goldstein, R. M., & Perruchet, P. (2009). Does unconscious thought improve complex decision mak-

ing? *Psychological Research*, *73*, 372–379.

Rouder, J. N., Speckman, P. L., Sun, D. C., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin, & Review*, *16*, 225–237.

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (Eds.). (2005). *Publication bias in meta-analysis. Prevention, assessment, and adjustments*. New York: WIley.

Simmons, J. P., Nelson, L. D., Simonsohn, U. (2011). False-positive psychology: flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*, 1359–1366.

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: A key to the file drawer. *Journal of Experimental Psychology: General, 143*, 534–547.

Smith, P. K., Dijksterhuis., A., & Wigboldus, D. H. J. (2008). Powerful people make good decisions even when they consciously think. *Psychological Science*, *19*, 1258–1259.

Sterne, J. A. C., & Egger, M. (2005). Regression methods to detect publication and other bias in meta-analysis. In H. R. Rothstein, A. J. Sutton & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 99-110). New York: Wiley.

Strick, M., Dijksterhuis, A., Bos, M. W, Sjoerdsma, A., & Van Baaren, R. B. (2011). A meta-analysis on unconscious thought effects. *Social Cognition*, *29,* 738-762.

Strick, M., Dijksterhuis, A., & Van Baaren, R. B. (2010). Unconscious-thought effects take place off-line, not online. *Psychological Science*, *21*, 484–488.

Terrin, N., Schmid, C. H., Lau, J., & Olkin, I. (2003). Adjusting for publication bias in the presence of heterogeneity. *Statistics in Medicine*, *22*, 2113-2126.

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.

Usher, M., Russo, Z., Weyers, M., Brauner, R., & Zakay, D. (2011). The impact of the mode of thought in complex decisions: Intuitive decisions are better. *Frontiers in Psychology*. http://dx.doi.org/10.3389/fpsyg.2011.00037.

Van Assen, M. A. L. M., Van Aert, R. C. M., & Wicherts, J. M. (*in press*). Meta-analysis using effect size distributions of only significant studies. *Psychological Methods*.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metaphor package. *Journal of Statistical Software*, *36*, 1–48.

Waroquier, L., Marchiori, D., Klein, O., & Cleeremans, A. (2009). Methodological pitfalls of the unconscious thought paradigm. *Judgment and Decision Making*, *4*, 601–610.

Waroquier, L., Marchiori, D., Klein, O., & Cleeremans, A. (2010). Is it better to think unconsciously or to trust your first impression? A reassessment of unconscious thought theory. *Social Psychological and Personality Science*, *1*, 112–118.

Wagenmakers, E. J. (2007). A practical solution to the pervasive problem of p values. *Psychonomic Bulletin, & Review*, *14*, 779–804.

Yang, H., Chattopadhyay, A., Zhang, K., & Dahl, D. W. (2012). Unconscious creativity: When can unconscious thought outperform conscious thought? *Journal of Consumer Psychology*, *22*, 573–581.

Zhong, C. B., Dijksterhuis, A., & Galinsky, A. D. (2008). The merits of unconscious thought in creativity. *Psychological Science*, *19*, 912–918.