Social Cognition, Vol. 29, No. 6, 2011, pp. 711–726

REVISING BELIEFS ABOUT THE MERIT OF UNCONSCIOUS THOUGHT: EVIDENCE IN FAVOR OF THE NULL HYPOTHESIS

Ben R. Newell
University of New South Wales

Tim Rakow
University of Essex

Claims that a period of distraction—designed to promote unconscious thought—improves decisions relative to a period of conscious deliberation are as multifarious as they are controversial. We reviewed 16 experimental studies from two labs, across a range of tasks (multi-attribute choice, creativity, moral dilemmas), only one of which found any significant advantages for unconscious thought. The results of each study were analyzed using Bayesian *t* tests. Unlike traditional significance tests, these tests allow an assessment of the evidence for the null hypothesis—in this case, no difference between conscious and unconscious thought. This is done by computing the likelihood ratio (or Bayes factor), which compares the probability of the data given the null against the probability of the data given a distribution of plausible alternate hypotheses. Almost without exception, the probability of the data given the null exceeded that for the alternate distribution. A Bayesian *t* test for the average effect size across all studies (N= 1,071) yielded a Bayes factor of 9, which can be taken as clear evidence supporting the null hypothesis; that is, a period of distraction had no noticeable improving effect on the range of decision-making tasks in our sample.

The virtues of engaging in unconscious thought have been extolled for a wide variety of complex social and cognitive tasks. A period of distraction which al-

The support of the Australian Research Council (Grant: DP0877510) awarded to the first author is gratefully acknowledged. Both authors contributed equally to this article, the order of authorship is arbitrary. We thank Jeff Rouder and Will Matthews for discussion and advice on the Bayesian analyses. We thank Selena Russo for comments on an earlier draft of the manuscript. For assistance with data collection we also thank Jeremy Cheung, Kwan Wong, Genevieve McReady, My Linh To, Heather Champion, Helen Paton, Selena Russo, Allison Bennett-Barnes, Amie Robinson, Valentina Cafaro, Aanchal Sood, Oleg Barakov, and Kelly Mannin.

Address correspondence to Ben Newell, School of Psychology, University of New South Wales, Sydney, 2052, Australia. Email: ben.newell@unsw.edu.au.

lows “cognitive and/or affective task-relevant processes [to] take place outside of consciousness awareness” (Dijksterhuis, 2004, p.586; i.e., unconscious thought) is claimed to benefit multiple-cue judgment (Dijksterhuis, 2004), multi-attribute choice (Dijksterhuis, Bos, Nordgren, & Van Baaren, 2006), creativity (Dijksterhuis & Meurs, 2006; Zhong, Dijksterhuis, & Galinsky, 2008), legal reasoning (Ham, Van den Bos, & Van Doorn, 2009), forecasting (Dijksterhuis, Bos, Van der Leij, & Van Baaren, 2009), moral decision making (Ham & Van den Bos, 2010), and medical diagnosis (de Vries, Witteman, Holland, & Dijksterhuis, 2010). The evidence to support these claims comes from experiments where participants’ data encoding and response (decision or judgment) are interpolated by either a period of conscious deliberation (“conscious thought condition”) or a period of distraction (“unconscious thought condition”).

Unconscious thought is but one variant of the intuitive processes that have often been claimed to play a guiding role in our judgments and choices (e.g., Glöckner & Witteman, 2010; Hogarth, 2010). However, claims about the abilities of unconscious thought lie at the more extreme end of the ‘intelligent unconscious’ spectrum. Not only are the merits of unconscious thought espoused for the range of tasks mentioned above (and probably more), but the process itself is claimed to harness huge computational brain power that would otherwise be left untouched by the poor (capacity limited) cousin—conscious thought (Dijksterhuis & Nordgren, 2006).

Many of these bold claims have been given the intense scrutiny they deserve. Some have found these claims lacking on theoretical grounds (e.g., Gonzalez-Vallejo, Lassiter, Bellazza, & Lindberg, 2008), or offered alternative (less startling) explanations for the apparent advantages of unconscious thought in complex tasks (e.g., Lassiter, Lindberg, Gonzalez-Vallejo, Belleza, & Phillips, 2009; Payne, Samper, Bettman, & Luce, 2008; Rey, Goldstein, & Perruchet, 2009; Srinivasan & Mukherjee, 2010). Others have challenged the claims of unconscious thought theory on empirical grounds—via failures to replicate advantages for unconscious thought (e.g., Acker, 2008; Calvillo & Penaloza, 2009; Mamede et al., 2010; Newell, Wong, Cheung, & Rakow, 2009; Thorsteinson & Withrow, 2009; Waroquier, Marchiori, Klein, & Cleeremans, 2010).

No doubt the debate will continue; the idea of a powerful unconscious has an enduring appeal, and given that intuition *appears* to be so central to much of our cognition, researchers will continue to expend their efforts to understand its role (cf. Hogarth, 2010). Indeed, we were sufficiently intrigued by this idea, and the accompanying debate, to conduct 16 experiments that compared the effectiveness of conscious and unconscious thought. We report the fruits of our labor in this article, with the aim of offering researchers (and fellow-travelers) on the quest to understand unconscious cognition a different perspective on their own data. To some, this perspective will be novel, to others familiar, but we believe that it is the first time this particular statistical tool—*Bayesian inference*—has been applied to the Unconscious Thought Theory (UTT) paradigm.

We argue that the use of Bayesian inference for evaluating the effects of unconscious thought is particularly apposite given the controversy surrounding the replicability of the basic effect (e.g., Acker, 2008; Newell et al., 2009; Waroquier et al., 2010). While it is clear that there is more than one research group or lab that *is* able to find the effect (see Strick et al., this issue), there are rather more published failures to replicate than one might be comfortable with given the bold claims and prescriptions for reliance on unconscious thought in various real-world situations (e.g., Ham et al., 2009). The appealing aspect of Bayesian statistical inference is that it provides a more coherent analytic framework than standard approaches (Dienes, 2011; Matthews, in press; Wetzels et al., 2011) and allows stronger conclusions to be drawn on the basis of *null* results. It provides us, arguably, with a more balanced view of the merits of unconscious thought by permitting principled evaluation of the theoretically interesting possibility that conscious and unconscious thought do not differ according to the quality of the decisions that they deliver.

BAYESIAN REANALYSES OF THE UNCONSCIOUS THOUGHT EFFECT

In psychological research, null results—failures to find a statistically significant effect—are generally deemed inconclusive: one is unable to draw any clear conclusion from the data. Arguably, this practice owes much to Ronald Fisher (see Hogben, 1970), one of the fathers of significance testing, for whom "every experiment may be said to exist only in order to give the facts a chance of disproving the null hypothesis" (Fisher, 1935, p. 19). Thus, the null (the hypothesis to be nullified) could be disproved, but never proved or established.[1] This position rapidly became the standard view in psychology, so much so, that by 1974 the *APA Publication Manual* equated "negative results" with failures to reject the null, and "positive results" with rejections of the null (Gigerenzer, 1993). Much has been written concerning the adverse consequences of this "prejudice against the null" (Greenwald, 1993, p. 419). These include: questionable research practices (e.g., recasting a secondary hypothesis or chance result as the main hypothesis), unfortunate dissemination practices (e.g., publication biases), incomplete interpretation of findings (e.g., ignoring effect size), and plain-and-simple misinterpretation of the *p* value (Cohen, 1990, 1994; Greenwald, 1993; Meehl, 1967, 1978; Rozeboom, 1960; Zilak & McCloskey, 2008).

However, just suppose that the null hypothesis were true (precisely, or within some degree of tolerance). For instance, imagine that, for the kinds of complex decisions explored via the UTT paradigm, there really was no difference between the conditions (i.e., conscious thought, when a period of deliberation is provided, and unconscious thought, when participants are distracted before making a decision). How would we ever learn that this were so? Presumably there would be a large accumulation of null results, and a small number of significant results in either direction.

Alternatively, suppose that the null hypothesis were true most of the time, but a genuine effect were present some of the time (e.g., in specific domains). Only after particularly extensive investigation of the effect through numerous studies would it be possible to assess whether a body of evidence was consistent with the null in each of several different domains. In either case, we cannot assess the evidence for the null in each individual study—we can only do so if we are prepared to wait for many studies to become available (and even then, the picture is approximate).

1. Fisher's later writing took a slightly softer line, whereby, although it is impossible to "establish" the null hypothesis, it coutld be said to be "confirmed" or "strengthened" by a nonsignificant result (Fisher, 1955, p. 73). However, Fisher's early view seems to have had greater influence over the use of significance tests in psychology (Gigerenzer, 1993; Gigerenzer, Swijtink, Porter, Daston, Beatty, & Krüger, 1989).

Moreover, where would we find these many studies? We know that there is some disinclination (from researchers and editors) to publish nonsignificant results (Rosenthal, 1979). Consequently, the road to assessing the evidence for the null hypothesis is a rocky one if we rely on published research reporting tests of significance. However, there are alternative approaches to statistical inference that allow a precise evaluation of the evidence for the null hypothesis in a single study. One of these is the Bayesian approach, which we illustrate here.

The recommendation to employ Bayesian statistical inference in psychology is not a new one (e.g., see Edwards, Lindman, & Savage, 1963), but the approach has attracted increased attention in recent years. For instance, Wagenmakers, Wetzels, Borsboom, and Van der Maas (2011) and Rouder and Morey (2011) reanalyzed experiments on psi or extra-sensory phenomena (Bem, 2011) using Bayesian inference, on the basis that this is exactly the kind of situation where one needs to assess the evidence for the (theoretically important) null hypothesis (reflecting the absence of an extra-sensory effect).

The Bayesian perspective treats probability as "personal" or "subjective," in other words, reflecting the strength of belief in a proposition or hypothesis (e.g., Lindley, 1971)—and this perspective is adopted in our exposition that follows.[2] However, many aspects of Bayesian theory are applicable to other definitions of probability. The basic result that underpins Bayesian inference is as follows. For a null hypothesis (H_0) and a specific alternate hypothesis (H_1), it is possible to assess the probabilities of these hypotheses given new data (D) using the equation:

$$(1) \qquad \underbrace{\frac{P(H_0|D)}{P(H_1|D)}}_{\text{Posterior odds}} = \underbrace{\frac{P(D|H_0)}{P(D|H_1)}}_{\substack{\text{Likelihood ratio} \\ \text{("Bayes factor")}}} \times \underbrace{\frac{P(H_0)}{P(H_1)}}_{\text{Prior odds}}$$

When thinking through the implications of this equation, it helps to consider a specific example of examining the difference between two means: H_0 states that the effect size is zero ($d = 0$); H_1 assumes an exact effect size in a specified direction (e.g., $d = +0.5$); and the data (D) would be the outcome of an experiment as summarized by the t value (or equivalently, by the effect size and sample size).

In equation (1), the prior odds reflect how strongly one's beliefs favor H_0 or H_1 (*before* the new data are obtained). Prior odds above 1 reflect the belief that H_0 has higher probability than H_1: for instance, prior odds of 10 (i.e., 10/1) reflect the belief that the null is ten times more probable than the alternate hypothesis. Prior odds below 1 reflect the belief that H_1 has a higher probability than H_0: for instance, prior odds of 0.1 (i.e., 1/10) reflect the belief that the alternate is ten times more probable than the null hypothesis.

The posterior odds result from revising the prior odds in light of the new data. Crucially, these odds reflect the probabilities of the hypotheses given the data—probabilities which significance testing does not generate, though people often

2. Berry (1996) provides an accessible introduction to Bayesian statistics, as does Dienes (2008, 2011) in the context of psychological research. Further explanation of, and justification for, the Bayesian t tests (that we used) can be found in Rouder et al. (2009), Rouder and Morley (2011), and Wagenmakers et al. (2011).

wish (or erroneously assume) that it does (Cohen, 1994). The likelihood ratio (also known as the Bayes factor) specifies the scale and direction of the revision to the prior odds, and reflects the probability of the data given each of the hypotheses. If the Bayes factor is above 1, the odds will revise to become favorable to the null (favoring H_0 even more strongly for prior odds > 1, or favoring H_1 less strongly for prior odds < 1). Conversely, a Bayes factor below 1 yields a revision in the opposite direction: reducing the odds, reflecting a lowering of the probability of H_0 relative to the probability of H_1.

It is a key point to appreciate that the Bayes factor tells you precisely how much, and in which direction, you should revise your beliefs in light of the data. Two experimenters may have different prior beliefs about the probabilities of H_0 or H_1, but given new data (D) both experimenters should revise their prior odds in the same direction and to the same extent, where the Bayes factor (likelihood ratio) specifies the direction and extent of this revision.

A logical extension to this approach is to consider a distribution of plausible alternate hypotheses, as opposed to a single point hypothesis (such as $d = +0.5$). It makes sense that, if there is an effect, the true effect size (d) might plausibly take one of quite a large number of possible values. If one is prepared to use familiar statistical distributions, such as the normal distribution, to represent such prior distributions for H_1, then the calculations involved in extending equation (1) to this more general case are tractable.

In this general case, the Bayes factor represents a kind of weighted average of individual Bayes factors: averaging across the Bayes factors for each individual (point) H_1 within the distribution, where most weight is given to the Bayes factors for the alternate hypotheses with the highest probability density in the prior distribution for H_1.

APPLYING BAYESIAN INFERENCE TO UNCONSCIOUS THOUGHT

This, then, is the novel approach that we took to reanalyzing our data on the deliberation-without-attention (DwA) effect (Dijksterhuis et al., 2006). For the purposes of this analysis, we operationalize the DwA effect as the effect size (d) for the mean difference between conscious and unconscious thought conditions for a continuous dependent variable in a complex cognitive task. These dependent measures reflect decision quality or some related aspect of the decision that UTT predicts will be enhanced by unconscious thought (e.g., utilitarian considerations; see Ham & Van den Bos, 2010).

Findings consistent with the proposal that unconscious thought is superior to conscious thought for such measures are denoted by a *positive* DwA effect. Therefore, a *negative* effect size indicates results where conscious thought was superior to unconscious thought. Table 1 summarizes all studies conducted in our labs that examined the DwA effect (as defined above).[3]

3. Two studies that we have conducted on unconscious thought were not included in Table 1 for the following reasons: no measure of decision quality was included in Experiment 4 of Newell et al. (2009); and a study on *non*-complex decision making was run alongside Study 10, for which UTT would *not* predict an advantage for unconscious thought.

TABLE 1. Summary of Effects and Bayesian *t* Tests (Analyses Using the Bayes Factor) for the Deliberation-without-Attention (DwA) Effect in Multi-attribute Choice (MAC), Creativity (C), and Moral Judgment (MJ) Tasks (Total *N* = 1,071)

			Sample sizes		Mean difference[b]			Bayes factor, Odds favoring H_0[c]	
Study	Domain, task, data encoding instruction	DwA Dependent variable	N_{unc}[a]	N_{con}[a]	Effect size (d)	*t* value	*p* value	H_1 Cauchy prior	H_1 Normal prior
1[d]	MAC, 4 apartments with 10 attributes presented sequentially, form impression	D(Best – Worst), generic	23	24	0.09	0.254	0.801	2.6	1.93
		D(Best – Worst), individual			0.04	0.484	0.631	2.44	1.81
2[d]	MAC, 4 apartments with 10 attributes presented simultaneously, form impression	D(Best – Worst), generic	23	46	–0.15	–0.596	0.553	2.56	1.9
		D(Best – Worst), individual			–0.07	–0.297	0.767	2.82	2.12
3[d]	MAC, 4 cars with 12 attributes presented sequentially, form impression	D(Best – Worst), generic	30	30	–0.18	–0.707	0.482	2.41	1.78
		D(Best – Worst), individual			–0.17	–0.666	0.508	2.46	1.82
4	MAC, 4 houses with 12 attributes presented sequentially, memorise information	D(Best – Worst), generice	23	49	–0.13	–0.499	0.62	2.69	2
		D(Best – Worst), individuale			–0.15	–0.604	0.548	2.57	1.91
5	MAC, 4 houses with 12 attributes presented simultaneously, memorise information	D(Best – Worst), generic	52	28	0.03	0.108	0.914	3.11	2.34
		D(Best – Worst), individual			–0.10	–0.426	0.671	2.91	2.19
6	MAC, 4 houses with 12 attributes presented simultaneously, memorise information	D(Best – Worst), generic	29	55	0.04	0.167	0.868	3.14	2.37
		D(Best – Worst), individual			0.16	0.688	0.494	2.64	1.97
7	MAC, 4 houses with 12 attributes presented simultaneously	a. D(Best – Worst), generic	20	20	–0.33	–1.029	0.31	1.75	1.28
	(7a) memorize information	a. D(Best – Worst), individual			0.14	0.443	0.661	2.35	1.74
	(7b) form an impression	b. D(Best – Worst), generic	20	20	0.03	0.082	0.935	2.51	1.87

		b. D(Best – Worst), individual			–0.10	–0.326	0.746	2.43	1.8
8	MAC, 4 houses with 12 attributes presented simultaneously, form impression	D(Best – Worst), generic	32	32	–0.13	–0.535	0.595	2.67	1.99
		D(Best – Worst), individual	(29)f	-30	0.17	0.642	0.524	2.47	1.84
9	MAC, 3 housemates with 12 attributes presented simultaneously, form impression	D(Best – Worst), generic	32	32	0.31	1.259	0.213	1.63	1.19
		D(Best – Worst), individual	-32	-30	0.14	0.563	0.576	2.61	1.94
10	MAC, 4 jobs with 11 attributes presented simultaneously, form impression	D(Best – Worst), generic	25	25	–0.37	–1.312	0.196	1.46	1.06
		D(Best – Worst), individual			–0.39	–1.363	0.179	1.39	1.01
11	MAC, 4 job candidates with 12 attributes presented simultaneously	D(Best – Worst), generic	15	30	–0.07	–0.224	0.824	2.47	1.83
12	C, creative thought task, give characteristics of X (e.g., "desk") in 90 seconds	Number of characteristics	22	22	–0.42	–1.384	0.175	1.32	0.96
		Number of categories used			–0.29	–0.973	0.336	1.86	1.36
		Creativity level per response			0.93	3.075	0.004	0.1	0.09
		Total creativity score			0.02	0.073	0.942	2.59	1.93
13	C, Remote Associates Test, produce missing associate (measure reaction time, RT)	a. Solution word RT	22	23	–0.72	–2.424	0.02	0.33	0.25
	(13a) With distracters to solution words	a. Solution-word RT-advantage			0.06	0.185	0.854	2.59	1.93
	(13b) No distracter words	b. Solution word RT	28	29	–0.04	–0.140	0.889	2.83	2.12
		b. Solution-word RT-advantage			0.31	1.185	0.241	1.69	1.23
14	C, Remote Associates Test, produce missing associate (measure reaction time, RT)	a. Solution word RT	21	21	–0.31	–1.002	0.322	1.8	1.32
	(14a) With low cognitive load	a. Solution-word RT-advantage			0.26	0.855	0.398	1.98	1.46
	(14b) With high cognitive load	b. Solution word RT	21	21	0	0.01	0.992	2.56	1.9

TABLE 1. (continued)

Study	Domain, task, data encoding instruction	DwA Dependent variable	Sample sizes N_{unc}[a]	Sample sizes N_{con}[a]	Mean difference[b] Effect size (d)	Mean difference[b] *t* value	Mean difference[b] *p* value	Bayes factor, Odds favoring H_0[c] H^1 Cauchy prior	Bayes factor, Odds favoring H_0[c] H^1 Normal prior
		b. Solution-word RT-advantage			0.09	0.302	0.764	2.48	1.84
15	MJ, manipulated mode-of-thought within-ss, multiple moral dilemmas (counterbalanced)	Utilitarian character of choice	72		0.24	1.394	0.168	2.3	1.74
16[d]	MJ, moral dilemmas, respond and state consequential/deontological considerations	a. Consequential-index	27	27	–0.43	–1.590	0.119	1.1	0.8
	(a) Trolley dilemma	a. Deontological-index			–0.27	–0.985	0.33	1.96	1.43
	(b) Fish and dam dilemma	b. Rational-index			0.35	1.286	0.205	1.52	1.11
		b. Consequential-index			0.25	0.914	0.366	2.06	1.51
		b. Deontological-index			0.19	0.702	0.487	2.33	1.73

Note. [a]All analyses collapse multiple unconscious thought conditions into a single unconscious thought condition, and similarly for multiple conscious thought conditions (where more than one condition of the relevant kind was included). [b]Positive effect and *t* value denotes direction of mean difference consistent with superior performance in the unconscious thought condition. [C]Odds above 1 favor the null (H0) over the prior distribution for the alternate hypothesis (H1), whereas odds below 1 favor H1 over H0. The Cauchy prior assumes a median magnitude of effect under H1 of |d| = 0.5, and a 30% chance that |d| exceeds 1. The normal prior assumes a median magnitude of effect under H1 of |d| = 0.34, and a 5% chance that |d| exceeds 1. To achieve this, the "Scale r on effect size" parameter is set to 0.5 on the Bayesian *t*-test calculator at the University of Missouri (http://pcl.missouri.edu/). [d]Studies 1, 2, and 3 are Experiments 1, 2, and 3 (respectively) of Newell et al. (2009). A fourth experiment from that paper is not included as it examined the influences on decisions under (un)conscious thought conditions but included no measure of decision quality. [e]For the generic measures, the "best" and "worst" option are experimenter-defined according to the number of attributes favoring an option, or via a weighted additive model using average attribute weightings. Individual measures reflect the personal values of the participant: the "best" and "worst" options are defined via weighted additive models using the participant's own attribute weightings. [f]Bracketed sample sizes indicate where *N* is reduced because a dependent variable could not be computed for all participants. Studies 1–3, 11, and 13–16 were conducted at the University of New South Wales (Sydney, Australia). Studies 4–10 and 12 were conducted at the University of Essex (Colchester, UK).

In all studies, task instructions and data encoding preceded a period of conscious or unconscious thought, after which participants made their response(s). Conscious thought was usually for a specified length of time (4 minutes was typical), though in some cases was self-paced (e.g., "think for as long as you like"). Unconscious thought conditions comprised a specified period of "cognitive busyness"—unrelated to the target task—requiring the participant to carry out one or sometimes two simultaneous, distracter tasks (e.g., anagram completion, Sudoku number puzzles, *n*-back tasks, or digit cancellation). For the dependent measures in studies of multi-attribute choice (MAC), we followed the practice used by Dijksterhuis (2004) and Dijksterhuis et al. (2006) of computing the difference in attractiveness ratings for the "best" and "worst" option. In creativity tasks, the dependent measures followed standard methods from creativity research (e.g., as per Dijksterhuis & Meurs, 2006; Zhong et al., 2008) to assess qualities such as the fluency of responses (e.g., solution generation times or frequency/variety of responses) or the (independently judged) creativity of responses. In studies of moral judgment, dependent measures reflected the utilitarian character of responses or the rational character of decision-relevant considerations (e.g., as per Ham & Van den Bos, 2010).

Our analysis follows that of Rouder, Speckman, Sun, Morey, and Iverson (2009), who give the name "Bayesian *t* test" to the analytic technique using the Bayes factor that we have described above. The analyses were conducted using the University of Missouri Web-based program for Bayesian *t* tests (http://pcl.missouri.edu/). This approach is explicitly designed to permit "rejection" or "acceptance" of the null hypothesis. For our data, either a positive DwA effect (unconscious outperforms conscious thought) or a negative DwA effect (conscious outperforms unconscious thought) in any given experiment could provide evidence against the null.

The two prior distributions for H_1 that we adopted (for separate analyses) are shown in Figure 1: a normal (or "scaled information") prior (dashed line), and a prior with a Cauchy distribution (which is a *t* distribution with 1*df*). We set the standard deviation (*SD*) of the normal prior to 0.5 (in units of *d*), and set the interquartile range (IQR) of the Cauchy prior to 0.5 (by setting its scaling factor at r = 0.5). Both priors reflect an assumption that, if there is a DwA effect, it is more likely to be "small" or "medium" in size than it is to be "large," and that (particularly in the case of the normal prior) the effect is unlikely to be "very large," This assumption was informed by our understanding that proponents of a positive DwA effect make no claim that the effect is likely to be large under laboratory conditions, though they do, of course, hold that an effect is there (e.g., Strick et al., this issue). On the other side of the fence, advocates for the superiority of conscious thought have argued that forgetting, memory interference, and the possibility of on-line decision making in the UTT paradigm may ameliorate the effect (Lassiter et al., 2009; Newell et al., 2009; Payne et al., 2008; Rey at al., 2009; Shanks, 2006). Therefore, while these "UTT-critics" assume that negative DwA effects are the norm, they too may not expect this (negative) effect to be large in laboratory experiments.

It is important to note that this assumption that a small DwA effect is more probable than a large one makes it more likely that the Bayes factor will favor the alternate than would be the case for a prior distribution that gave more density to larger effect sizes.[4] In simple terms, we are letting the null hypothesis fight it out with the alternate hypothesis, and we have set fight rules that give the alternate every opportunity to show its mettle.

Inspection of the mean difference columns of Table 1 reveals that almost all of the DwA effects obtained in our labs were small or small-to-moderate in size (Cohen, 1992), and that positive and negative effects were similarly frequent. One of the (standard) *t* tests detected a significant and positive DwA effect in Study 12 (creative thought task), and a significant negative DwA effect was found in Study 13 (remote associations task). However, in the other 14 studies (and indeed, for other dependent variables in those two studies), the evidence against the null from these standard *t* tests was inconclusive.

What of the evidence in support of the null? The Bayesian *t* tests almost always yielded Bayes factors greater than 1. In other words, in almost every study, $P(D \mid H_0)$ exceeded $P(D \mid \text{prior distribution for } H_1)$. Recall that the Bayes factor tells us how much to revise our opinion in light of the evidence. Thus, given the evidence from a typical experiment from this set, with a Bayes factor of about 2, we should double our odds in favor of H_0 over H_1. Such a revision might not be considered substantial. Indeed, Wagenmakers et al. (2011), following Jeffreys (1961), propose that Bayes factors in the range 1–3 provide "anecdotal" evidence for H_0, whereas a Bayes factor above 3 is required for "substantial" evidence in favor of H_0.

However, this is how we might interpret a single result. A Bayesian *t* test for the average effect size in Table 1 of $d = -.034$ (total $N = 1{,}071$) yields a Bayes factor of 9 for the Cauchy prior and 7 for the normal prior. The same result (Bayes factor of 9) is obtained using the technique for calculating a meta-analytic Bayes factor described in Rouder and Morey (in press).[5] The simple interpretation of this result is that the accumulated evidence of Table 1 points decisively toward revising beliefs about DwA in the direction of favoring the null hypothesis.

DISCUSSION

We introduced a method of analyzing results that has been argued to provide a more coherent approach to statistical inference (Dienes, 2011; Matthews, in press) and is better able to address the frustration of a null finding—a pattern that has plagued several researchers interested in UTT (e.g., Acker, 2008; Newell et al.,

4. The reason for this is that if a small effect is observed (e.g., $d = 0.2$) this is much more probable under the H_0 of $d = 0$ than it is under an alternate hypothesis that assumes a large effect such as $d = 0.9$ or $d = 1.5$. Therefore, if a prior gives more density to such possibilities for H_1, the Bayes factor will indicate that the data are more likely given H_0 than they are given the prior distribution for H_1—thereby indicating that the evidence favors the null over the prior distribution for H_1. We, therefore, stress that our choice of prior distribution(s) for H_1 (reflecting a "real" DwA effect) makes it harder to recruit evidence for H_0 than would be the case under standard defaults of unit *SD* for the normal prior or unit IQR for the Cauchy prior (Rouder et al., 2009).

5. We used two different methods to calculate an overall Bayes factor. The first involved effectively pooling the data to obtain an average effect (*d*), which was computed by finding the median effect of each study in Table 1, and then taking the mean of these effects. This avoided placing undue weight on those studies that had a greater number of dependent measures than the other ones. The second involved taking one *t* value per independent study, and following the methods detailed in Rouder and Morey (2011) to obtain a meta-analytic Bayes factor. This latter method could be considered more technically correct (as it takes the different sample sizes across studies into account)—though, in this instance, both methods yielded the same result.

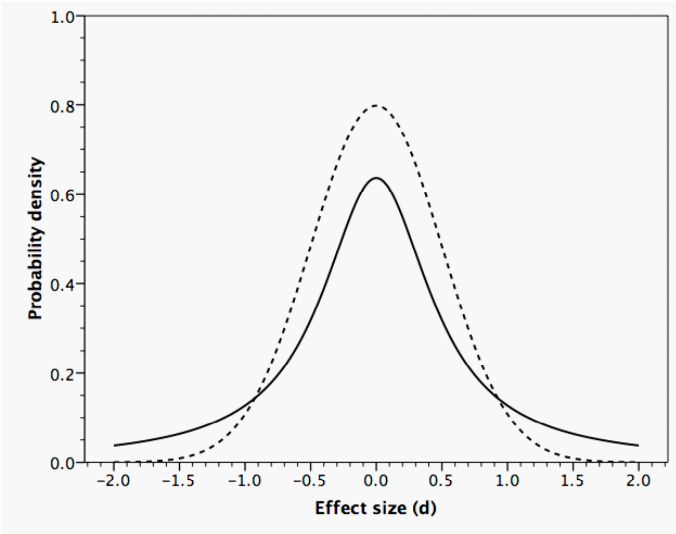

FIGURE 1. Two possible prior distributions for the alternate hypothesis, each shown for a scaling factor for the dispersion of the distribution set to r = 0.5. The normal prior (dashed line) assigns relatively high probabilities to small effect sizes, and (for this example, where r = 0.5) assumes that effects of size $|d| > 1.0$ are unlikely, and effects of size $|d| > 1.5$ are highly unlikely. In contrast, the Cauchy prior (solid line) assigns more density to larger effect sizes.

2009; Waroquier et al., 2010). Our reanalysis suggests that rather than treating such findings as inconclusive, they point in the direction of supporting the null hypothesis. Here we discuss briefly the questions and implications that follow from this conclusion.

IS THERE SOMETHING UNUSUAL ABOUT THIS SET OF EXPERIMENTS?

Some readers might be tempted to conclude that the support for the null found in these experiments was something specific to the location, participants, or materials that were used. We have two responses. First, as indicated in Table 1, these experiments come from two labs separated by approximately 17,000km, and, of course, participants, experimenters, and so forth, were all different at the two locations. Many of the studies were designed independently by the authors, and the starting point for these independently run studies were the methods described in published tests of UTT (that did find positive DwA effects). Second, ours are not the only labs that have been unable to replicate the basic effect (e.g., Acker, 2008; Calvillo & Penaloza, 2009; Mamede et al., 2010; Thorsteinson & Withrow, 2009; Waroquier et al., 2010), and we would hazard a guess (on the basis of comments received when we have discussed our data, and the well-documented reluctance of journals to publish null results) that there are others out there with file drawers similar to our own.

WAS THIS FORM OF BAYESIAN ANALYSIS SUITABLE?

The approach that we have taken is not the only one that Bayesian methods permit. One could test the null against an alternate prior distribution that only allowed for positive DwA effects (consistent with UTT), or against one that only allowed for negative DwA effects (consistent with "traditional" models of decision making). If one were to take this approach, a "half-normal" or "half-Cauchy" one-tailed distribution would be a suitable form for the alternate prior distribution (Dienes, 2011; Rouder & Morey, 2011). An advantage of this approach is that it tests the null against an alternate distribution that reflects one specific theory—thereby providing a test of that theory. However, any result in the opposite direction to that anticipated by the theory under test would constitute greater evidence for the null than for the alternate,even if the null were a poor representation of the true state of the world. Our goal was to assess the evidence for and against the null hypothesis; therefore, we were content to undertake a strong test of the null by utilizing a two-tailed alternate prior (which permit positive or negative effects to contribute evidence against the null). Inspection of Table 1 indicates that the alternative "half normal" approach would not have been more favorable to UTT, because many of the studies showed a negative DwA effect.

HOW CAN THE BAYESIAN REANALYSIS BE RECONCILED WITH A META-ANALYSIS SHOWING A RELIABLE DWA EFFECT?

This special issue includes a meta-analysis of the DwA effect (Strick et al., this issue), the final version of which was not available to us at the time of writing. We understand that it shows small but nonetheless reliable overall benefits of unconscious over conscious, and unconscious over immediate thought conditions (Dijksterhuis, personal communication). Because we are unable to comment on the specifics of the new meta-analysis, we note simply that our conclusions are in line with Acker (2008), who found in an earlier meta-analysis of 17 data sets that there was "little evidence" (p. 292) for an advantage of unconscious thought. He also found that the largest unconscious thought effects were in the studies with the smallest sample sizes. Although it is clear that the Strick et al. meta-analysis draws the opposite overall conclusion, it will be intriguing to explore the relationship between effect size and sample size in this larger set of studies.

Regardless of the take-home message from the meta-analysis, our results offer an unequivocal conclusion. As we noted in the introduction, the Bayesian approach treats probability as subjective, reflecting the strength of belief in a proposition or hypothesis. How should one's belief be revised in light of our data? For instance, perhaps you read the Strick et al. article before reading this one and that led you to have a strong prior belief that there is a genuine DwA effect. On the basis of the evidence from our labs, you should now hold to that position less strongly. To illustrate with some numbers: if your prior odds were 10:1 in favor of a DwA effect, your posterior odds (on the basis of a Bayes factor of 9) would now be 1.1:1 (i.e., favoring H_1 but at close to even odds). Alternatively, if you came to this article as a UTT skeptic with a strong leaning toward the null hypothesis (perhaps you are yet to read Strick et al.), our data should push you to believe more strongly in the null hypothesis of no DwA effect. For illustration, if your prior odds were 10:1 in favor of the null, your posterior odds would now be 90:1 in favor of the null (see equation 1).

Crucially, this prescription to favor the null more strongly applies not only to skeptics who suspect that unconscious thought is no different to conscious thought in UTT experiments, but also to those who anticipate that conscious deliberation will produce superior performance to a period of distraction. Thus, the proponents of a positive DwA effect and the proponents of a negative DwA direction should each now show more sympathy for the null hypothesis.

WHY DOESN'T "CONSCIOUS" THOUGHT IMPROVE DECISION MAKING?

The evidence for the null is clearly problematic for proponents of UTT, but it can also be viewed as inconsistent with traditional models of decision making that emphasize the importance of deliberation (Newell, Lagnado, & Shanks, 2007). The extent of the problem for such traditional views, however, is dependent on whether one views the standard UTT-paradigm as providing a sound basis for adequately assessing decision making. There are two separate issues to consider.

First, there is clear evidence that many participants make their decisions before they enter into the different modes of thought, thereby potentially rendering the manipulation peripheral to the decision-making process itself (e.g., Lassiter et al., 2009; Newell et al., 2009). Even Strick, Dijksterhuis, and Van Baaren (2010) in challenging such an explanation noted that the majority of their participants (60%) made a decision during information encoding. If the decision is relatively trivial and nonconsequential (which is true for most of those made in the UTT paradigm), then it would not be surprising if few participants reevaluated their initial decision during a period that allowed deliberation. Such a practice would lead to the null effects we have often observed (e.g., Table 1).

Second, when reliable positive DwA effects do occur, care needs to be taken to establish that these are not due to deleterious effects on conscious thought rather than superiority of unconscious thought. Many researchers (e.g., Payne et al., 2008; Shanks, 2006) have noted that the conditions for deliberative reasoning are extremely poor in most UTT experiments (e.g., randomly and briefly presented attributes, no information present during deliberation, forced and fixed amount of time to think). Payne et al. (2008) demonstrated that allowing participants to think consciously for as long they liked (rather than for a forced amount of time) led to decisions that were superior to those made following distraction. In a similar vein, Mamede et al. (2010) showed that expert doctors given a structured diagnosis-elicitation tool during the deliberation period produced more accurate diagnoses in complex cases than when they were distracted or made an immediate diagnosis. Interestingly, in the same study novice doctors made poorer diagnoses in complex cases following deliberation compared to an immediate judgment (the accuracy of deliberative and distracted diagnoses did not differ). This suggests that the period of structured deliberation is only useful if particular key facts are already part of one's knowledge base (Mamede et al., 2010).

CONCLUDING THOUGHTS

The debate over the merits of unconscious thought will no doubt continue; the ideas are intriguing and have strong intuitive appeal and anecdotal relevance, even if the methods are not ideal. As methods advance, we hope our understanding of the boundary conditions will improve. For example, Usher, Russo, Weyers, Brauner, and Zakay (2011) recently presented a set of experiments that addressed a number of the limitations raised above and still found results consistent with UTT (although Usher et al. were unable to conclude whether an *active* unconscious thought process during distraction was critical for explaining their effects). In whatever way the debate over UTT is finally resolved, our hope is that by urging researchers to adopt a Bayesian approach to analyzing data from the UTT paradigm, other researchers whose file drawers might rival our own can help us to revise our beliefs about the merits (or otherwise) of unconscious thought.

REFERENCES

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.

Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology, 100,* 407–425.

Berry, D. A. (1996). *Statistics: A Bayesian perspective.* Belmont, CA: Wadsworth/ Duxbury.

Calvillo, D. P., & Penazola, 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.

Cohen, J. (1990). Things I have learned (so far). *American Psychologist*, *45*, 1304–1312.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159.

Cohen, J. (1994). The earth is round (p < .05). *American Psychologist, 49*, 997–1003.

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*. Basingstoke, UK: Palgrave Macmillan.

Dienes, Z. (2011). Bayesian vs. 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 in decision making. *Journal of Personality and Social Psychology, 87*, 586–598.

Dijksterhuis, A., Bos, M. W., Nordgren, L. F., & Van Baaren, R. B. (2006). On making the right choice: The deliberation-without-attention effect. *Science, 311*, 1005–1007.

Dijksterhuis, A., Bos, M. W., Van der Leij, A., & Van Baaren, R. (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,* 95–109.

Edwards, W., Lindman, H., & Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological Review, 70,* 193–242.

Fisher, R. A. (1935). *The design of experiments.* Edinburgh: Oliver and Boyd.

Fisher, R. A. (1955). Statistical methods and scientific induction. *Journal of the Royal Statistical Society (B), 17,* 69–78.

Gigerenzer, G. (1993). The superego, the ego, and the id in statistical reasoning. In G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences: Methodological issues* (pp. 311–339). Hillsdale, NJ: Lawrence Erlbaum.

Gigerenzer, G., Swijtink, Z. G., Porter, T. M., Daston, L., Beatty, J., & Krüger, L. (1989). *The empire of chance: How probability changed science and everyday life.* Cambridge: Cambridge University Press.

Glöckner, A., & Witteman, C.L.M. (2010). Beyond dual-process models: A categorisation of processes underlying intuitive judgement and decision making. *Thinking & Reasoning, 16*, 1–25.

Gonzalez-Vallejo, C., Lassiter, G. D., Bellezza, F. S., & Lindberg, M. J. (2008). "Save angels perhaps": A critical examination of unconscious thought theory and the deliberation-without-attention effect. *Review of General Psychology, 12*, 282–296.

Greenwald, A. G. (1993). Consequences of prejudice against the null hypothesis. In G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences: Methodological issues* (pp. 419–448). Hillsdale, NJ: Lawrence Erlbaum.

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

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

Hogarth, R. M. (2010). Intuition: A challenge for psychological research on decision making. *Psychological Inquiry, 21*, 338–353.

Hogben, L. (1970). Significance as interpreted by the school of R. A. Fisher. In D. E. Morrison & R. E. Henkel (Eds.), *The significance test controversy: A reader* (pp. 41–56). London: Butterworths.

Jeffreys, H. (1961). *Theory of probability*. Oxford, UK: Oxford University Press.

Lassiter, G. D., Lindberg, M. J., Gonzalez-Vallejo, C., Bellezza, F. S., & Phillips, N. D. (2009). The deliberation-without-attention effect: Evidence for an artifactual interpretation. *Psychological Science, 20,* 671–675.

Lindley, D. V. (1971). *Making decisions*. London: Wiley.

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 at least when you are an expert. *Psychological Research, 74,* 586–592.

Matthews, W. J. (in press). What might judgment and decision making research be like if we took a Bayesian approach to hypothesis testing? *Judgment and Decision Making*.

Meehl, P. E. (1967). Theory testing in psychology and physics: A methodological paradox. *Philosophy of Science, 34,* 103–115.

Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Karl, Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology, 46*, 806–834.

Newell, B. R., Lagnado, D. A., & Shanks, D. R. (2007). *Straight choices: The psychology of decision making.* Hove, UK: Psychology Press.

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

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

Rey, A., Goldstein, R. M., & Perruchet, P. (2009). Does unconscious thought improve complex decision making? *Psychological Research*, *73,* 372–379.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86,* 638–641.

Rouder, J. N., & Morey, R. D. (2011). A Bayesfactor meta analysis of Bem's ESP claim. *Psychonomic Bulletin & Review, 18,* 682-689.

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

Rozeboom, W. W. (1960). The fallacy of the null-hypothesis significance test. *Psychological Bulletin, 57,* 416–428.

Shanks, D. R. (2006). Complex choices better made unconsciously? *Science, 313*, 760–761.

Srinivasan, N., & Mukherjee, S. (2010). Attribute preference and selection in multiattribute decision making: Implications for conscious and unconscious thought. *Consciousness & Cognition, 19,* 644–652.

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

Thorsteinson, T. J., & Withrow, S. (2009). Does unconscious thought outperform conscious thought on complex decisions? A further examination. *Judgement 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, 2,* 1–13.

Wagenmakers,E-J., Wetzels, R., Borsboom, D., & Van der Maas, H.L.J. (2011). Why psychologists must change the way they analyze their data: The case of psi. *Journal of Personality and Social Psychology, 100,* 432–436.

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,* 111–118.

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E.-J. (2011). Statistical evidence in experimental psychology: An empirical comparison using 855 t tests. *Perspectives on Psychological Science, 6*, 291–298.

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

Zilak, S.T., & McCloskey, D.N. (2008). *The cult of statistical significance: How the standard error costs us jobs, justice and lives*. Ann Arbor: University of Michigan Press.