Deliberation, Distraction and the Role of the Unconscious in Multiple Cue Probability Learning

by
Mike Yeomans

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Arts
in
Psychology

Waterloo, Ontario, Canada, 2009

© Mike Yeomans 2009

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Many findings in cognitive psychology suggest that many decisions and judgments rely on processes that are unconscious, that these processes can be disrupted by conscious input, leading to poor decision making. A commonly paradigm has shown that decision makers who are distracted while deciding make better to make quick decisions . The distraction is thought to facilitate spontaneous unconscious processing, called the "deliberation without attention effect", that lead to better decisions when the question is later revisited. This effect was tested in three studies on diagnostic judgments in a multiple cue probability learning paradigm. In three studies, the effect of rushed decision making, forced deliberation, and distraction were tested on probability judgments, and on choices and confidence judgments. Evidence suggested that subjects in the immediate decision were less accurate than the other conditions, in decisions and judgments, despite higher levels of confidence, but equaled performance in other conditions when given make-work tasks in between decisions, which may have primed more careful or more deliberate thinking. The results did not make any strong theoretical implications for the deliberation without attention effect.

Acknowledgments

I would like to acknowledge the many people at the University of Waterloo who have done a great deal during his time to educate, enrich, and guide, among them my advisor Derek Koehler, my thesis readers Jonathan Fugelsang and Colin McLeod, as well as Grainne Fitzsimons and Dan Smilek for their patience and support, and Rita Cherkewski and Yvonne Weppler, both of whom were pivotal for many stages of the process.

In dedication to John Yeomans

Table of Contents

List of Figures

List of Tables

Introduction

The suggestive allure of the unconscious mind has seduced philosophers and psychologists for centuries. While humans have the most active conscious mind in the animal kingdom, much of our behavior is still at the mercy of unconscious decisions and action. Rene Descartes (1637) described the most basic model of this, defining as a "reflex" behavior triggered by certain stimuli and subject to little or no conscious control. Seemingly every new branch and tradition of psychology in the past 100 years has uncovered new roles for mental capacities that can and do operate automatically, without conscious control - this trend has continued and unconscious processes have been discovered in increasingly complex behaviors (Gladwell, 2005). These cases are particularly interesting because despite influences from the unconscious, they are still experienced by consciousness, and the role of the unconscious can go undetected (Nisbett & Wilson, 1977).

Daniel Kahneman (2003) has made an analogy between decision making and perception. In that field, the Gestalt tradition is notable for its demonstrations that what is perceived in consciousness is not an exact replica of the physical world, but an interpretation of what is sensed by the nervous system. Most of the time, this interpretation is virtually indistinguishable from the real thing, but research has shown that particular stimuli (visual illusions, such as Mach bands) can expose the shortcuts the perceptual system takes, and a distorted version of reality is perceived instead. Likewise with a complex decision – though the final choosing may be under conscious control, it rests on the assumptions and approximations of numerous unconscious operations. And just as with visual illusions, certain kinds of decisions and contexts can elicit irrational behavior by exposing weaknesses in the shortcuts and algorithms of the unconscious.

Despite these weaknesses, unconscious processing is relied upon in many decisions, for different reasons. First, conscious deliberation can be slow, constrained as it is by the amount of information that can be held active in working memory at any given time (Miller, 1956; Evans, 2003). Many decisions must be made quickly, and in these cases a purely conscious approach could take into account only the most obvious information. For example, decision making in social situations requires near-instantaneous consideration of many sources of information and decision alternatives – accordingly, much of that has been found to be unconsciously driven (Bargh & Chartrand, 1999). Expert decision making often involves not only better deliberation, but also a well-trained unconscious that can quickly organize information to lighten the burden on conscious mind (Klein, 1999). Deliberation is also resource intensive, and the difference in utility between an optimal and a sub-optimal choice may be outweighed by the time, effort, and glucose spent to engage conscious processes (Baumeister, Muraven, & Tice, 1999; Masicampo & Baumeister, 2008). So even if deliberation might always improve any single decision, this may be an unrealistic long-term decision-making strategy.

Does deliberation always improve a decision? Lay theories assume that in cognitive tasks, much like in physical tasks, more effort will lead to a better outcome. Economists similarly assume a lot about the decision makers that they model, who are seen to be carefully pricing and weighing every prospect for even small decisions. Behavioral economists often consider conscious deliberation as the more rational strategy. Consider a commonly cited model, Keith Stanovich and Richard West's (2000) System 1 (unconscious) and System 2 (conscious). System 2 is cast as a manager as well as a problem solver, and is held to account for the mistakes of System 1, as a failure of "cognitive reflection" (Frederick, 2005). System 1, however, is the source of "predictable irrationality" (Ariely, 2008). Prescriptions are offered to, if not prevent unconscious decision-making altogether, then at least foolproof it with "nudges" (Thaler & Sunstein, 2008). And indeed, the heuristics and biases literature is filled with paradigms that highlight the failings of the unconscious, where a more deliberative approach would succeed (Kahneman, 2003). Research focus on unconscious biases is indeed warranted because they can affect decisions undetected. It is romantic to think that all decisions can be improved by trying harder, that a little elbow grease is sufficient to achieve rationality. That, however, would be an oversimplification.

For one thing, many common decision making errors are not ameliorated by deliberation. Others get worse. Take the phenomenon of "verbal overshadowing" – studies in several paradigms show that subjects who are forced to articulate their thoughts out loud during certain types of decisions fare worse than those who are allowed to think in silence. This has been demonstrated in a wide range of tasks, such as recognizing faces (Schooler & Engstler-Schooler, 1990), wine tasting (Melcher & Schooler, 1996), spatial mapping (Fiore & Schooler, 1998), and picking a decorative poster (Wilson et al., 1993). In all of these cases, the skill impaired by forced verbalization is typically dominated by unconscious factors, such as perceptual processing.

Verbal overshadowing goes beyond strictly unconscious processes, though. It has also been found to affect tasks that have verbal components, such as insight problem solving (Schooler, Ohlsson & Brooks, 1993) and analogical reasoning (Lane & Schooler, 2004). Here the optimal recruitment is likely some combination of unconscious and conscious contributions, but forced deliberation throws the balance off, and conscious processes dominate. It is cases like these where verbal overshadowing can be at its most insidious – the superficially verbal problem can initially encourage deliberation, which recruits more and more conscious capabilities at the expense of the unconscious, and potentially helpful unconscious strategies are drowned out in a "transfer inappropriate processing shift" (Schooler, 2002). This kind of cascade is likely to happen in the real world when some amount of deliberation is already suggested by the context.

Deliberation can change not only how information is processed, but also what information is processed. Studies show that when deliberation is engaged, information that is easy to express or quantify verbally is more likely to be considered. The availability heuristic then leads us to give more weight to such easily-processed information (Tversky & Kahneman, 1974). An example of this is the "distinction bias" (Hsee & Jiang, 2004), that occurs when subjects predict their affective reaction to several related outcomes – say, reading 10 or 25 negative words, or 10 or 25 positive words. Subjects expect that reading 10 negative words will feel a little worse than reading 10 positive words, but a little better than 25 negative words and much worse than 25 positive words. When people actually read words and rated their feelings, they indeed liked positive words more than negative words, but the number of words had no effect on either set. This mistake only occurs when all options are evaluated at once – if they are evaluated separately, then predictions fall in line with experiential data. The process of comparison overweights the importance of easy-to-quantify information (25 is 2.5 times bigger than 10) over hard-to-quantify information (how much better are positive words than negative words?). Other examples of the same phenomenon show that choices that are more easily justified are sub-optimally over-preferred, and many paradigms can trap System 2 into a bias known as "reason-based choice" (Shafir, Simonson & Tversky, 1992).

The availability bias is further exacerbated by the capacity constraints of the conscious mind. Information held in working memory will always be more easily retrieved, and this ease of retrieval is cyclically used as a cue for its relevance. Focal hypotheses are typically ascribed more support and are judged to be more likely than residual hypotheses (Koehler, White & Grondin, 2003). This can result in impossibly high frequency estimates when a prospect (e.g. "How many science majors are at Waterloo?") is unpacked into its constituent possibilities (e.g. "How many biology, physics, chemistry, and other science majors are at Waterloo?"). Focusing attention on each component prospect individually increases the combined perceived likelihood of the whole set, which can lead the total probability for all outcomes well above 100% (Tversky & Koehler, 1994). More generally, many problems that require divergent thinking, or consideration of a lot of alternatives, are not suited to the constrained scope of conscious thought, and decision makers are adversely affected by these "narrow frames" (Larrick, 2009). Deliberation, however, is selfperpetuating, and many decisions that might benefit from a step back and a broader frame suffer without.

A full account of these results leads to the conclusion that lay theory is wrong, and that conscious deliberation is not always the best strategy. Rather, conscious capabilities have strengths and weaknesses just like unconscious capabilities, and many types of problems are dealt with better in the absence of deliberation. Conscious involvement is determined by several things, often by what is ideal for a task, but also by irrelevant factors, such as the demands of the task immediately before. And just as it is possible to deliberate too little on a task that demands it (such as tax returns) so one can deliberate too much when a task does not demand deliberation (face recognition). The metacognitive decision to recruit conscious and unconscious resources for decision making is an important one, and many questions remain about how recruitment occurs, and in what situations each are more capable.

A recent series of studies by Ap Dijksterhuis (2004) has drawn considerable attention, and a similar amount of skepticism, for demonstrating a large role for unconscious processes in increasingly complex decisions. The initial experiment modeled different ways to approach an important decision - in this case, between four apartments to rent. Subjects were first shown bits of either positive or negative information about each apartment (for example, “Apartment B has in-house laundry” or “Apartment D has a long commute”) for ten seconds each. Some saw four bits of information about each house (16 total) and some saw 12 bits (48 total). Of the four apartments, one had more positive bits than the other three, and after seeing all the bits one at a time, subjects had to decide which apartment they liked the best.

At this point subjects were divided into three conditions, called “Blink”, “Think”, and “Sleep” (these names, not used in Dijksterhuis’ original study, were adapted from Newell et al., 2008). In Blink, they were given ten seconds to decide, which was meant to mimic a snap decision. In Think, they were given four minutes and told to deliberate through the whole time, which was intended to simulate a typical “think as hard as you can” strategy. In Sleep, subjects were supposed to make a snap decision, but only after four minutes of distraction – either anagrams, or the *n*-back task (Jonides et al., 1997). Subjects given only four bits about each apartment were equally likely to pick the best apartment of the four, regardless of condition. However, when subjects were given twelve bits, then the Sleep condition chose the best apartment more often than either the Blink or the Think condition.

Subjects in a follow-up study were asked just to rate each apartment on a 10-point scale, and the same pattern emerged – Sleep gave higher ratings to the best apartment than the other conditions, butonly when there were twelve bits of information per apartment. A final study showed than when subjects were asked to recall the bits of information instead of making any judgment or decision, subjects in Sleep recalled more positive bits and fewer negative bits about the best apartment, compared to the other conditions. Further work has taken this effect out of the lab and shown it in real consumer choices (Dijksterhuis et al., 2006). Following up on initial reports, one meta-analysis has been published that found a less-than-significant effect in the predicted direction (Acker, 2008). However, this analysis included only 17 experiments, and did not consider any moderating factors. A more comprehensive review of 37 studies shows a much stronger effect, and one that predictably varies depending on task specifics, such as the type of decision, and how the bits of information are presented (Strick et al., 2009). Doubts about whether Dijksterhuis' initial findings replicate have been, if not put to rest, then at least pacified.

The authors' proposed mechanism, however, has come under considerable controversy. Dijksterhuis (2004) described it as the "deliberation-without-attention effect", and thought that while the conscious was occupied with anagrams, unconscious processes slowly worked their way through the problem. Note that this is different from merely avoiding the common pitfalls of deliberation – rather, the theory is that unconscious processes are actively helping the solution along. Dijksterhuis proposed that wide association-driven networks, which can be employed by conscious thought, can also turn on spontaneously and search for problem-relevant information without entering awareness until the problem is again consciously revisited (for a neuroscience perspective on these networks, see Christoff et al., 2010; Mason et al., 2007). Thus, when a subject is finished solving anagrams and returns to the problem, much of the organization of relevant information is already done. The authors believe that the unconscious processing sorts out which information is most relevant for the problem, as evinced by subjects' disproportionately remembering the discriminating features (good features for the good apartment, and vice versa) in Sleep than in Think (2004, study 5).

Based on this mechanism, Dijksterhuis and colleagues (2006) proposed a "theory of unconscious thought" which delineated the practical differences between deliberation with, and without, attention. According to their theory, unconscious thought works bottom-up, and can handle a lot of information at once. Conscious thought, however, must rely on top-down rules and strategies to condense information to a manageable size, or else must focus on only small parts of the problem at once. Because unconscious thought can hold a lot of information at once, it is less susceptible to availability-based overweighting. As well, unconscious thought is capable of more divergent thinking, because it does not need to constrain the scope of relevant information for processing.

These principles can be used to predict situations where conscious and unconscious thought would be more appropriate for a decision. For example, if a task does not depend on enough information to overburden working memory, then conscious thought would be better able to handle it – as the informational burden increases, however, conscious thought would be less and less able to manage the relevant factors, and decision quality would be impaired accordingly (see Figure 1). As well, linear, rule-based thinking (such as financial planning) would be better left to deliberation, whereas divergent thinking (such as brainstorming) would be better solved using unconscious thought.

These predictions have garnered some further experimental support. Dijksterhuis and Meurs (2006) applied the conditions (Blink, Think, and Sleep) from earlier studies to creative problems, asking subjects to name different cities in The Netherlands starting with an "a", or to list different uses for a brick. Although the total number of answers was not different between conditions, Sleep prompted less typical, more divergent answers than

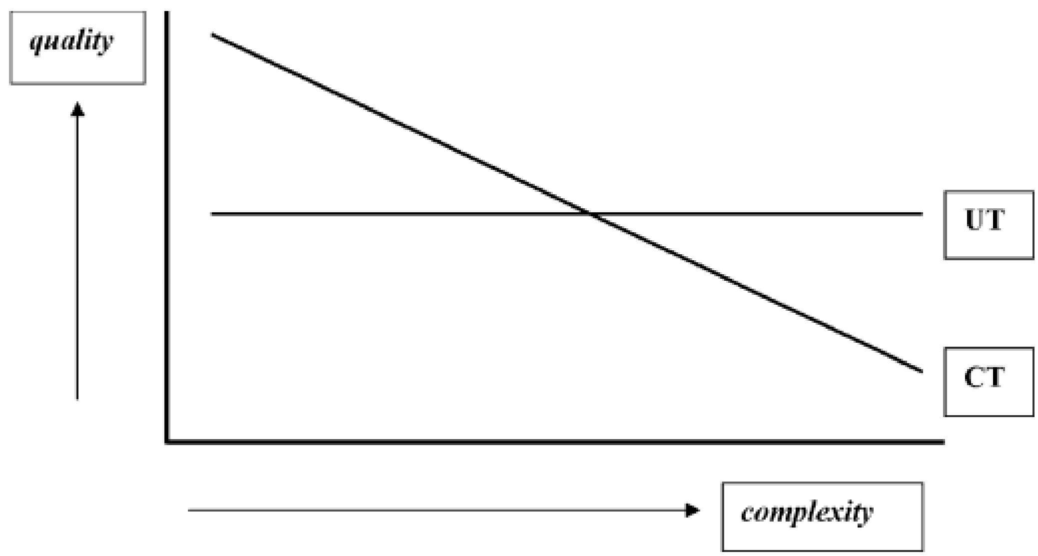

Figure 1: How complexity and quality of a decision interact in different styles of thinking (UT = unconscious thought, CT = conscious thought). Reprinted from Dijksterhuis & Nordgren (2006).

Think or Blink. Chen-Bo Zhong and colleagues (2009) attempted to reveal the workings of the unconscious networks in action. Their study used the remote associates test, which is solved by linking three seemingly unrelated words with another word commonly paired with each of the three – for example, “room”, “foot”, and “meat” would all be paired with “ball” (Bowden & Jung-Beeman, 2003). Subjects were given a set of nine problems, and had either five minutes of deliberation (Think) or five minutes of the *n*-back task (Sleep). Before solving the problems, however, subjects performed a lexical decision task, and the solutions to the remote associate problems were included in the word list. While there was no significant difference in the number of problems solved between the two conditions, subjects who did the *n*-back task responded to the remote associate solutions faster in the lexical decision task than did subjects who deliberated. In a second study using harder problems, Sleep solved fewer problems, but was just as fast as Think in recognizing the answers on the lexical decision task.

A third condition in both studies was similar to the Sleep condition – however, subjects did not know they would later have to solve the problems during the n-back task. This “Mere Distraction” condition performed worse than both Think and Sleep, suggesting that unconscious thought must be goal-directed, that the simple passage of time was not enough to improve decisions. The same effect was found in a replication of Dijksterhuis' initial four-item-by-twelve-bit experiment with Mere Distraction, where subjects who were not told that they would have to solve the problem later were no better at making decisions than were those who deliberated (Bos et al., 2008). This Mere Distraction condition was included to rule out a simple alternative account of the effect, that of "path switching" (Schooler & Melcher, 1995). This account suggests that the distraction task helps decisions not because it occupies conscious thought while the unconscious organizes relevant information, but because it allows irrelevant but activated information to drift from working memory. Vul and Pashler (2007) have shown that a period of interruption during solving time can improve performance for remote associate problems, but only when subjects are misdirected with a plausible but incorrect solution first. When subjects were not initially misdirected, the interruption was not helpful. A similar effect is found in consumer decisions - interruption can move the focus of decisions away from easy-to-quantify feasibility information, towards hard-to-quantify desirability information, which can improve post-choice satisfaction (Liu, 2009). There is a large literature on incubation that needs not draw on Dijksterhuis' unconscious networks to explain how interrupted decisions are made. However, the Mere Distraction condition could distinguish deliberation without attention from simple path switching.

Building on this, a more comprehensive alternative account was drawn from a study by Payne and colleagues (2008). The authors used a superficial modification of Dijksterhuis' initial paradigm – instead of learning 12 pieces of information about 4 apartments for rent, they learned the 12 potential outcomes of four lotteries (the underlying calculations were the same as the original). This experiment employed all three of Dijksterhuis' conditions (Blink, Think, and Sleep) for the same time periods (4 minutes; 10 seconds; 4 minutes of anagrams, respectively), and added another - "Self-Paced". In Self-

Paced, subjects were allowed to take as much time as they wanted to give an answer. The average subject spent much less than four minutes (mean=48s, median=24s). In the first study, they found that while Sleep outperformed Blink and Think, replicating Dijksterhuis, it was equaled by Self-Paced. A second study used a more complex lottery system (outcomes were unequally weighted, unlike the evenly weighted outcomes in Study 1 of Dijksterhuis, 2004) and showed that Self-Paced was better than all three original conditions.

Payne et al. took this as evidence against an active unconscious. For Self-Paced to match Sleep in less than a fifth of the distraction time undermines the necessity of unconscious processes for optimal decision making. Rather, it implies that Dijksterhuis' effect is driven by the relative under-performance of the Think condition. A case can be made that forcing subjects into four minutes of deliberation for a simple decision is an "unusual" imposition, and unrepresentative of any real-world scenario (Larrick, 2009). Dijksterhuis' conclusions implicitly assume that forced deliberation is the optimal standard to which unconscious decision making should be compared. The results of Payne et al. suggest otherwise.

This account is comparable to verbal overshadowing - in both cases, a forced and atypical over-reliance on conscious thought has a destructive influence on decision-making. A meta-analysis of verbal overshadowing in face recognition showed that that the effect was strongest when instructions for verbalization emphasized forced elaboration, rather than merely asking for a free recitation of the subject's thoughts (Meissner & Brigham, 2001). But face recognition is usually done without deliberation - in domains where subjects have experience verbalizing their thoughts, the effects are eliminated. Members of a wine tasting club showed no verbal overshadowing on wine tasting, because they had practice describing wines, and had inoculated their palates from the distorting effects of deliberation (Melcher & Schooler, 1996). A follow-up study taught subjects to identify mushroom species based on photos - some were told to describe their thoughts and theories as they learned, while others could learn in silence. When everyone was tested again later, subjects who had been forced to verbalize were no better on average than those who learned in silence, but importantly they were not affected by verbalization at test (Melcher & Schooler, 2004). The relevance of Dijksterhuis' effect is underscored by the fact that choosing an apartment is a common decision, but certainly his subjects were not practiced in deciding on apartments in four minutes, out loud, based on randomly distributed bits. While this does not discount the existence of his effect, it calls into question both the proposed mechanism and whether any of this is applicable to real-world decision making.

In light of Payne et al. (2008), the evidence for the "goal-directedness" of unconscious thought may not hold. Consider that, in the Sleep condition, a subject might stray from the distractor task and revisit the initial problem, or simply rehearse the most important information. Subjects know they have to answer the question afterward, and the median subject would have to spend only five percent of the four minute distractor task (fourteen seconds) drifting, combined with the initial ten seconds, to match the amount of time subjects voluntarily spent in Payne et al's Self-Paced condition. No task is distracting enough to make fourteen seconds of off-task thought seem unreasonable, and none of

Dijksterhuis' studies report any accounting of subjects' performance on the intermittent anagrams or *n*-back task, that might suggest they were on task. This puts Bos and colleagues' (2008) claim about the “goal-directedness of unconscious thought” in a different light. A less charitable explanation would be that subjects who know they will be tested later end up drifting more often, or for longer periods. Or they drift for the same amount of time but use that time to reconsider the problem.

To sum up, there are many unanswered questions about Dijksterhuis' Theory of Unconscious Thought. First, the evidence for “polarization” of memory traces is indirect – it has been shown as priming for lexical decision, and in memory, but not in the decision itself. This is primarily a limitation of the paradigms used - Dijksterhuis' effect has been shown only in preference decisions (picking apartments) and problem solving (remote associates), where the direction of evidence is essential, and the strength of that evidence needs not be calibrated. An inference-based judgment task, one where calibrating the intensity of support for various options is just as important as identifying the direction towards which the evidence points, may sort this out and show polarization in a more influential (and potentially harmful) role. Second, Dijksterhuis’ paradigm fails to differentiate between unconscious facilitation and conscious impedance – a task that allowed for multiple decisions that varied the degree to which these factors might influence a decision might better parse the two competing accounts. Third, the role of conscious thought during the distraction has not been quantified in any way – some attempt must be made to measure performance on the distraction task to at least get a sense of how “goal directedness” affects conscious thought, as well.

The current study attempted answers to these questions by applying the effect of Dijksterhuis' Blink, Think, and Sleep conditions to a multiple cue probability learning (MCPL) paradigm (based on White, 2006). This paradigm is an inferential diagnostic learning task – subjects see a set of case files that are representative samples of a distribution of causally related information (cues that can predict the prevalence of outcomes). From the set of case files, the underlying relation can be understood by comparing the co-occurrence of cues and outcomes. Subjects later have to apply these learned relations and diagnose the likelihood of all possible outcomes based on a set of cues.

This task meets many needs. First it satisfies many requirements for the deliberation-without-attention effect laid out by Dijksterhuis and Nordgren (2006). Each diagnosis requires distilling a large set of case files, which would benefit the relatively unconstrained capacity of the unconscious. As well, this task is driven by unconscious-favored bottom-up processing. The learning is from experience and not calculation, and the entire relational structure is learned in the task, whereas much previous information is relied upon in Dijksterhuis' original task (i.e., pre-existing knowledge about what apartment features are good or bad). Also in MCPL, the relative importance of each cue must be weighted properly, so that deliberation might cause attentional overweighting and give the unconscious a further boost.

MCPL is flexible, and many aspects of the paradigm were tweaked to encourage a bottom-up exemplar-based strategy during judgment (Juslin et al., 2003a; 2003b). This was thought to favor an unconscious cognitive style, and to provide ample opportunity for unconscious networks to outperform deliberation. For example, because the task is abstract, pre-existing knowledge will be relied upon less. A probabilistic structure was used, and the learning phase was quick, with no guessing and no feedback, which made it harder for subjects to form hard-fast rules to aid their later judgments. Further, probability judgments require understanding of both the direction and magnitude of evidence in support of one outcome, making it possible to measure polarization of the choice directly. These many judgments can also employ a diverse set of cue patterns to test a broader scope of a subjects' understanding.

The following three experiments attempted to find a deliberation without attention effect in multiple cue probability learning. All studies assigned subjects to one of three conditions, each a modified version of Dijksterhuis' Blink, Think, and Sleep conditions, and measured the effect on subjects' diagnostic judgments. All studies used the same cue structure and training set. Study 1 asked subjects for probability judgments and measured both their accuracy and perceived strength of evidence, based on the implied predictions of unconscious thought theory. Study 2 had subjects choose the most likely outcome and give confidence ratings, and measured reaction times. This was an attempt to subdivide the process of probability judgment into separate phases for evaluating strength and direction. Study 3 again used frequency judgments and had subjects in the Blink condition solve anagrams between judgments to compare against subjects in the Sleep condition and address questions about the effectiveness of the distractor.

Study 1

Methods

Subjects

Seventy-three subjects were recruited. Seven were excluded from the analysis for performance that indicated they didn't understand the task, defined as chance-or-below accuracy of implicit choices (see below for an explanation of the measure). This left 66 subjects included in the analysis.

Training Set

The MCPL structure was based on a similar one in White (2006), and was chosen because it had a broad but shallow level of complexity. It was broad because it had six cues and three outcomes, and the training set had 246 case files, meaning subjects would have to handle a lot of information at once. It was shallow because cue diagnosticity was predictable, and easy to estimate if the right evidence from the training set is recalled. The cues were conditionally independent, i.e., conditioned on a particular outcome, the probability of observing a particular cue value was independent of the value of the other cues. All cues were either high diagnostic (paired with an outcome 4.0 times as often as without) or low diagnostic (paired 1.7 times as often) with a high and low diagnostic cue associated with each of the three outcomes. It was thought that had the problem space been less broad and involved fewer cues or outcomes, it might not have overloaded the capacity of conscious thought. Further, had the problem space been deeper, with a more complex cue structure (cue dependence, or other non-linear relations), it would have been unsolvable without active hypothesis testing and rule generation, which are typically associated with conscious thought (Juslin et al., 2003b)[1]. The probabilities of each outcome given each cue are summarized in Table 1.

According to the predetermined diagnosticity levels, the probabilities of each cue given each outcome were calculated, as well as the probability of every potential combination of the six cues. Each outcome was equally likely. The probability of each potential case file (the set of six binary cue values plus one of three diagnoses) was compiled and extrapolated to a total set of 300. The occurrence of each case file was rounded to a whole number (because it is impossible to show only a fraction of a case file). As well, the 39 case files in the training set that matched those used in the multi-cue test set (which are described below) were removed from the final set. These case files all contained the same cue pattern (two high diagnostic cues, one low diagnostic cue) so low diagnostic cues are present slightly more often (131 cases each) than high diagnostic cues (114 cases each)[2]. Table 2 shows the resulting cue dependence matrix. Values indicate the ratio of the actual hypothetical likelihood assuming cue independence (which in this case means before rounding and removal of test cases). The only ratio noticeably different from 1 results from the set of missing cues – because the removed exemplars all have two high diagnostic cues and an outcome agrees with one of them, the remaining outcome that disagrees with both

1Initially the diagnosticity levels were set lower (high: 2.0 times as likely; low: 1.3 times as likely) but pilot testing indicated that this was too difficult, and a majority of subjects did not perform above chance level.
2These cases were left out due to an initial interest in whether novel test case files would be handled differently than those that were included in the training set. The results, however, were not informative, so this particular comparison was excluded from the analysis below.

	Flu1	Flu2	Flu3	Total
CueA	76	19	19	114
CueB	19	76	19	114
CueC	19	19	76	114
CueD	61	35	35	131
CueE	35	61	35	131
CueF	35	35	61	131
Total	82	82	82	246

Table 1. Frequency of each cue given each outcome in the final training set. Each outcome has two associated cues – #1 with cues A and D, #2 with cues B and E, and #3 with cues C and F. The cues are organized, with one of each pair a highly diagnostic cue (cues A, B and C; p(Flu/Cue)/p(Flu|not(Cue) = 4) and the other a low diagnostic cue (cues D, E and F; p(Flu/Cue)/p(Flu|not(Cue) = 1.74)

Cue Pairing	Flu Strain 1	2	3
AB	1.02	1.02	1.36
AC	1.02	1.36	1.02
BC	1.36	1.02	1.02
AD	0.97	0.99	0.99
BE	0.99	0.97	0.99
CF	0.99	0.99	0.97
AE	1.02	0.99	1.11
AF	1.02	1.11	0.99
BD	0.99	1.02	1.11
BF	1.11	1.02	0.99
CD	0.99	1.11	1.02
CE	1.11	0.99	1.02
DE	1.04	1.04	1.00
DF	1.04	1.00	1.04
EF	1.00	1.04	1.04

Table 2: Cue independence ratios in the training set. Values indicate deviations of flu likelihood given each cue pairing, from assumed independence. All values fall close to 1 (no deviation) indicating that rounding and case file removal would have little impact on subjects' perceived cue-outcome relationships. Note that the largest deviation, 1.36, is for the single least likely cue-flu combination, and while the ratio of the deviation is high, the absolute value is very small.

becomes slightly more likely than the training set should predict – but the overall magnitude is trivial (even with this boost the disagreeing outcome would happen 0.79 times out of 100). The rest of the comparisons the distortion imposed on the training set from the ideal distribution, due to rounding and the removal of certain case files, is minimal.. Thus the impact of these distortions on measured performance was considered negligible.

Procedure

All subjects completed the study in the lab using a program developed with the E-Prime 1.5 Suite (Schneider et al, 2008). The program first detailed a hypothetical scenario – subjects were to imagine they were doctors in a small town during an outbreak of flu, which has been shown to be caused by three separate virus strains, labeled “Russian Flu”, “Brazilian Flu”, and “Nigerian Flu”. They were told that six symptoms had been consistently observed in patients – upset stomach, dizziness, fever, headache, cough, and sore throat – and that while the presence or absence of any one symptom did not wholly determine the flu strain present, each was more likely to appear for some strains than others. It was further explained that the training set was a randomly selected series of case files, where the symptoms present at time of checkup were listed with laboratory analysis that had confirmed the flu strain. Their task, as doctors, was to use these case files to estimate the likelihood of each flu strain given information about a patients' symptoms. They were told that once they had seen the training set, they would later diagnose patients based on their symptoms, although the specific phrasing of the diagnosis (frequency estimates instead of single choices or probability estimates) was not described until later.

The design of the case file, like the design of the training set itself, was guided to maximize the potential for exemplar-based decisions later on. Each cue was presented in the same part of the screen each time. Present cues were printed in navy blue and all-caps, whereas absent cues were printed in maroon and lower case. Each flu strain was color coded – Brazilian in yellow, Nigerian in green, and Russian in red. An example case file is presented in Figure 2. Each case file was shown on-screen for six seconds, and nothing separated each case file from the next. Every subject saw the same training set in the same pseudo-random order - no significant trends were present, i.e. particular cues, outcomes, and cue-outcome pairings were not clustered towards the beginning or the end of the set.

When the 246 case files had all been shown, subjects diagnosed a test set of case files. The test set is shown in Table 3. Three types of symptom patterns were chosen: six “multi-cue” case files with three cues present, six “single cue” case files with one cue present, and a final case file with all six cues present. Each multi-cue case file had an agreeing low and high diagnostic cue paired with a contrasting high diagnostic cue, while the single cue case files had each cue present once. Further, each set of six was orthogonally balanced to equally sample knowledge for all cues. The order was constant for all subjects, and all the single cue test cases were presented after the multiple cue cases - it was thought that diagnosing a symptom with no other cues present would be simple to combine and would thus set a stronger anchor for later multiple cue judgments, than in the reverse order, where multi-cue judgments would require more deconstruction to serve as an effective anchor for single cue judgments.

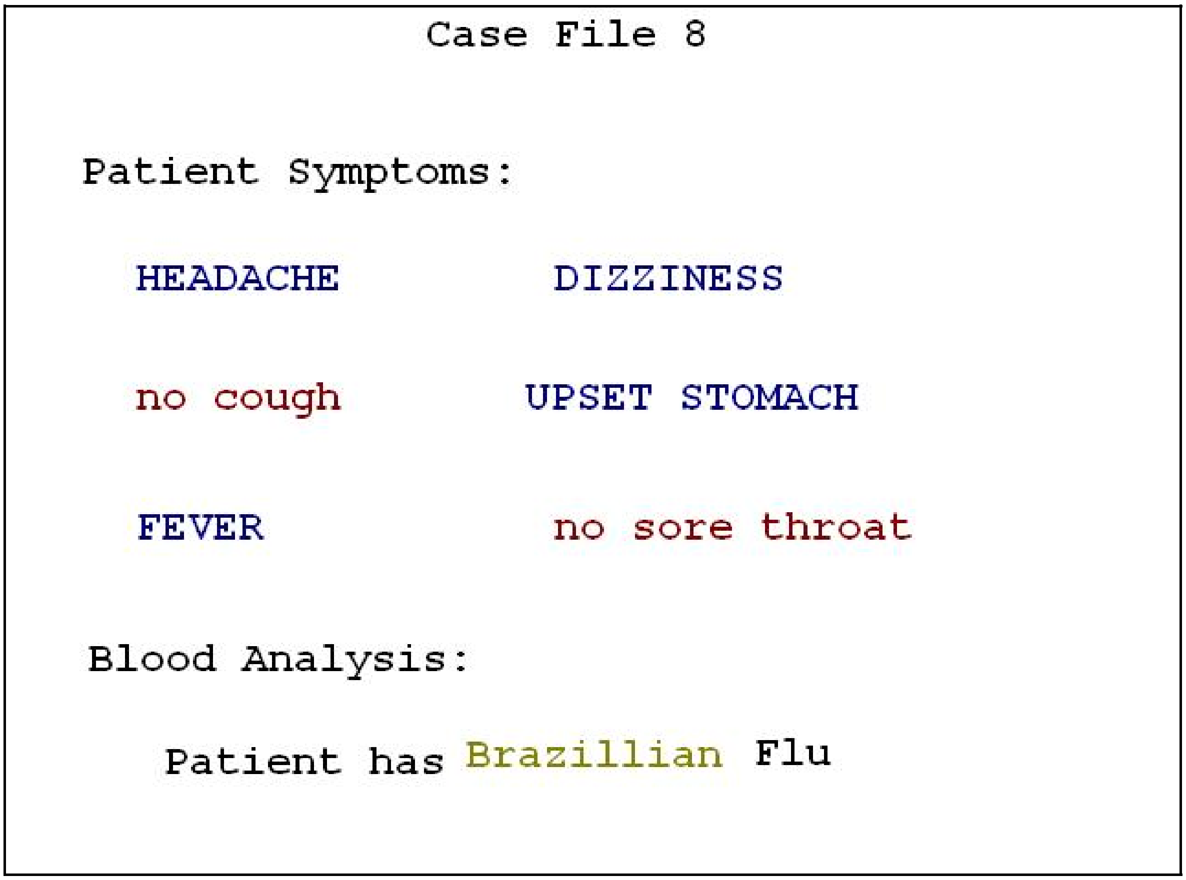

Figure 2: A screen shot example of a case file from the training set.

	Cues Present						Correct Answer		
Test Order	A	B	C	D	E	F	p (Flu1)	p (Flu2)	P (Flu3)
1		x	x		x		.8	82.7	16.5
2	x		x	x			82.7	.8	16.5
3	x		x			x	16.5	.8	82.7
4		x	x			x	.8	16.5	82.7
5	x	x		x			82.7	16.5	.8
6	x	x			x		16.5	82.7	.8
7			x				4.3	4.3	91.3
8	x						91.3	4.3	4.3
9				x			71.4	14.3	14.3
10					x		14.3	71.4	14.3
11						x	14.3	14.3	71.4
12		x					4.3	91.3	4.3
13	x	x	x	x	x	x	33.3	33.3	33.3

Table 3: Summary of test set case files. The most likely outcome, and all cues that agree with it, are in dark gray, while disagreeing cues and the associated outcomes are in light gray. The set is presented in order, and divided in two – first six multi-cue trials and then six single cue trials.

Diagnoses were elicited as frequencies - “Imagine 100 patients had this set of symptoms. How many of the 100 would have (Brazilian/Nigerian/Russian) Flu? _________”. It has been argued that frequencies are a more intuitive way of representing fractions of outcome distributions than probabilities (Gigerenzer & Hoffrage, 1995). Each of the three values was elicited sequentially, and subjects were prompted to re-enter all three if the total did not match 100. The test case files were presented in exactly the same format as the training case files, except that where previously the flu diagnosis had been given, instead the screen read “unknown” in all-caps and bright red. This was to encourage any spatially-based exemplar memory retrieval – information given visually has been shown to encourage association-based thinking (Juslin et al., 2003a; Strick et al., 2009)

The test phase was divided into three between-subject conditions. These were designed to mimic the three conditions used in Dijksterhuis' (2004) initial studies, but were adapted along three lines to accommodate MCPL. These adaptations are summarized in Table 4. First, as mentioned earlier, subjects made multiple decisions. Second, the cues were left on screen for the duration of the Think condition. It was thought that keeping the information on screen would encourage subjects to stay on task. However, Dijksterhuis had given no information during this time. That being said, remembering the cue values is not the difficult part of MCPL – the cue-outcome relations learned from the training set are more focal. The amount of assistance given by this information is small – the analogy to Dijksterhuis would be if he had written on a piece of paper “low rent apartments are better than high rent apartments” and given it to subjects in Think. Finally, the time allotted to the

	Current Study 1	Dijksterhuis (2004)
Number of judgments	Thirteen	One
Blink condition	10 s w/ cue values	10 s, no information
Think condition	100 s w/ cue values	240 s, no information
Sleep condition	10 s w/cue values, 90 s of anagrams	240 s of anagrams

Table 4: Summary of changes in decision context to adapt the manipulations of Dijksterhuis (2004) to the multiple cue probability learning task used in Study 1.

Sleep and Think conditions was shortened – instead of deliberating for 4 minutes (240 seconds), Think subjects were given 100 seconds. Sleep subjects were shown the case file for 10 seconds and then were asked to solve anagrams for 90 seconds, which balanced the total time subjects in the two conditions had between initial case file exposure and the prompt for solution.

The anagrams were presented on an otherwise blank screen, and subjects were given a text box to enter their guesses. They advanced to the next puzzle on the list only once they had solved the previous one. The anagrams were all common five- and six-letter words, and they were chosen to be particularly easy. The reason for this was two-fold – the anagrams were meant merely to distract the subjects and not to tax them to any great deal. Also, it was also thought that if subjects spent a long amount of time on any one anagram, they were more likely to drift away
from the task, and possibly re-engage the diagnosis task. Thus, is was decided that subjects would be more focused while solving a lot of easy anagrams, rather than fewer hard anagrams. The length of the block of anagrams was decided upon for practical reasons (the entire experiment had to fit into hour-long blocks for the subject pool). However, unlike the original experiment, subjects here made multiple decisions, so over time the compounded processing of information from the training set would well exceed 4 minutes. No studies to date have systematically compared how these factors interact with the deliberation-without-attention effect, so there were no *a priori* expectations about interaction.

Results

MCPL Performance

Performance in the test phase was the main dependent variable, but it can be measured in several ways. Two are used for analysis here. The first is "mean absolute deviation", or MAD. This was calculated by first tallying the absolute differences between the correct probability of a given outcome and the subject's answer. These were averaged across all three outcomes on a given test case to produce a final MAD score. This method employs a linear loss function, so that the difference between missing the correct answer by 50 and 51 is the same as the difference between missing by 2 and 3. Other measures, such as averaging within-subject correlations, or scoring deviation with different loss functions, were considered but are not reported here – there were no meaningful differences in the qualitative pattern of results based on these other scoring measures and the one used.

The other performance measure, "implied choice accuracy" made fairly basic assumptions about subjects' frequency estimates. That is, if subjects were asked not to give a frequency estimate for 100 people, but to guess the single most likely flu for a case file, it is assumed that they would choose the flu for which they had assigned the largest frequency estimate. Implied choice accuracy is scored as the proportion of case files for which this assumed choice matches the correct most likely flu. If a subject had judged two outcomes as equally frequent in a given case file population (e.g. a 50-50-0 response), and one of the two was the correct most likely flu, it was further assumed that had they been forced to pick a single flu, they would be equally likely to choose the correct one as the incorrect one. Thus, these responses were scored as half correct (0.5 out of 1).

MAD score was intended to represent both a subjects' ability to identify the valence of the case file (towards which flu the evidence pointed), and the strength of that evidence (the difference in frequency between the least and most likely outcomes). Implied choice accuracy, on the other hand, was intended to represent subjects' ability to identify the valence of the evidence while ignoring the strength of that evidence. These different measures could potentially be used to test Dijksterhuis' proposed mechanism, memory polarization. If a more polarized memory of the training set were recalled during a judgment, this should be reflected in the strength of evidence in favor of a given hypothesis, but not necessarily the direction of the evidence. So by parsing these two components of probability judgment, polarization might be measured in addition to (and perhaps, in the absence of) a general benefit in accuracy. Attempts to capture the strength alone are detailed below.

Mean Absolute Deviation

The average MAD across all thirteen test cases is plotted in Figure 3. The results were statistically compared using between condition unpaired Student's T tests, which showed a significant effect of condition – Blink had a higher MAD score (m=18.81; MSE=1.33), and thus poorer accuracy, than Think (m=14.15; MSE=1.31; $t(48)=2.42$, $p=0.015$, two-tailed) or Sleep (m=15.04; MSE=1.63; $t(38)=1.81$, $p=0.078$, two-tailed). This pattern was consistent across both types of case files – the trend of effects by condition was similar in both single cue test cases (Blink: m=17.07, MSE=1.39; Think: m=12.7, MSE=1.36; Sleep: m=14.24, MSE=1.70) and multi cue test cases (Blink: m=21.03,

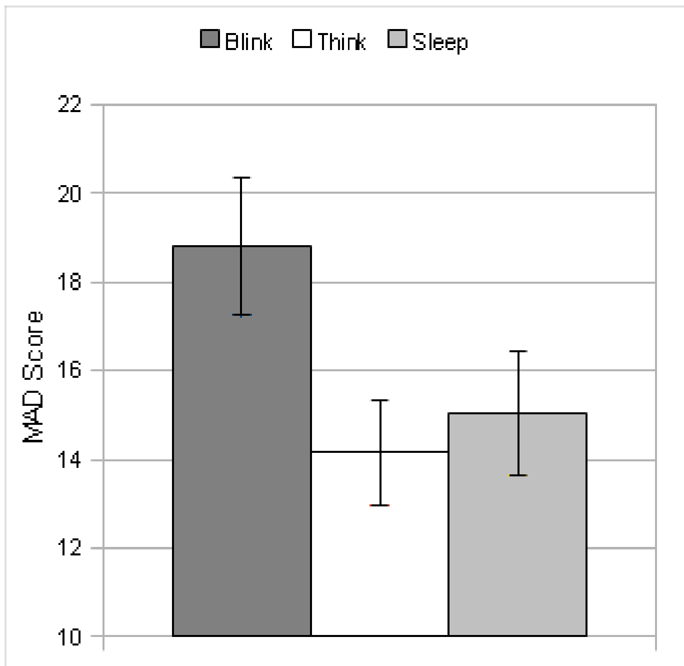

Figure 3: Mean absolute deviation of judgments in Study 1. Blink had a higher MAD score, and was thus less accurate (m=18.81; MSE=1.33) than Think (m=14.15; MSE=1.31) or Sleep (m=15.04; MSE=1.63)

MSE=1.78; Think: m=16.80, MSE=1.74; Sleep: m=16.64, MSE=2.18), and no statistical tests of interaction effects approached significance.

Implied Choice

The proportion of correct implied choices across conditions is shown in Figure 4. The results were again analyzed using between-condition t-tests, and similar to MAD scores, Blink (m=0.724; MSE=0.037) was marginally less likely to answer test cases correctly than Think (m=0.81; MSE=0.036; $t(48)=1.71$, $p=0.093$, two-tailed) or Sleep (m=0.83; MSE=0.045; $t(38)=1.81$, $p=0.078$, two-tailed). The effect of condition on accuracy of implied choices did not vary across type of test case, with similar differences in single cue test cases (Blink: m=0.77, MSE=0.039; Think: m=0.86, MSE=0.039; Sleep: m=0.84, MSE=0.048) and multiple cue test cases (Blink: m=0.68, MSE=0.048; Think: m=0.77, MSE=0.047; Sleep: m=0.81, MSE=0.059).

Extremity

Apart from comparing subjects' answers to the correct Bayesian probabilities for each test case, across-case patterns were analyzed to get a sense of consistent judgment errors or tendencies. Specifically, interest was directed towards the degree of diagnosticity that subjects attributed to case files, that is, how much the predicted probability distribution differed from the base rate (which was equal frequency for all three flu strains). Because this analysis attempted to measure the strength of evidence, not the direction, only responses from subjects who knew

which flu was most likely for a particular case (that is, were correct in the implied choice)

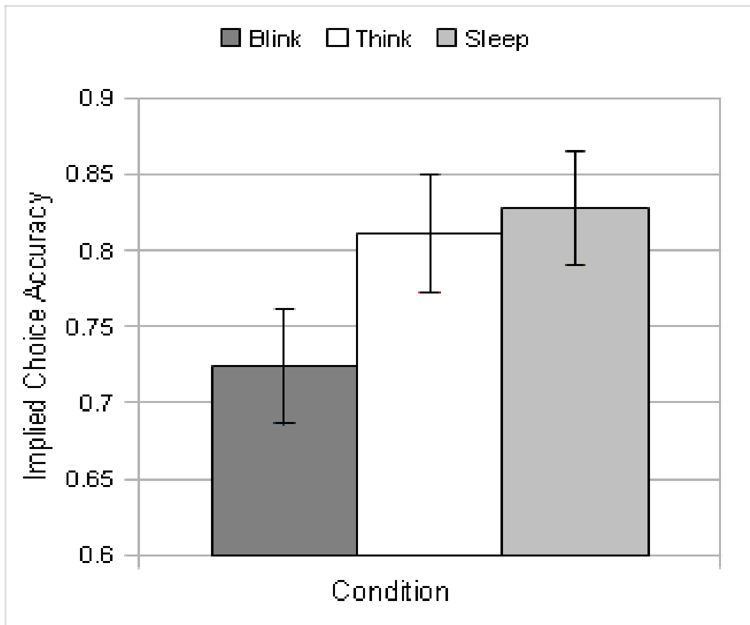

Figure 4: Accuracy of implied choices in Study 1, measuring subjects' ability to determine towards which outcome the evidence pointed. Blink (m=0.72; MSE=0.04) was worse at this than Think (m=0.81; MSE=0.04) and Sleep (m=0.83; MSE=0.04).

were included. Extremity scores were calculated by subtracting the correct likelihood from the judged likelihood of the most likely flu. This was done to take into account the fact that some case files were more diagnostic than others, and had different correct likelihoods A score of zero would indicate exact identification of the strength of the evidence towards the focal hypothesis. Positive numbers indicate overextremity of judgments, and negative numbers underextremity.

The analysis of extremity in Study 1 was suggestive. There were no significant differences between conditions on extremity averaged across all test case types (Blink: m=-7.19, MSE=2.40; Think: m=-5.46, MSE=2.38; Sleep: m=-0.74, MSE=3.31; all *p*s >0.12). Because what was considered over- and under-extreme varied with each test case type, the extremity of judgments was analyzed by test case type, of which there were three – six multiple cue cases; three single high-diagnosticity cue cases; and three single low-diagnosticity cue cases. These comparisons are shown in Figure 5. No significant differences were found between conditions in single, high-diagnosticity cue case files (Blink: m=-5.37, MSE=2.77; Think: m=-1.43, MSE=2.38; Sleep: m=-0.74, MSE=3.31, all *p*s >0.3), or in multiple cue case files (Blink: m=-12.97, MSE=2.82; Think: m=-9.36, MSE=2.46; Sleep: m=-6.65, MSE=3.72; all *p*s >0.18); however on single, low-diagnosticity cue test cases, Sleep made more extreme judgments (m=12.08, MSE=3.99) than both Blink (m=0.9 MSE=3.38; t(38)=2.06, *p*=0.046, two-tailed) and Think (m=-2.89, MSE=3.99; t(40)=2.57, *p*=0.014, two-tailed).

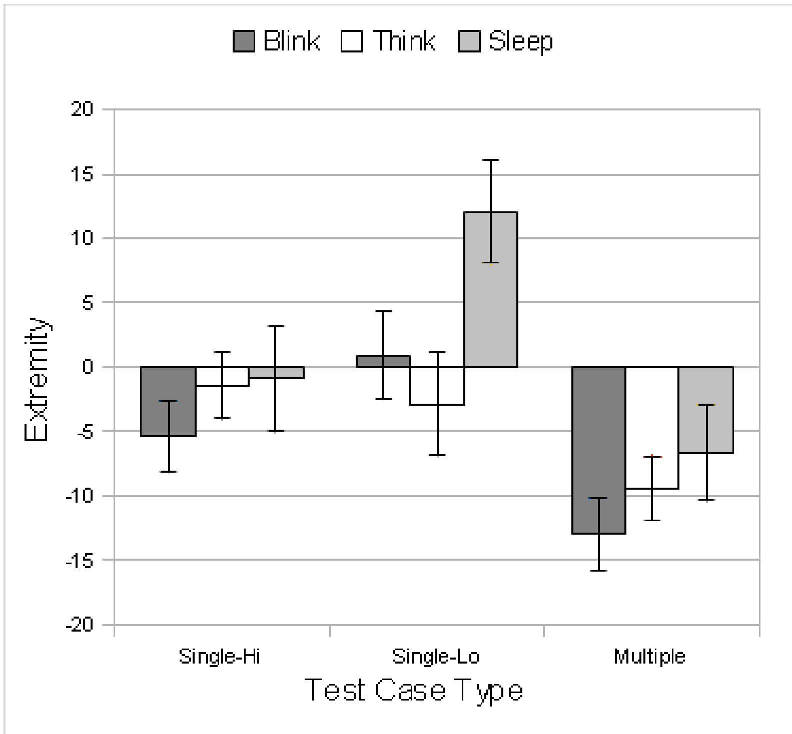

Figure 5: Between-condition comparisons of judgment extremity in Study 1. Only test cases on which the implied choice would be accurate were included. No significant differences were found in high-diagnosticity single cue case files (Blink: m=-5.37, MSE=2.77; Think: m=-1.43, MSE=2.38; Sleep: m=-0.74, MSE=3.31), or multiple cue case files (Blink: m=-12.97, MSE=2.82; Think: m=-9.36, MSE=2.46; Sleep: m=-6.65, MSE=3.72), however on low-diagnosticity single cue test cases, Sleep made more extreme judgments (m=-0.94, MSE=4.06) than both Blink (m=0.9 MSE=3.38) and Think (m=-2.89, MSE=3.99).

Discussion

The results of Study 1 do not suggest that multiple cue probability learning benefits from a deliberation-without-attention effect. A 90-second distraction did not foster better judgments than a similar period of deliberation. However, both were better than judgments made after only ten seconds of deliberation. These results were consistent both for the simple criterion of identifying the most likely outcome, and for calibrating judgments based on the diagnosticity of the cue pattern. This was contrary to the prediction, based on Dijksterhuis' theory and experiments, that subjects who were distracted would be more accurate than those who we given that time to deliberate.

While accuracy was similar between Think and Sleep, the pattern of judgment was different – subjects who were distracted produced probability estimates that were more extreme than those who deliberated, either for a short or a long amount of time. While this was predominantly found in low-diagnostic single cue case files, this was reasoned to be the result of a ceiling effect. There was little margin for subjects to overestimate the extremity of the high-diagnosticity single cue case files - p(most likely flu) = 0.91 – or the multiple cue case files – p(most likely flu) = 0.82. This finding fit the predictions based on Dijksterhuis' theory of unconscious thought - extremity of probability judgments might be the result of the same unconscious processes that polarize the memories of Dijksterhuis' subjects in Sleep. Subjects might often try to figure out the most likely flu first – once one is identified, supporting information might be given more weight than contradictory information. Thus, the roots of extremity might lie in a process secondary to identifying in which direction the evidence points. Note that Dijksterhuis' paradigm had no conception of the "correct" amount of polarization – the magnitude of difference in preference between the apartments was of no concern, merely that the best ones were chosen more often than the others. As such, polarization may facilitate choices, but not necessarily help with probability judgments. This may explain the presence of polarization but absence of better decision-making in Study 1.

To square these results with a deliberation-without-attention effect, consider the differences between this task and Dijksterhuis' original task. Probability judgment requires a deeper level of analysis than choosing a preferred apartment, or rating apartments on a scale from 1 to 9 on how they are liked. In these preference-based tasks, the relation between the cue (apartment feature) and outcome (amount of liking) is straightforward – everyone knows that a shorter commute or lower rent is better. However, in the current task memory of the cues is not enough, and the weighing of the evidence is a crucial step in producing the final judgments. This extra step may engage conscious thought on relatively friendly ground, and the interference of attentional overweighting may apply to all three conditions, since subjects in each must apply some conscious direction towards producing final point estimates of probability on each trial. Thus the task may not, allow for much unconscious influence no matter what condition.

Study 2 attempted to address these questions by modifying the MCPL paradigm to separate this two-stage process into separate decisions, instead of deriving both extremity and accuracy from a single set of judgments. First, subjects would make choices instead of probability judgments and then, as a measure of the strength of evidence, would give confidence judgments. Both of these were elicited in a non-numerical way, so that Sleep and Blink would command as little analytical thinking as possible. It was hoped that this would allow deliberation without attention to improve choices unimpeded. An added advantage of choice elicitation was that responses could be given with a single keystroke, and did not require any math (summing three numbers to 100), so that reaction times could be measured and analyzed as a meaningful representation of the amount of time required for a decision, and not merely an artifact of the process of entering an answer.

Study 2

Methods

Subjects

67 subjects were recruited through the University of Waterloo Research Experiences Group. Subjects were compensated with course credit. 8 subjects were excluded from the analysis for exceptionally poor performance, defined as chance-or-below accuracy of implicit choices(as in study 1) so that 59 subjects remained.

Materials

Study 2 was similar to Study 1, with some alterations. The underlying cue structure and training set were the same as in Study 1. However, changes were made to the test phase of the paradigm. The set of test cases was almost identical, with the only change being that the final case, where all cues are present, was removed. The remaining twelve were presented in the same order as in Study 1. A minor change was implemented in the Sleep condition – during the distraction phase, subjects were allowed to skip an anagram if, after ten seconds, they did not produce a solution. This was intended to keep subjects engaged, even if they got stuck on a particularly difficult anagram.

In Study 2 subjects made choices instead of frequency estimates. They were asked to pick the single flu that was most likely given a case file. The amount of time allowed for each choice was the same as in Study 1: subjects in the “blink” condition were presented each case file for 10 seconds before the decision screen came on; Subjects in the “think” condition were given each case file for 100 seconds before they were able to enter their decision; and subjects in the “blink” condition saw each case file for 10 seconds and then solved anagrams for 90 seconds before the decision screen. The decision screen was modified slightly from Study 1 – because subjects only had to choose one of the three outcomes, and not enter any numbers, responses were entered by pressing individual keys - “r” for Russian flu, “n” for Nigerian flu, or “b” for Brazilian flu.

In addition to choosing the most likely flu for each case file, subjects were also asked for confidence ratings. Immediately after each choice, subjects were asked to indicate their confidence in the previous choice. Confidence was mapped onto a non-numerical scale based on the middle row of the keyboard, where a response of “a” was labeled “a total guess” and a response of “l” was labeled “absolutely sure”. The non-numerical scale was used to make confidence judgments as intuitive and quick as possible. There was concern that any extra time subjects spent contemplating their knowledge outside of the window in which they actually deliberated their choices might dilute the effects of the manipulation. As well, the confidence measure was supposed to represent momentary post-choice confidence, and not be tainted by overall task confidence, or personal feelings of competence. That being said, no explicit time pressure was placed on confidence judgments. A summary of relevant comparisons to Dijksterhuis’ initial study can be found in Table 5.

Finally, reaction times for both the initial answer and the confidence judgment were recorded, unlike in Study 1. This was to address the concern that subjects in the Blink or Sleep conditions were using as much time as the Think conditions, which we could not have controlled for in Study 1 because of the nature of the response.

	Current Study 2	Dijksterhuis (2004)
Number of judgments	Twelve	One
Blink condition	10 s w/ cue values	10 s, no information
Think condition	100 s w/ cue values	240 s, no information
Sleep condition	10 s w/cue values, 90 s of anagrams	240 s of anagrams

Table 5: Summary of changes in decision context to adapt the manipulations of Dijksterhuis (2004) to the multiple cue probability learning task used in Study 2.

Results

Choice

Accuracy was the fraction of the 12 test cases for which a subject correctly identified the most likely flu given the set of symptoms. Analysis applied the Student's T-test to the between condition pairwise comparisons and found no significant differences between conditions (Blink: m=0.74; MSE=0.04; Think: m=0.75; MSE=0.07; Sleep: m=0.72; MSE=0.05; all ps >0.1). Choice accuracy was broken down by case type – again no significant differences were found between conditions in accuracy on either the multi-Cue (Blink: m=0.68; MSE=0.05; Think: m=0.70; MSE=0.08; Sleep: m=0.69; MSE=0.05; all ps >0.1), the Single-High (Blink: m=0.92; MSE=0.05; Think: m=0.91; MSE=0.05; Sleep: m=0.83; MSE=0.04; all ps >0.1), or the Single-Low (Blink: m=0.67; MSE=0.06; Think: m=0.67; MSE=0.11; Sleep: m=0.67; MSE=0.06; all ps >0.1) trials.

Confidence

Although responses for confidence were given on a non-numeric scale, the scale was re-coded for analysis onto a linear numerical scale from 1 to 9 (where "a"=1, "s"=2, "d"=3, etc.). Overall group means are shown in Figure 6. Confidence judgments were analyzed using between-condition Student's T tests, which showed Blink (m=6.64; MSE=0.36) had higher confidence than Think (m=5.80; MSE=0.34; t(32)=1.71, p=0.097, two-tailed) and Sleep (m=5.73; MSE=0.29; t(39)=1.98, p=0.055, two-tailed). These results were broken down by case file – Figure 7 shows that the Blink condition increased confidence on single cue trials (m=6.93; MSE=0.42) over Think (m=5.73; MSE=0.40; t(32)=2.12, p=0.042, two-tailed) and Sleep (m=5.43;MSE=0.34; t(39)=2.85, p=0.007, two-

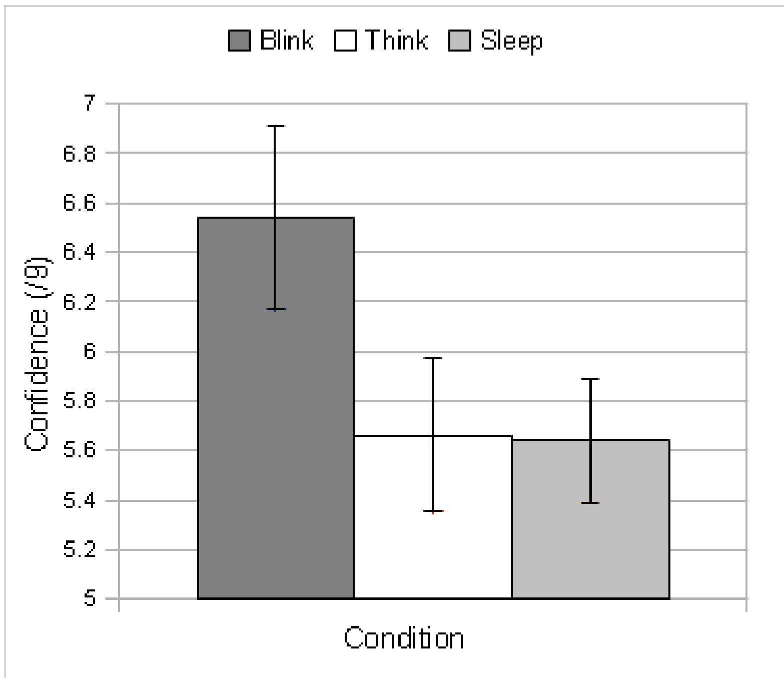

Figure 6. Average confidence ratings by condition in Study 2. Blink (m=6.54; MSE=0.37) had higher confidence than Think (m=5.66; MSE=0.31) or Sleep (m=5.64; MSE=0.25).

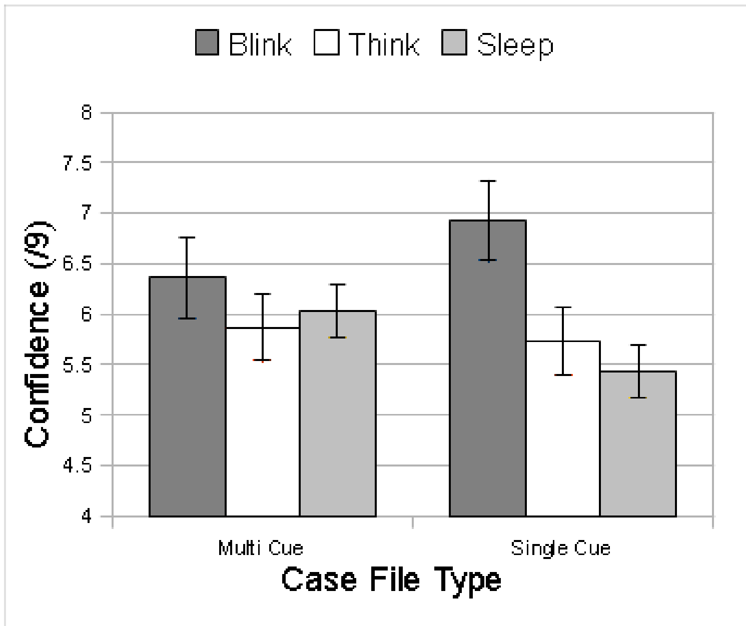

Figure 7: Confidence by condition and case file type in Study 2. Blink condition significantly increased confidence on Single-cue judgments (m=6.93; MSE=0.42) over Think (m=5.73; MSE=0.44) and Sleep (m=5.43; MSE=0.28). However, there were no significant differences between conditions on the multiple cue trials (Blink: m=6.37; MSE=0.4; Think: m=5.87; MSE=0.33; Sleep: m=6.03; MSE=0.26)

tailed). However, there were no significant difference between conditions on the multiple cue trials (Blink: m=6.34; MSE=0.38; Think: m=5.87; MSE=0.35; Sleep: m=6.03; MSE=0.30; all *p*s >0.3).

Choice Reaction Times

Reaction times for each choice question are plotted in Figure 8. Subjects responded much slower to the first question (m=9429ms; MSE=2467ms) than to questions 2-12 (m=3056ms; MSE=107ms). This is likely an artifact, as subjects took extra time the first go around to get familiar with the response keys. For this reason, question 1 was excluded from reaction time analysis. Subjects in the Blink condition took significantly less time to respond

(m=1554ms; MSE=245ms) than subjects in the Think (m=3290ms; MSE=3849ms; t(32)=2.89, $p<0.01$, two-tailed), and Sleep (m=3849ms; MSE=342ms; t(39)=3.32, $p<0.01$, two-tailed) condition. This pattern was consistent across both multi-cue trials (Blink m=1815ms; Think m=2908ms; Sleep m=4293ms) and single cue trials (Blink m=1336ms; Think m=3608ms; Sleep m=3479ms).

Confidence Reaction Times

Reaction times for confidence judgments were also measured. The keyboard mapping for confidence was novel, and while it was explained prior to confidence judgment 1, subjects were clearly still learning how to express their confidence appropriately. As such, the learning curve was even steeper than for choice questions – reaction times for question 1 (m=11423ms; MSE=6932ms) were much higher than for the remaining questions

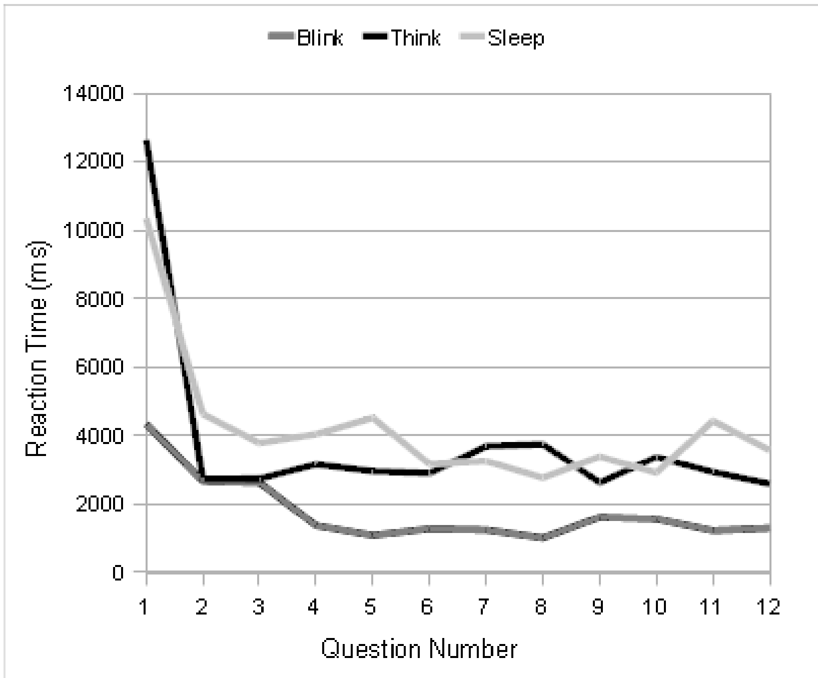

Figure 8: Reaction times for choice questions in Study 2. Subjects spent extra time on question 1 (m=9429ms; MSE=2467ms) learning the keyboard mapping for responses, so only questions 2-12 (m=3056ms; MSE=107ms) were analyzed.

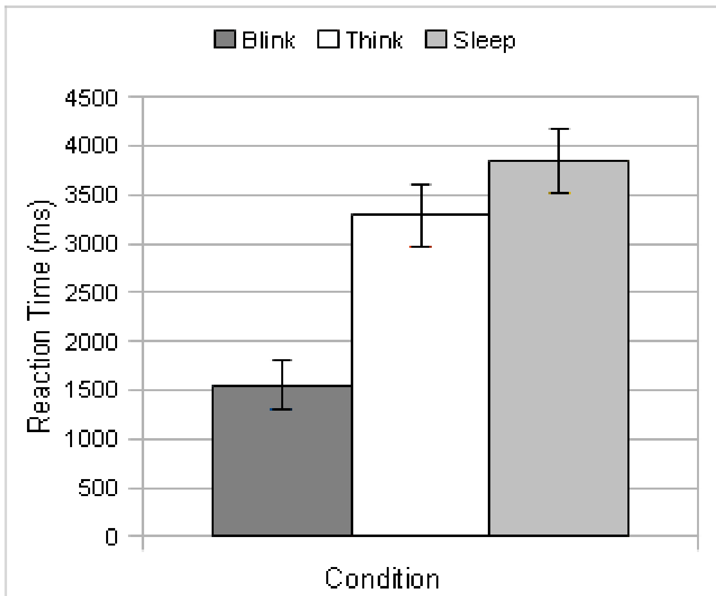

Figure 9: Reaction times for confidence judgments by condition in Study 2. Subjects in Blink condition responded faster (m=1277ms; MSE=167ms) than subjects in Think (m=2029ms; MSE=302ms) or Sleep conditions (m=2217ms; MSE=140ms).

(m=1905ms; MSE=1484ms). To minimize this issue, question 1 was again excluded from analysis. Group means for the remaining 11 confidence judgments are shown in Figure 9. Blink subjects again answered faster (m=1277ms; MSE=167ms) than subjects in Think (m=2029ms; MSE=302ms; $t(32)=2.55$, $p=0.016$, two-tailed) or Sleep (m=2217ms; MSE=140ms; $t(39)=2.74$, $p<0.01$, two-tailed) conditions. This pattern was consistent across both multi-cue trials (Blink m=1551ms; Think m=2514ms; Sleep m=2558ms) and single cue (Blink m=1049ms; Think m=1626ms; Sleep m=1933ms) trials.

Discussion

Study 2, like Study 1 before it, did not find evidence for a deliberation without attention effect. Subjects were equally likely to correctly identify the most likely flu in all three conditions, and even the advantage that deliberation and distraction had over quick decisions was erased. This lends even more skepticism that this particular MCPL paradigm benefits from distraction. As well, the relation between polarization and confidence was not as predicted. Subjects in Blink were the most confident, while there was no difference between Think and Sleep. The high confidence for Blink was primarily found in single cue trials. This is likely the result of the surprisingly short amount of time that subjects spent thinking about the question. The experimental paradigm allowed them to take as long as they wanted after the prompt, but Blink subjects responded fast even compared to the Think condition, which had been staring at the information for the problem for a full 100 seconds before the prompt. This fast responding carried over even into confidence judgments. This did not affect decision performance, however. This is likely because subjects were instructed to be “as accurate as possible” on the decisions, but this was not explicitly stated. There was likely a carry-over effect, where subjects who had been working on anagrams or the choice would take extra time to calibrate their confidence and those who had been going through the experiment quickly would not take care to make accurate confidence judgments– this effect is not unlike transfer inappropriate processing.

The results of Study 2 were disconcerting, as the expected carry-over of extremity into confidence did not hold. As such, it was decided to replicate Study 1 to get a better sense of what caused extremity differences between conditions. In addition, several methodological concerns would be addressed that had not been addressed in Study 1. The modified anagram task used in Study 2 would be applied in Study 3. Also, anagram performance was measured, to address a longstanding concern about the use of subjects' time over the distraction period. As a control group to compare against, Blink subjects were asked to do anagrams in between test case files. This also had the effect of (roughly) balancing the total time that each subject spent in the experiment, forcing subjects in the Blink condition to spend a level of effort and time similar to that of subjects in the other conditions.

<u>Study 3</u>

Methods

Subjects

103 subjects were recruited through the University of Waterloo Research Experiences Group. Subjects were compensated with course credit. Eleven subjects were excluded from analysis for exceptionally poor performance, defined as chance-or-below accuracy of implicit choices (as in Study 1), leaving 92 subjects in the final analysis.

Materials

Study 3 was a modified replication of Study 1. The training phase was exactly the same. As well, subjects were asked for frequency judgments for each case file. However, like Study 2, but not like Study 1, the final "all cues present" case file was dropped, leaving the six single cue case files and the six three-cue case files, for a total of twelve. Also, like in Study 2 but not Study 1, subjects in the Sleep condition were allowed to skip an anagram if, after ten seconds, they did not produce a solution. However, unlike either previous study, subjects' anagram performance was measured. This was to account for concern that there was no record that subjects in the Sleep condition were actually solving the word problems. As well, unlike either previous study, subjects in the Blink condition performed 90-second blocks of anagrams before each test case file. A summary of relevant comparisons to Dijksterhuis' initial study can be found in Table 6.

Results

Anagrams

The average number of anagrams solved by each participant across all 12 90-second blocks is plotted in Figure 10. A between-subjects t-test shows that Sleep solved marginally more anagrams (m=127.8; MSE=9.6) than Blink (m=106.5, MSE=8.6; $t(58)=1.65$, $p=0.105$, two-tailed).

Mean Absolute Deviation

Comparing the average MAD score across all twelve test case files, between-condition Student's T tests showed no significant effect of condition (Blink: m=17.89, MSE=1.65; Think: m=17.37, MSE=1.14; Sleep: m=19.14, MSE=1.03; all *p*s >0.2). The results were broken down by test case type, shown in Figure 11. Single cue test cases showed no overall difference between conditions (Blink: m=18.18, MSE=1.96; Think: m=17.40, MSE=1.16; Sleep: m=15.90, MSE=1.20; all *p*s >0.3). However, on multiple cue test cases Sleep (m=20.62, MSE=1.19) performed marginally worse than Blink (m=16.80, MSE=1.74; $t(58)=1.51$, $p=0.136$, two-tailed) and Think (m=17.35, MSE=1.41; $t(64)=1.77$, $p=0.081$, two-tailed).

Implied Choice Accuracy

Implied choices were derived along the same lines as in Study 1. Similar to MAD scores, Student's T-tests analyzing the pairwise between-condition comparisons showed that overall, there were no differences between conditions (Blink: m=0.78, MSE=0.036; Think: m=0.81, MSE=0.025; Sleep: m=0.76, MSE=0.024; all *p*s >0.15). These results were further

	Current Study 3	Dijksterhuis (2004)
Number of judgments	Twelve	One
Blink condition	90 s of anagrams, 10 s w/ cue values	10 s, no information
Think condition	100 s w/ cue values	240 s, no information
Sleep condition	10 s w/cue values, 90 s of anagrams	240 s of anagrams

Table 6: Summary of changes in decision context to adapt the manipulations of Dijksterhuis (2004) to the multiple cue probability learning task used in Study 3.

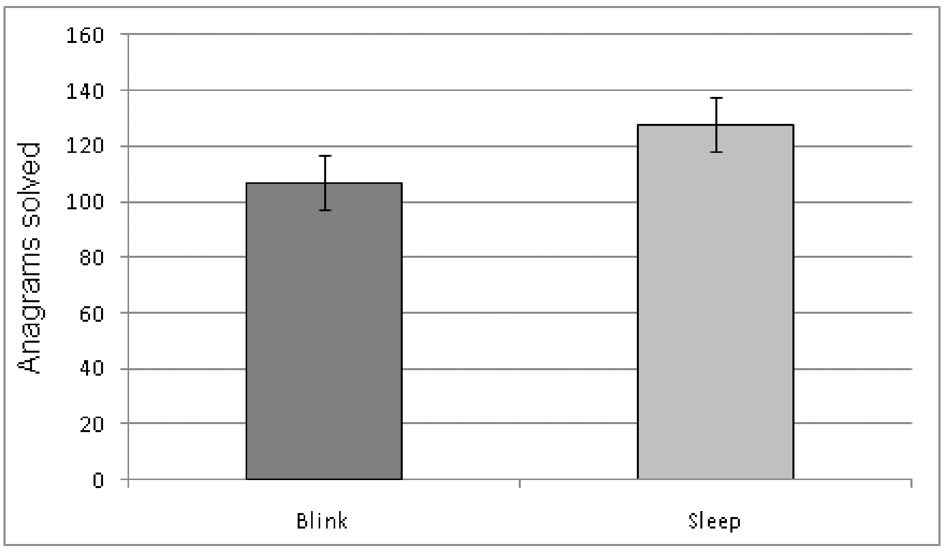

Figure 10: Average number of anagrams solved across 12 90-second blocks in Study 3. Sleep solved marginally more puzzles (m=127.8; MSE=9.6) than Blink (m=106.5, MSE=8.6).

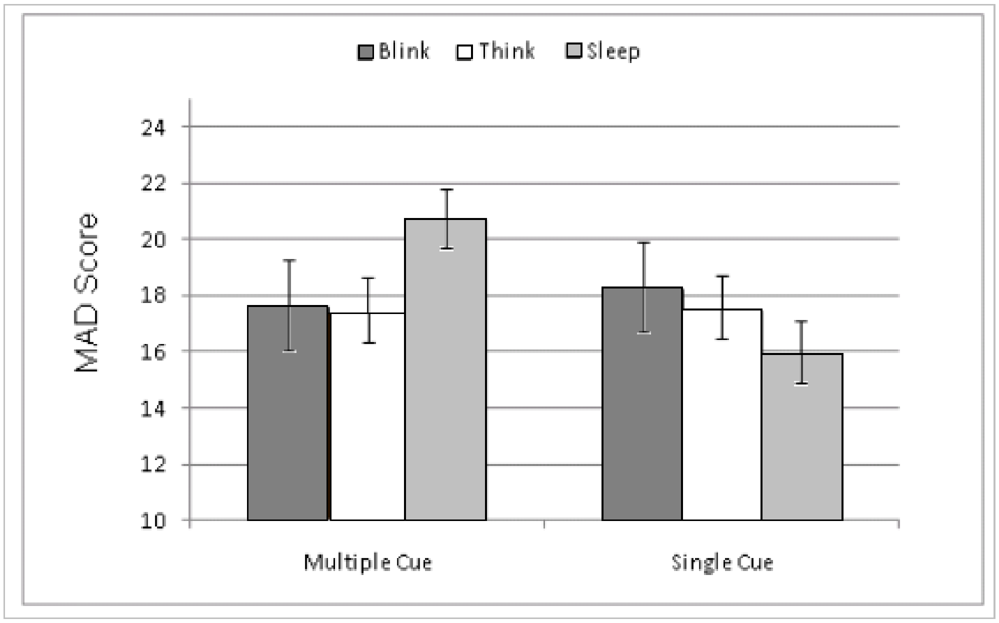

Figure 11: Average MAD score of judgments by condition and test case type in Study 3. Single cue test cases showed no overall difference between conditions (Blink: m=18.18, MSE=1.96; Think: m=17.40, MSE=1.16; Sleep: m=15.90, MSE=1.20). However, on multiple cue test cases Sleep (m=20.62, MSE=1.19) performed marginally worse than Blink (m=16.80, MSE=1.74) and Think (m=17.35, MSE=1.41).

analyzed as a function of test case file types, which is shown in Figure 12. While there were no significant differences between conditions for single cue test cases (Blink: m=0.77, MSE=0.048; Think: m=0.81, MSE=0.028; Sleep: m=0.81, MSE=0.026; all *p*s >0.4), on multiple cue test cases Sleep (m=0.72, MSE=0.035) was poorer at this measure of performance than both Blink (m=0.79, MSE=0.033; t(58)=1.49, *p*=0.145, two-tailed) and Think (m=0.82, MSE=0.032; t(64)=2.04, *p*= 0.045, two-tailed).

Extremity

Extremity was calculated in the same manner as in Study 1. Between-condition Student's T tests showed no significant differences between conditions (Blink: m=-6.95, MSE=2.43; Think: m=-5.52, MSE=2.42; Sleep: m=-6.25, MSE=2.29; all *p*s >0.6). Again the results were broken down by case file type, shown in Figure 13. No significant differences were found between conditions in single high-diagnosticity cue case files (Blink: m=-7.01, MSE=3.47; Think: m=-4.71, MSE=2.91; Sleep: m=-4.93, MSE=3.23; all *p*s > 0.6), multiple cue case files (Blink: m=-10.06, MSE=2.00; Think: m=-9.65, MSE=2.60; Sleep: m=-10.73, MSE=2.66; all *p*s > 0.7), or single low-diagnosticity cue case files (Blink: m=3.36, MSE=3.63; Think: m=2.84, MSE=3.52; Sleep: m=-3.58, MSE=3.80; all *p*s >0.19).

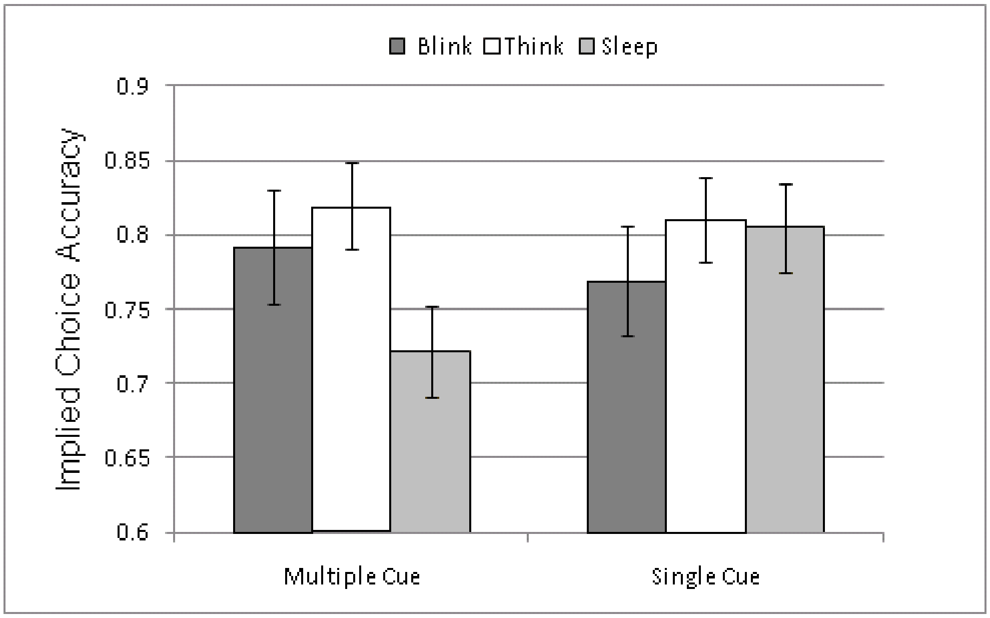

Figure 12: Implied choice accuracy by test case file type in Study 3. While there were no significant differences between conditions for single cue test cases (Blink: m=0.77, MSE=0.048; Think: m=0.81, MSE=0.028; Sleep: m=0.81, MSE=0.026), on multiple cue test cases Sleep (m=0.72, MSE=0.035) was poorer at this measure of performance than both Blink (m=0.79, MSE=0.033) and Think (m=0.82, MSE=0.032).

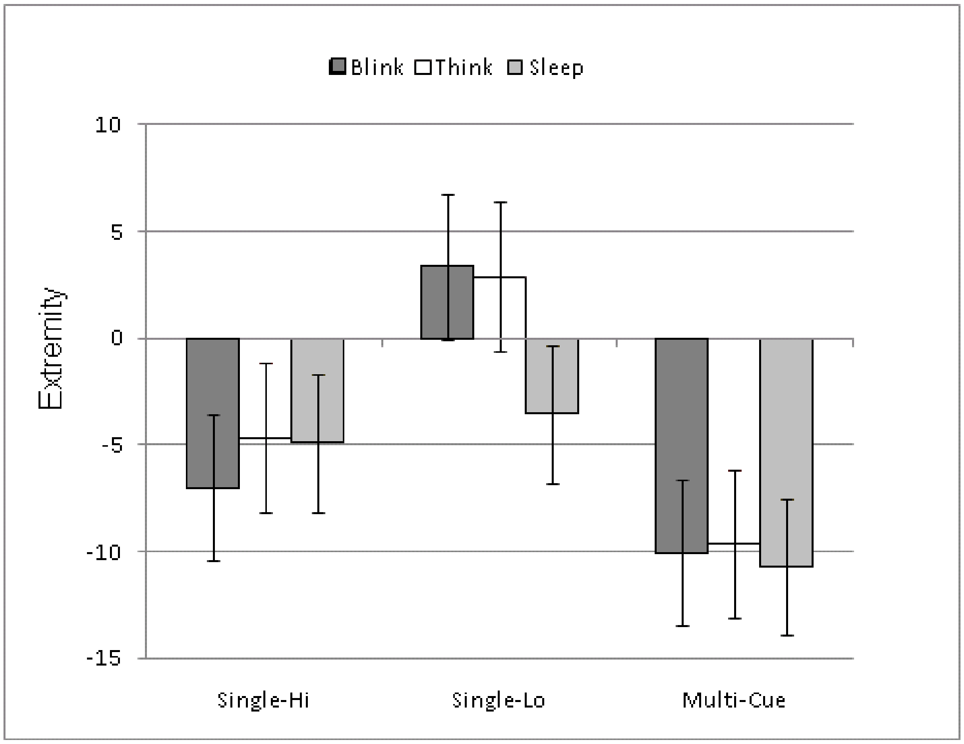

Figure 13: Between-condition comparisons of average judgment extremity in Study 3. Only test cases on which the implied choice would be accurate were included. No significant differences were found between conditions in single high-diagnosticity cue case files (Blink: m=-7.01, MSE=3.47; Think: m=-4.71, MSE=2.91; Sleep: m=-4.93, MSE=3.23), multiple cue case files (Blink: m=-10.06, MSE=2.00; Think: m=-9.65, MSE=2.60; Sleep: m=-10.73, MSE=2.66), or single low-diagnosticity cue case files (Blink: m=3.36, MSE=3.63; Think: m=2.84, MSE=3.52; Sleep: m=-3.58, MSE=3.80).

Discussion

The results of Study 3 stand in stark contrast to those of Study 1. Blink was no longer performing worse than Think and Sleep, and Sleep was marginally worse at multiple cue judgments than the other conditions, although the trend is not strong, especially in light of the first study. The only substantive change in the Sleep condition from Study 1 to Study 3 – that subjects who were stuck on one anagram were able to switch to a new puzzle after ten seconds – was not predicted to have any effect on performance, and certainly not to only influence multiple cue judgments. These results are most parsimoniously explained as the result of random variation – replication is the hallmark of a relevant effect in experimental psychology and, in these studies, the combination of results from the two studies leads to an average effect size far too close to zero to be worth interpreting. Similar conclusions can be reached about the analysis of the extremity data. Having said that, the relative improvement in performance of the Blink condition compared to Think, may be of some interest. By having subjects in Blink solve anagrams before approaching the judgments, they may have been "warmed up" and better prepared to make probability judgments using analytical faculties. Alternatively, because they were now held in the experiment room , they may have seen less benefit from rushing though decisions (in terms of percentage of time spent in the lab) and so may have conceded to experimenters demands and given a better effort. Neither of these effects was predicted, but both agree with prevailing theories about problem solving, and may have theoretical implications.

General Discussion

The three studies in this thesis demonstrate no evidence for a deliberation without attention effect. This was always a potential hazard for the experiment – Many studies, in paradigms even closer to Dijksterhuis' original design, including direct replications, have failed to find an effect (see Acker, 2008). This task requires much more analytical processing than preference-based tasks that are typically used, and may not be amenable to the purely association-based, non-analytical unconscious networks to which the deliberation without attention effect is commonly ascribed. Payne et al (2009) have shown that introducing tiered diagnosticity in the cue structure eliminates the deliberation without attention effect, which suggests that Dijksterhuis' (2006) initial prediction that increases in the complexity of a task lead to better relative performance of unconscious thought may be a simplistic view. A simpler, one-tiered multiple cue probability learning paradigm may better answer these questions. It may also be that deliberation without attention is vulnerable to repeated decisions. The same evidence may be useful (and therefore put to use) many times in this experiment, and this repeated retrieval and processing of stored memories about the task may block out any temporary benefits from unconscious organization of case file memories by the unconscious mind. If so, an effect might have been found had the manipulation been introduced before the test phase rather than during individual test case files. That is, if the learning phase was immediately followed by five minutes of deliberation (Think), five minutes of anagrams (Sleep), or nothing (Blink) right before the test cases (which would have to be identical across conditions), then memory organization might have been more sensitive to the difference between forced deliberation and distraction. Deliberation without attention might require a minimum amount of time to make a practical difference on decisions, and the stop-and-start nature of the manipulations used, over the course of many decisions, may not have allowed the difference to emerge. Whatever the case, though, there is little room for conclusions to be made from the above studies about deliberation without attention.

However, some theoretical implications may be warranted. Comparing Studies 1 and 3, the performance of those subjects given only ten seconds before being prompted for judgment improved, relative to the other groups, when they were given anagram puzzles for 90 seconds in between test cases. Likewise, results from Study 2 suggest that Blink subjects responded much quicker than they had to, despite not having nearly as much pre-prompt time to go over their decision as other groups. This suggests that quick decisions suffer not only from performance deficits in more complex problems, but also for the meta-cognitive awareness, both to judge confidence and to allocate an appropriate amount of time to a problem. The improvement of judgments in this quick decision condition by having subjects do anagrams beforehand (in Study 3) may result either from activation of analytical resources for later use, akin to transfer inappropriate processing shifts (Schooler, 2002) or a more basic interpretation that, subjects who were forced to solve anagrams had less to gain from rushing through the judgments, as a percentage of the total time in the experiment. Whatever the underlying mechanism, it appears that cognitive effort may have momentum, and unlike a muscle, where previous effort depletes later ability, decisions may be improved from expended effort beforehand. Obviously the data here are only suggestive, and further study is required.

The attempt to measure anagram performance was inconclusive, but did at least rule out the possibility that subjects were entirely neglecting the distractor task, something that, if it had been measured in previous studies, had not been mentioned in the resulting publications. However, the results are not conclusive, given the inherent variation in anagram solution speed over any amount of time. Some work recently has explored mind-wandering and off-task problem solving, but most of this work has focused on the prevalence and predictors of mind wandering episodes, and such work often uses thought probes and depends on subjects' accurate and honest self-report of off-task thinking (Mason et al., 2007; Smallwood & Schooler, 2006). However, trying to identify the content of those episodes might trigger a Heisenberg's Uncertainty Principle dilemma whereby introspection might change how those episodes might otherwise be integrated into the decision. The sustained attention to response task – or SART – has a very high temporal resolution, with responses required roughly once a second, so drifting can at least be distinguished from thinking hard about the task (Robertson et al., 1997). Future study might instead use the SART as a distractor task.

A different multiple cue probability learning paradigm might also have been more fruitful. A more diverse set of test case files would have allowed for modeling of the decision strategies (see White, 2006). This would have required much more time and attention from subjects, but might have at least opened a larger window into subjects' understanding of the training set. As well, a different cue structure might be considered. The current structure was chosen for logistical reasons above all else (a similar version having been previously used in the same laboratory). Using Figure 1, taken from Dijksterhuis et al. (2006), as a guide, the task was intended to be as complex as was manageable given the time constraints on an experiment. This may have been a mistake, as even the author himself did not have enough of a grasp of the cue structure *a priori* to appreciate how various factors would come into play during the analysis.

That being said, the deliberation without attention effect was explored in a paradigm very different from previous studies, and a determination of the exact reason why Sleep did not outperform Think may not be warranted given the limited data for comparison. This investigation applied a reasonable approximation of Dijksterhuis' principles of unconscious thought to a complex and well-known task, and found no evidence of the expected benefit of distraction in decision making. While some hypothetical multiple cue probability learning task might potentially be influenced by the effect, the above paradigms have outlined several tasks that are not. It is not unreasonable, however, to conclude that this is not a final word on the potency of deliberation without attention in multiple cue probability learning, and more study is needed to understand this fascinating phenomenon.

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(4),* 292-303.

Ariely, D. (2008). *Predictably Irrational*. New York: HarperCollins Publishers.

Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. American Psychologist, 54, 462-479.

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.

Bowden E. M. & Jung-Beeman, M. (2003). Normative data for 144 compound remote associate
problems. Behavior Research Methods, Instruments, & Computers, 35, 634-639.

Christoff, K., Gordon, A. & Smith, R. (2010). The Role of Spontaneous Thought in Human Cognition. In O. Vartanian & D. R. Mandel (Eds.), Neuroscience of Decision Making. New York, NY: Psychology Press.

Descartes, R. (1637). *Discourse on the Method of Rightly Conducting One's Reason and of Seeking Truth in the Sciences.* Retrieved August 2, 2009, from Project Gutenberg: http://www.gutenberg.org/etext/59

Dijksterhuis, A. (2004). Think Different: The merits of unconscious thought in preference development and decision making. *Journal of Personality and Social Psychology, 87(5),* 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. & 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(2),* 95-109.

Epley, N. & Gilovich, T. (2001). Putting Adjustment Back in the Anchoring and Adjustment Heuristic. *Psychological Science, 12,* 391-396.

Evans, J.St.B.T. (2003). In Two Minds: Dual process accounts of reasoning. *Trends in Cognitive Sciences, 7(10),* 454-459.

Gigerenzer, G. & Hoffrage, U. (1995). How to Improve Bayesian Reasoning Without Instruction: Frequency Formats. *Psychological Review, 102(4),* 684–704.

Gladwell, M. (2005). *Blink: The Power of Thinking Without Thinking*. New York: Back Bay Books.

Frederick, S. (2005). Cognitive Reflection and Decision Making. *Journal of Economic Perspectives, 19(4),* 25-42.

Herzog, S.M. & Hertwig, R. (2009). The Wisdom of Many in One Mind: Improving Individual Judgments with Dialectical Bootstrapping. *Psychological Science, 20(2),* 231-237.

Hsee, C.K. & Zhang, J. (2004). Distinction Bias: Misprediction and mischoice due to joint evaluation.
Journal of Personality and Social Psychology, 86(5), 680-695.

Jonides, J., Schumacher, E.H., Smith, E.E., Lauber, E.J., Awh, E., Minoshima, S., & Koeppe, R.A. (1997). Verbal working memory load affects regional brain activation as measured by PET. *Journal of Cognitive Neuroscience, 9,* 462–475.

Juslin, P., Jones, S., Olsson, H. & Winman, A. (2003a). Cue Abstraction and Exemplar Memory in Categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29(5),* 924–941.

Juslin, P. Olsson, H. & Olsson, A.-C. (2003b) Exemplar Effects in Categorization and Multiple-Cue Judgment. *Journal of Experimental Psychology: General, 132(1),* 133–156.

Larrick, R. (2009). *Broaden the Frame*. Unpublished manuscript, Duke University, Durham, NC.

Liu, W. (2009). Focus on Desirability: The Effect of Decision Interruption and Suspension on Preferences. *Journal of Consumer Research, 35(4),* 640-652.

Kahneman, D. (December 8, 2002). *Maps of Bounded Rationality.* Nobel Prize Lecture. Stockholm: Stockholm University.

Kane, M.J., Brown, L.H., McVay, J.C., Silvia, P.J., Myin-Germeys, I. & Kwapil, T.R. (2007). For Whom the Mind Wanders, and When: An Experience-Sampling Study of Working Memory and Executive Control in Daily Life. *Psychological Science, 18(7),* 614-

621.

Klein, G. (1999). *Sources of Power: How People Make Decisions*. Boston: The MIT Press.

Koehler, D.J., White, C.M. & Grondin, R. An Evidential Support Accumulation Model of Subjective Probability. *Cognitive Psychology, 46,* 152-197.

Mason, M. F., Norton, M. I., Van Horn, J. D., Wagner, D. M., Grafton, S. T., Macrae, C. N. (2007). Wandering minds: The default network and stimulus-independent thought. *Science, 315,* 393-395.

Melcher, J.M. & Schooler, J.W. (1996). The misremembrance of wines past: verbal and perceptual expertise differentially mediate verbal overshadowing of taste. *Journal of Memory and Language, 35,* 231–245.

Melcher, J.M. & Schooler, J.W. (2004). Perceptual and conceptual training mediate the verbal overshadowing effect in an unfamiliar domain. *Memory & Cognition, 32(4),* 618-631.

Meissner, C.A. & Brigham, J.C. (2001). A Meta-Analysis of the Verbal Overshadowing Effect in Face Recognition. *Applied Cognitive Psychology, 15,* 603-616.

Miller, G.A. (1956). The magical number seven, plus or minus two. *Psychological Review, 63(2),* 81-97.

Newell, B.R., Wong, K.Y., Cheung, J., & Rakow, T. (2008). Think, Blink, or Sleep on It? The impact of modes of thought on complex decision making. *The Quarterly Journal of Experimental Psychology, 62(4),* 707-732.

Nisbett, R.E. & Wilson, T.D. (1977). Telling More Than We Can Know: Verbal reports on mental processes. *Psychological Review, 84(3),* 231-259.

Payne, J.W., Samper, A., Bettman, J.R., & Luce, M.F. (2009). Boundary conditions on Unconscious Thought. *Psychological Science, 19(11),* 1118-1123.

Robertson. I.H., Manly, T., Andrade, J., Baddeley, B.T., & Yiend, J. (1997). 'Oops!': Performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia, 35(6),* 747-758.

Schneider, W., Zuccolotto, A., Gilkey, D., Christopher, D., Carper, C., Basler, L., Cernicky, B., Cotchen, M., Kuntz, L., & Eschman, A. (2008). E-Prime Suite (Version 1.5) [computer program]. Pittsburgh, PA: Psychology Software Tools, Inc.

Schooler, J.W. (2002). Verbalization produces a Transfer Inappropriate Processing Shift. *Applied Cognitive Psychology, 16,* 989-997.

Schooler, J.W. & Engstler-Schooler, T.Y. (1990). Thoughts beyond words: Some things are better left unsaid. *Cognitive Psychology, 22(1),* 36-71.

Schooler, J.W., Ohlsson, S. & Brooks, K. (1993). Thoughts beyond words: When language overshadows insight. *Journal of Experimental Psychology: General, 122(2),* 166-183.

Shafir, E., Simonson, I. & Tversky, A. (1992). Reason-Based Choice. *Cognition, 49,* 11-36.

Smallwood, J. M. & Schooler, J. W. (2006). The restless mind. *Psychological Bulletin, 132*, 946-958.

Soll., J, Larrick, R. & Zhu, X. (2005) The Wisdom of Many in One Mind.

Stanovich, K.E. & West, R.F. (2000). Individual Differences in Reasoning: Implications for the Rationality Debate? *Behavioral and Brain Sciences. 22(5)*, 645–726.

Strick, M., Dijksterhuis, A., Bos, M.W., Sjoerdma, A., van Baaren, R.B., & Nordgren, L.F. (2009). *A Meta-Analysis on Unconscious Thought Effects.* Unpublished manuscript, Radboud University, Nijmegen, The Netherlands. Retrieved July 25, 2009, from Nijmegen Unconscious Lab: www.unconsciouslab.com

Surowiecki, J. (2004). *The Wisdom of Crowds*. New York: Random House.

Thaler, R. & Sunstein, C. (2008). *Nudge: Improving decisions about health, wealth, and happiness.* New Haven: Yale University Press.

Tversky, A. & Kahneman, D. (1974). Judgment Under Uncertainty: Heuristics and Biases. *Science, 185(4157),* 1124-1131.

Tversky, A. & Koehler, D.J. (1994). Support Theory: A Non-Extensional Representation of Subjective Probability. *Psychological Review, 101(4),* 547-567.

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

Vul, E., & Pashler, H. (2008). Measuring the Crowd Within: Probabilistic Representations Within Individuals. *Psychological Science, 19(7),* 645-647.

Wegner, D.M., (1994). Ironic Processes of Mental Control. *Psychological Review, 101(1),* 636-647.

White, C.M. (2005) Deterministic versus Probabilistic Responding in Multiple Cue Probability Learning. Doctorate thesis: University of Waterloo, Waterloo, Canada.

Wilson, T.D., & Brekke, N. (1994). Mental contamination and mental correction: unwanted influences on judgments and evaluations. *Psychological Bulletin and Review, 116(1),* 117-42.

Wilson, T. D., Lisle, D., Schooler, J. W., Hodges, S. D., Klaaren, K. J., & LaFleur, S. J. (1993). Introspecting about reasons can reduce post-choice satisfaction. *Personality and Social Psychology Bulletin, 19,* 331–339.

Wilson, T.D. & Schooler, J.W. (1991). Thinking Too Much: Introspection can reduce the quality of preferences and decisions. *Journal of Personality and Social Psychology, 60,* 181-192.

Zhong, C.-B., Dijksterhuis, A., & Galinsky, A. (2009). The merits of Unconscious Thought in creativity. *Psychological Science, 19(9),* 912-918.