Woolley Amanda, Kostopoulou Olga, Delaney Brendan C. Can medical diagnosis benefit from “unconscious thought”? *Medical Decision Making*. doi: 10.1177/0272989X15581352. First published online 7 April 2015.

TITLE: Can medical diagnosis benefit from "unconscious thought"?

Running head: Medical diagnosis and unconscious thought

AUTHORS:

Amanda Woolley, BA

Olga Kostopoulou, PhD

Brendan C. Delaney, MD

Department of Primary Care and Public Health Sciences, Faculty of Life Sciences & Medicine, King's College London

Correspondence: Olga Kostopoulou, King's College London, Faculty of Life Sciences & Medicine, Department of Primary Care & Public Health Sciences, Capital House, 42 Weston Street, London SE1 3QD, UK

E-mail: olga.kostopoulou@kcl.ac.uk

Tel: +44 (0)20 7848 6757

Fax: +44 (0)20 7848 6620

Prior presentations of this work: 1st SMDM Asia Pacific conference (Singapore, 6-8 January 2014)

Financial support: Amanda Woolley was supported by a departmental PhD studentship.

Word count: 4637

Number of tables: 2

ABSTRACT

The Unconscious Thought Theory argues that making complex decisions after a period of distraction can lead to better decision quality than deciding either immediately or after conscious deliberation. Two studies have tested this Unconscious Thought Effect (UTE) in clinical diagnosis with conflicting results. The studies used different methodologies and had methodological weaknesses. We attempted to replicate the UTE in medical diagnosis by providing favorable conditions for the effect, while maintaining ecological validity.

Family physicians (N=116) diagnosed three complex cases in one of three thinking modes: immediate, unconscious (UT) and conscious (CT). Cases were divided into short sentences, which were presented briefly and sequentially on computer. After each case presentation, the immediate response group gave a diagnosis; the UT group performed a 2-back distraction task for 3 minutes before giving a diagnosis; and the CT group could take as long as necessary before giving a diagnosis.

We found no differences in diagnostic accuracy between groups (*P*=0.95). The CT group took a median of 7 seconds to diagnose, which suggests that physicians were able to diagnose "online", as information was being presented. The lack of a difference between the immediate and UT groups suggests that the distraction had no additional effect on performance.

To assess the decisiveness of the evidence of this null result, we computed a Bayes Factor (BF_{01}) for the two comparisons of interest. We found a BF_{01} of 5.76 for the UT vs. immediate comparison, and of 3.61 for the UT vs. CT comparison. Both BFs provide substantial evidence in favor of the null hypothesis: physicians' diagnoses made after distraction are no better than diagnoses made either immediately or after self-paced deliberation.

INTRODUCTION

According to the Unconscious Thought Theory (UTT), there are two modes of thought: a conscious mode, where thinking about a task happens while the task is in the focus of conscious attention; and an unconscious mode, where thinking occurs while conscious attention is directed elsewhere.[1] The unconscious mode of thought has a much greater capacity for holding, weighting and integrating information than the conscious mode. Therefore, decisions made after a period of distraction with an unrelated task can be better than those made after conscious deliberation. This phenomenon is termed the “Unconscious Thought Effect” (UTE), previously known as the “Deliberation Without Attention Effect”.[2]

In typical demonstrations of the UTE, participants are presented with several options, e.g., apartments, which vary on a number of attributes, e.g., size, location, cleanliness, and price amongst others. Each attribute is presented sequentially on a computer screen in either a random or a fixed order and for a few seconds only. Following this, and depending on the experimental condition, participants are either instructed to think consciously for several minutes (usually 4 minutes) or are distracted to prevent any conscious thought for the same length of time or, in some studies, are asked to respond immediately. Those who are distracted are found to choose the best option (defined either objectively or subjectively) more frequently than in the other conditions and to form more favorable impressions of that option.

The Unconscious Thought Theory has been strongly criticized[3 4] and has been the topic of numerous replication attempts, both successful and unsuccessful, with conflicting findings even at the level of meta-analysis. [5 6 7] Although choice has been the domain par excellence for demonstrations of the UTE, there is some evidence that it may not be the only one. The effect has also been investigated in sports predictions,[8] moral dilemmas,[9] evaluating the persuasiveness of messages,[10] and clinical diagnosis.[11 12]

Two studies have investigated the UTE in clinical diagnosis, with conflicting results. Mamede and colleagues compared diagnoses of written cases by internal medicine residents and 4th-year medical students under conscious, unconscious and immediate thought modes, in a within-subjects design.[12] The UTE was found with medical students but not with doctors and only in simple rather than in complex cases. This is in contrast to the Unconscious Thought Theory, which would predict the effect for complex cases and experienced participants. There are two methodological concerns with this study. First, the conscious thought condition consisted of an unusual procedure that required participants to list on paper all relevant diagnostic hypotheses, and for each hypothesis to list the expected cues that were present in and those that were absent from the case. This is clearly more systematic than what clinicians would naturally do if they were asked to think about a case before making a diagnosis. Furthermore, it is likely that this procedure would alleviate the working memory load of the conscious thinkers, which is proposed to limit the success of the conscious thought mode. Second, the complex cases consisted of uncommon diseases or atypical presentations. It is not clear whether 4th-year students had the necessary knowledge to diagnose such cases, which would undermine their ability to encode properly all the information necessary to diagnose, as required for the UTE to occur.

Another attempted replication of the UTE in clinical diagnosis was in the domain of mental health, where participants (clinical psychology students) diagnosed two cases in either an unconscious or a conscious thought mode.[11] The cases were described as “complex” because each included two diagnoses (DSM classifications). Participants were asked to give two diagnoses per case. The study found a sizeable UTE (η^2=0.15) in a one-way ANOVA, where accuracy (range 0-4) was treated as a continuous variable. Inspection of Table 1 in the publication (p. 580) shows more instances of 2 correct diagnoses in the unconscious than the conscious thought mode. There are however methodological concerns here too. First, there was no immediate response condition. This means that we cannot know whether the effect was due to unconscious thought improving performance or conscious thought impairing it. Neither can we know to what extent participants' responses reflected a diagnosis made before rather than during the period of distraction or deliberation. Second, diagnoses for the two cases were collected after participants had read both cases. This possibly confounded the distraction manipulation, as even those in the conscious thought mode would have had their attention taken away from the first case (distraction) while reading the second case. Finally, neither of the two studies had an ecologically valid deliberation condition, thereby comparing UT with a more natural thinking mode.

We sought to conduct an improved replication of the UTE in medical diagnosis with family physicians as participants and an ecologically valid deliberation condition. The study aimed to investigate differences in the diagnostic accuracy of physicians diagnosing complex cases either immediately or after distraction or after deliberation.

METHODS

Materials

We constructed three clinical vignettes based on real patient cases. In an earlier study, we asked family physicians to describe cases where they felt that they knew the diagnosis but did not know how they knew.[13] We selected three of these cases for use in the current study. Case selection was based on the following criteria:

1) Case descriptions contained sufficient information in order for the final diagnosis (i.e., the diagnosis of the original physician, which was also the patient's actual diagnosis) to be objectively the "best answer".

2) Cases represented different conditions/disease areas. Limiting materials to a single case or to cases from a single disease area might advantage participants with specialist knowledge in that area.*

3) Cases represented different types of diagnostic difficulty. Case 1 involved a final diagnosis that was unlikely for the patient and a more common, competing diagnosis that partially explained the symptoms. Case 2 had multiple, competing, potential causes. Case 3 could easily elicit an unlikely but life-threatening diagnosis that needed urgent treatment, and a more likely, competing diagnosis that explained the symptoms (the final diagnosis).

An experienced family physician (BCD) helped to analyze the cases and write them in the form of vignettes for use in the study (see Appendix). He also confirmed that the final diagnosis was the best answer in each case.

Sample size and recruitment

To calculate the required sample size, we used the data in Table 1 of the article by De Vries and colleagues (p. 580).[11] The table presents the number of correct DSM classifications (diagnoses), by case and thinking mode (UT and CT). Our vignettes had only one diagnosis each; therefore our accuracy variable was binary (correct/incorrect). On the other hand, each case in the De Vries et al. study had 2 diagnoses (co-morbidities); therefore, the researchers scored the responses per case as 0 (no correct diagnosis), 1 (1 correct diagnosis) or 2 (both diagnoses correct). To calculate sample size for a dichotomous outcome variable, we converted the 0-2 scores in the article's Table 1 into ether 1 or 0, where 1 meant that both diagnoses

* Although family physicians are "generalists", they often develop specialist interests and expertise in specific medical conditions or areas of medicine.

were correct. We also added up the responses from the two cases, which resulted in proportional diagnostic accuracy of 0.56 (45/80) for the UT group and 0.34 (27/80) for the CT group. To capture this difference in proportions between two groups with a power of 80% and alpha of 0.05, we would need 88 responses per group, i.e. a total of 264 responses (115 responses per group for 90% power). As the three cases varied widely in their content and type of difficulty, we expected independence of responses within participants. This is a common finding in studies of the diagnostic accuracy of general physicians, where responses to one case do not predict responses to other cases ("case specificity").[14-16] Therefore, we estimated that we would need 29 participants per group, providing 3 responses each, i.e., a total of 87 participants providing 261 responses. Nevertheless, we sought to recruit as many participants as possible within the time and resources available, in the event that the independence assumption was not met.

Participants were practicing family physicians and residents in family medicine in the UK. They were recruited through professional networks and at conferences. Participants were offered a £10 gift voucher in return for their time.

Procedure

Study materials were presented and responses were collected using a macro-enabled Microsoft PowerPoint presentation. Data collection took place over the Internet, in a single session, while participants were in concurrent phone communication with a researcher (AW), who facilitated the session.

The study followed a mixed factorial design, where each participant diagnosed all the vignettes in one of three thinking modes: immediate, conscious thought (CT) or unconscious thought (UT). Participants in the immediate response group gave their diagnosis immediately after each vignette presentation. In the CT group, participants were instructed to think carefully about their diagnosis after each vignette and respond when they were ready. We chose a self-paced rather than a fixed-time conscious thought condition, as more representative of the medical diagnostic process. In the UT group, participants completed a 2-back task for 3 minutes after each vignette, and then diagnosed. In the 2-back task, participants see a single digit presented in the middle of the screen. They first press “next” for a new digit to be presented, then “next” again for a third digit. At this point, “yes” and “no” buttons are provided and the participant needs to decide for each new digit whether it is the same as the digit shown 2 places previously (2 back). N-back tasks (recalling digits seen “n” places previously), although simple to perform, require continuous conscious attention, and thus use all available working memory preventing deliberation of the diagnosis task.[17]

All participants completed a short practice case, including the 2-back task for those in the UT group. Subsequently, they saw the three vignettes in a random order. Each vignette was broken down into individual cues, which were presented in the middle of the screen, one at a time, for 4 seconds each. The sequence of cue presentation was the same for all participants, as shown in the Appendix. Participants were instructed to assume that any symptom, sign or test result not included in the vignettes was absent/normal.

At the end of each vignette presentation and when appropriate for each condition (i.e., after physicians in the UT condition had completed the n-back task and physicians in the CT condition had indicated that they were ready to diagnose), a new screen appeared that instructed participants to “Please enter your diagnosis NOW”. In all conditions, responses on this screen were time-restricted to 20 seconds to prevent extensive thinking. A countdown clock was displayed on the screen.

As a manipulation check, we recorded the 2-back scores of the UT group. We also recorded time spent thinking before a diagnosis was given in the CT group.

Analyses

Diagnoses were scored as either correct or incorrect, depending on whether they matched the final diagnosis of each case. When participants gave more than one diagnosis, their response was scored as correct if it included the final diagnosis. We also employed a more detailed accuracy measure by scoring responses as correct, partially correct, or incorrect. We then collapsed correct with partially correct responses for a more generous measure of accuracy.

In case 1 (rectal cancer), partially correct diagnoses included inflammatory bowel conditions, such as Crohn's disease and ulcerative colitis, which could explain some of the symptoms but not the spontaneous rectal bleeding and visible rectal mass. Diagnoses such as hemorrhoids and anxiety were scored as incorrect, as they did not acknowledge the possibility of serious disease.

In case 2 (heart failure), partially correct responses included angina, ischemic heart disease, or simply "heart problem" or "cardiac problem", as well as any mention of cancer. Ischemic heart disease is a pre-existing condition in this patient but could not cause fatigue without heart failure. Cancer, specifically lung cancer, is a possible diagnosis in this case but much less likely than heart failure. Diagnoses such as depression/low mood/psychological issues, obese/unfit patient, and viral illness were considered incorrect, as they ignored signs of serious disease.

In case 3 (migraine), subarachnoid hemorrhage (SAH) and meningitis could be considered partially correct, although the lack of any neurological signs to indicate raised intracranial pressure make SAH unlikely, and the lack of any neck stiffness or rash make meningitis unlikely.

We tested the assumption of independence of responses (diagnostic accuracy) per physician by calculating the intraclass correlation coefficient (ICC). We tested the effect of thinking mode on diagnostic accuracy using logistic regression and the association between thinking mode and the 3-level measure of diagnostic accuracy using a chi square test. Given that the theory predicts stronger UTE for more experienced participants and more difficult problems, we also tested for differences in diagnostic accuracy between the three cases and for associations between diagnostic accuracy and physician experience using logistic regressions. We report results from univariate regression models. Thinking times in the CT group were log transformed to correct for skewness and compared between correct and incorrect responses using a two-tailed t-test. We used STATA 13.1 to analyze the data.

Finally, we calculated Bayes factors for the two comparisons of interest: UT vs. Immediate, and UT vs. CT. This allows us to draw conclusions on the basis of null results, which would be considered inconclusive in the classical inferential statistics tradition. Bayes Factor (BF_{01}) is a likelihood ratio: it is the ratio of the probability of the data under H_0 (no difference in accuracy) to the probability of the data under H_1 (UT is more accurate than Immediate/CT modes). Bayes factor values have been classified into grades of evidence ranging from “decisive” (BFs > 100) to “anecdotal” (BFs range from 3 to 1).[18] To compute the Bayes factors, we used the web-based applet by Rouder and colleagues (http://pcl.missouri.edu/bf-two-sample) and kept the scale factor r on effect size to the default (1.0).[19] For the BF calculations, we used the number of respondents per group and calculated t values based on the mean accuracy per respondent (correct diagnoses out of 3).

RESULTS

One hundred and sixteen family physicians participated, including 15 residents in family medicine. Half of the participants were female and the mean experience of the sample was 11 years in family medicine (SD 11, median 6, range 0-44 years). We obtained data on two rather than three cases from five respondents due to a computer failure, resulting in a total of 343 responses. There were 40 participants in the UT group providing a total of 118 responses, 38 participants in the immediate response group (112 responses) and 38 in the CT group (113 responses). The assumption of independence of responses was verified: the ICC for diagnostic accuracy approached 0 and was not significant (P=0.50).

The mean accuracy on the 2-back task (correct responses over all responses) was 0.88 (SD 0.09). No participant scored less than 0.50, suggesting that they were paying attention to the task, which fulfilled its role as a distractor.

Table 1 presents mean diagnostic accuracy (proportion of correct diagnoses over all diagnoses) per thinking mode and case. We found no effect of thinking mode on diagnostic accuracy: odds ratio (OR) 0.99 [95% CI 0.56 to 1.75] for immediate and OR 0.70 [0.39 to 1.28] for CT compared to UT. Case 3 (migraine) was diagnosed correctly more often than both case 1 (rectal cancer) (OR 3.26 [95% CI 1.77 to 6.01], P<0.001) and case 2 (heart failure) (OR 2.47 [1.38 to 4.42], P=0.002), and this was observed across all thinking modes (Table 1). We detected no relationship between physician experience and accuracy (OR 1.00 [0.98 to 1.03]).

Table 2 presents frequencies and percentages of correct, partially correct and incorrect responses per thinking mode and case. We found no relationship between thinking mode and this 3-level measure of diagnostic accuracy (P=0.40). When we collapsed partially correct and correct diagnoses for a more generous measure of accuracy, we found no effect of thinking mode: OR 1.00 [95% CI 0.58 to 1.72] for immediate and OR 0.78 [0.45 to 1.32] for CT compared to UT.

Thinking times in the CT group ranged from 2 seconds to 2:87 minutes (median 7 seconds), with only four instances where respondents took longer than a minute. No differences in thinking times were detected between correct and incorrect diagnoses (*P*=0.24).

The Bayes Factor (BF_{01}) for the UT vs. immediate comparison was 5.76; the data are almost 6 times more likely to occur if there is no difference between UT and immediate thinking modes than if there is a difference. This provides "substantial" evidence in favor of the null hypothesis, according to the proposed grades of evidence (BFs range from 10 to 3 in the "substantial" category).[18] The BF_{01} for the UT vs. CT comparison was 3.61, also providing substantial evidence in favor of the null hypothesis. Given the independence of responses within physicians, as verified by the low ICC, we repeated the BF_{01} calculations, this time basing the t values on the raw accuracy data, rather than the physician means, and using the number of responses as each group's sample size. This resulted in higher BFs for both comparisons: 9.65[†] for UT vs. immediate and 5.02 for UT vs. CT.

DISCUSSION

We aimed to replicate the Unconscious Thought Effect (UTE) in medical diagnosis. We used cases based on real patients that reflected different types of diagnostic difficulty and family physicians as study participants. We used a between-participants design, which ensures that there are no carry-over effects from one thinking mode to another, as a participant diagnoses cases in alternate modes. Our distraction manipulation was successful and we can thus be confident that UT participants'

[†] Bayes factors between 30 and 10 provide "strong" evidence in favor of the null hypothesis.[18]

attention remained away from the diagnostic task during distraction. We included an immediate response condition, so that any extra benefits of UT could be properly assessed. Finally, we made the deliberation mode self-paced and thus, more ecologically valid.

We did not find any differences in accuracy between thinking modes, hence no evidence that one mode is more beneficial than another. In fact, the Bayes factors indicated that we found substantial evidence for the null hypothesis: physicians' diagnoses made after a period of distraction with an unrelated task are no better than diagnoses made either immediately or after self-paced deliberation. Accuracy depended on the case rather than on thinking mode.

We provided auspicious conditions to allow the UTE to manifest:

1. We gave participants a general goal, central to their decision making, i.e., the goal to diagnose, rather than a more specific and peripheral goal, e.g., to identify patient risk factors. We gave them this goal before rather than after the distraction, as required for the effect to occur. The meta-analysis by Strick and colleagues found larger UTEs when central-general goals were induced.[6]

2. Diagnostic accuracy for all our cases was low, suggesting that they were indeed difficult. The meta-analysis found that complex decision problems led to larger effect sizes than simple decision problems.

3. We used experts rather than novices as study participants. There is evidence that experts benefit from unconscious thought more than non-experts.[8] At the same time, the meta-analysis cautions that too much expertise could inhibit UTE by making the task too easy. This was not the case in our study (see point 2 above).

4. We kept the distraction time relatively short (3 minutes); the meta-analysis by Strick and colleagues found that longer intervals of 7 or 8 minutes hampered the UTE.

5. Our materials were ecologically valid. Our cases were based on real patients, were complex, and contained the information that the family physicians, who had diagnosed the cases correctly, had elicited. They were therefore likely to have induced motivation and interest in our participants. The meta-analysis found that ecologically valid tasks tend to generate larger UTEs.

Our study thus adds to the number of studies that have failed to find any evidence of the superiority of "unconscious thought".[5 7 20-26] Could we have provided even more auspicious conditions for the UTE to occur?

1. The meta-analysis by Strick and colleagues found that word-search puzzles produced larger UTEs than more cognitively demanding distraction tasks but this made a difference only for the UT vs. CT comparison and not for the UT vs. immediate comparison. The authors suggested that such tasks potentially compete with UT for resources. We expected physicians to be motivated to provide accurate diagnoses and not stop thinking about the cases during relatively easy distraction tasks, therefore, we opted for a cognitively demanding distraction task.

2. We allowed CT participants to give a diagnosis when they wished and did not impose on them a fixed time for thinking. Payne and colleagues did not find a UTE when using a self-paced CT condition vs. a UT condition and argued that the effect is due not to the superiority of UT but to overthinking and interference resulting from unnatural thinking times in the CT mode.[22]

3. Unconscious thought is purported to deal well with large amounts of information that exceed the processing capacity of conscious thought. We attempted to load participants' working memory by presenting information quickly on screen. Nevertheless, it is likely that participants diagnosed "online", despite the quick succession of time-limited information. The sequence of information had a narrative structure, which would have facilitated online processing and integration. Participants in the CT mode, who had the opportunity to think about their diagnosis, gave it within a few seconds, suggesting that online processing was possible. The lack of differences in accuracy between the immediate and UT modes suggests that diagnoses after distraction were simply the diagnoses that participants made during or at the end of each case's presentation and that no further "unconscious" processing took place during distraction. The same argument was made recently by Nieuwenstein and colleagues, whose study participants (psychology undergraduates) took on average only 23 seconds in a multi-attribute choice task, even though they had the opportunity to deliberate for as long as they needed.[27]

Using self-report type methods, Heneghan and colleagues attempted to chart family physicians' diagnostic strategies.[28] They found different variants of pattern-recognition and rule-based strategies, and no analytical reasoning. Although self reporting may not provide reliable evidence about thinking processes, it shows that physicians consider themselves to rely on fast, automatic or semi-automatic processes almost exclusively. In a study by Flores and colleagues, clinical psychologists read clinical reports, consisting of a number of sentences sequentially presented on screen (the pace of presentation was determined by the clinician).[29] One sentence per report was inconsistent with the disorder named at the beginning of the report. Reading times for inconsistent sentences about highly relevant symptoms were longer than for inconsistent sentences about less relevant symptoms. The authors suggested that participants were able to detect inconsistencies by employing fast, online reasoning processes. The text comprehension literature assumes that readers make online inferences during reading, and that this enables readers to maintain text coherence.[30 31] Similarly, theories of cognitive consistency[32] assume that people do not wait to receive all the information before making a judgment, but process information actively, as it arrives, sometimes distorting it to support their emerging judgments[33] (this "predecisional information distortion" has also been established with family physicians[34 35]). Finally, there is evidence that it is the order in which information is received rather than the thinking mode that determines final judgments, with "unconscious thought" not adding anything to and not enhancing the final judgment in any way.[36] These multiple sources of evidence converge to suggest that fast, online processing represents an important hurdle for the UTE. Strick and colleagues *"suspect that unconscious thought thrives most when the presentation time is just long enough to encode the information, but also short enough to prevent further on-line processing"* (p. 757).[6] Thus, the UTE appears extremely sensitive to experimental manipulations, and the ideal conditions for it seem hard to achieve, and may not be achievable in real life decision making. We should also mention a very recent attempt to replicate the UTE in multi-attribute choice: the study by Nieuwenstein and colleagues was both adequately powered and maintained all the conditions found in Strick and colleagues' meta-analysis to favor the UTE. Nevertheless, no differences were found between the UT and deliberation conditions (there was no immediate response condition), leading the authors to suggest that the UTE is found only in studies with small sample sizes, which produce unreliable results. This hypothesis was subsequently supported by their meta-analysis of 69 studies of the UTE in multi-attribute choice, which found that the UTE was confined to studies of low precision.[27]

Study weaknesses and future research

Although we used cases based on real doctor-patient encounters, our materials could still be criticized as not sufficiently ecologically valid. Our case vignettes were provided entirely in verbal form. Some of the information would have been obtained visually by the physician; for example, the appearance of the hemorrhoids upon examination (case 1), and the general appearance of the obese patient complaining of exhaustion (case 2). This could explain the low diagnostic accuracy obtained in these two cases – the original physicians who had diagnosed them accurately and later recounted them to us had seen these patients, not their vignettes. It is also likely that different physicians would have requested different information from their patients. Therefore, despite the control afforded by presenting the same verbal descriptions to participants – which also enabled comparability with the two previous studies of UTE in medical diagnosis[11 12] – we need to be cautious when applying our conclusions to real life clinical encounters. Future research could test the UTE paradigm with more realistic clinical materials, perhaps even with standardized patients (i.e. actors). Distraction in such realistic situations could be operationalized not with anagrams and n-back tasks but with more plausible interruptions, such as phone calls and requests for prescriptions or medical advice, unrelated to the diagnosis at hand.

At the expense of ecological validity, future studies attempting to replicate the UTE in medical diagnosis could employ a “forced” CT condition, where participants are given a specific time to think about the problem, as in the original experimental paradigm.[2] However, we note that UK family physicians have 10-minute consultations; forcing them to think for 4 whole minutes about the diagnosis would be alien to their experience and entirely impractical for real life implementation. Future studies may also use simpler distraction tasks and develop manipulations that impede online processing of case information; for example, shorter presentation times, cues that convey the necessary meaning using symbols and abbreviations rather than complete sentences, or non-narrative information structures. We should however note that even if the UTE were eventually achieved with such manipulations, its value for real life clinical practice would be minimal, unless it were shown to provide some advantage over the natural way that physicians diagnose.

Finally, we should acknowledge an important difference between choice tasks, the domain par excellence of the UTE, and diagnostic tasks. In choice tasks, respondents are presented with explicit, alternative options and their attributes. In diagnosis, respondents must generate their own alternatives, while the cues presented are not pre-identified as supportive or not of the alternatives and must be interpreted by the participants. Previous research has suggested that diagnostic errors often occur because the physician has failed to even generate the correct hypothesis,[14 37 38] whilst information search and interpretation depend on the hypothesis considered. [39 34] Therefore, in situations where the correct hypothesis has already been formulated, and diagnosis more closely resembles a choice between options, the UT effect may be more likely to occur. In other words, had we presented our participants with a list of diagnostic alternatives to consider, our results might have been different. This hypothesis could be tested in future studies.

Acknowledgements

The computer program used for the presentation of study materials was adapted from a program developed by Stefan Mangold and technicians at the University of Göttingen. We would also like to thank Dr Miguel Vadillo and five anonymous reviewers for their helpful comments on an earlier draft of this paper.

Ethical approval for the study was obtained from London – Surrey Borders Research Ethics Committee, reference 11/LO/1112.

References

1. Dijksterhuis A, Nordgren LF. A Theory of Unconscious Thought. Perspect Psychol Sci 2006;**1**(2):95-109.
2. Dijksterhuis A, Bos MW, Nordgren LF, et al. On making the right choice: The deliberation-without-attention effect. Science 2006;**311**(5763):1005-07.
3. Gonzalez-Vallejo C, Lassiter GD, Bellezza FS, et al. "Save Angels Perhaps": A critical examination of Unconscious Thought Theory and the deliberation-without-attention effect. Review of General Psychology 2008;**12**(3):282-96.
4. Newell BR, Shanks DR. Unconscious influences on decision making: A critical review. Behavioral and Brain Sciences 2014;**37**:1-61.
5. Acker F. New findings on unconscious versus conscious thought in decision making: additional empirical data and meta-analysis. Judgment and Decision Making 2008;**3**(4):292-303.
6. Strick M, Dijksterhuis A, Bos MW, et al. A Meta-Analysis on Unconscious Thought Effects. Social Cognition 2011;**29**(6):738-62.
7. Newell BR, Rakow T. Revising Beliefs About the Merit of Unconscious Thought: Evidence in Favor of the Null Hypothesis. Social Cognition 2011;**29**(6):711-26.
8. Dijksterhuis A, Bos MW, van der Leij A, et al. Predicting Soccer Matches After Unconscious and Conscious Thought as a Function of Expertise. Psychological Science 2009;**20**(11):1381-87.
9. Ham J, van den Bos K. On Unconscious Morality: The Effects of Unconscious Thinking on Moral Decision Making. Social Cognition 2010;**28**(1):74-83.
10. Handley IM, Runnion BM. Evidence That Unconscious Thinking Influences Persuasion Based on Argument Quality. Social Cognition 2011;**29**(6):668-82.
11. De Vries M, Witteman CLM, Holland R, et al. The Unconscious Thought Effect in Clinical Decision Making: An Example in Diagnosis. Med Dec Making 2010:578-81.
12. Mamede S, Schmidt HG, Rikers RMJP, et al. Conscious thought beats deliberation without attention in diagnostic decision-making: at least when you are an expert. Psychol Res-Psych Fo 2010;**74**(6):586-92.
13. Woolley A, Kostopoulou O. Clinical intuition in family medicine: more than first impressions. Ann Fam Med 2013;**11**(1):60-6.
14. Elstein AS, Shulman LS, Sprafka SA. *Medical problem solving: An analysis of clinical reasoning*. Cambridge, MA: Harvard University Press, 1978.
15. Elstein AS, Shulman LS, Sprafka SA. Medical Problem Solving: A Ten-Year Retrospective. Evaluation and the Health Professions 1990;**13**(1):5-36.
16. Kostopoulou O, Oudhoff J, Nath R, et al. Predictors of diagnostic accuracy and safe management in difficult diagnostic problems in Family Medicine. Med Decis Making 2008;**28**(5):668-80.
17. Jonides J, Schumacher EH, Smith EE, et al. Verbal working memory load affects regional brain activation as measured by PET. J Cognitive Neurosci 1997;**9**(4):462-75.
18. Wetzels R, Matzke D, Lee MD, et al. Statistical Evidence in Experimental Psychology: An Empirical Comparison Using 855 t Tests. Perspect Psychol Sci 2011;**6**(3):291-98.
19. Rouder JN, Speckman PL, Sun D, et al. Bayesian t tests for accepting and rejecting the null hypothesis. Psychon Bull Rev 2009;**16**(2):225-37.
20. Calvillo DP, Penaloza A. Are complex decisions better left to the unconscious? Further failed replications of the deliberation-without-attention effect. . Judgment and Decision Making 2008;**4**(6):509-17.
21. Newell BR, Wong KY, Cheung JCH, et al. Think, blink or sleep on it? The impact of modes of thought on complex decision making. Q J Exp Psychol 2009;**62**(4):707-32.

22. Payne JW, Samper A, Bettman JR, et al. Boundary Conditions on Unconscious Thought in Complex Decision Making. Psychological Science 2008;**19**(11):1118-23.
23. Thorsteinson TJ, Withrow S. Does unconscious thought outperform conscious thought on complex decisions? A further examination. Judgment and Decision Making 2009;**4**(3):235-47.
24. Waroquier L, Marchiori D, Klein O, et al. Methodological pitfalls of the Unconscious Thought paradigm. Judgment and Decision Making 2009;**4**(7):601-10.
25. Waroquier L, Marchiori D, Klein O, et al. Is It Better to Think Unconsciously or to Trust Your First Impression? A Reassessment of Unconscious Thought Theory. Soc Psychol Pers Sci 2010;**1**(2):111-18.
26. Nieuwenstein M, van Rijn H. The unconscious thought advantage: Further replication failures from a search for confirmatory evidence. Judgment and Decision Making 2012;**7**(6):779-98.
27. Nieuwenstein MR, Wierenga T, Morey RD, et al. On making the right choice: A meta-analysis and large-scale replication attempt of the unconscious thought advantage. Judgm Decis Mak 2015;**10**(1).
28. Heneghan C, Glasziou P, Thompson M, et al. Diagnostic strategies used in primary care. BMJ 2009;**338**:b946.
29. Flores A, Cobos PL, Lopez FJ, et al. Detecting fast, online reasoning processes in clinical decision making. Psychological assessment 2014;**26**(2):660-5.
30. Albrecht JE, O'Brien EJ. Updating a mental model: Maintaining both local and global coherence. Journal of Experimental Psychology-Learning Memory and Cognition 1993;**19**(5):1061-70.
31. Kendeou P, Smith ER, O'Brien EJ. Updating during reading comphrehension: Why causality matters. Journal of Experimental Psychology: Learning, Memory & Cognition 2013;**39**:854-65.
32. Simon D, Snow CJ, Read SJ. The redux of cognitive consistency theories: Evidence judgments by constraint satisfaction. J Pers Soc Psychol 2004;**86**(6):814-37.
33. Russo JE, Carlson KA, Meloy MG, et al. The goal of consistency as a cause of information distortion. Journal of Experimental Psychology-General 2008;**137**(3):456-70.
34. Kostopoulou O, Russo JE, Keenan G, et al. Information distortion in physicians' diagnostic judgments. Med Dec Making 2012;**32**(6):831-9.
35. Nurek M, Kostopoulou O, Hagmayer Y. Predecisional information distortion in physicians' diagnostic judgments: Strengthening a leading hypothesis or weakening its competitor? Judgment and Decision Making 2014;**9**(6):572-85.
36. Gonzalez-Vallejo C, Cheng JQ, Phillips N, et al. Early Positive Information Impacts Final Evaluations: No Deliberation-Without-Attention Effect and a Test of a Dynamic Judgment Model. Journal of Behavioral Decision Making 2014;**27**(3):209-25.
37. Kostopoulou O, Devereaux-Walsh C, Delaney BC. Missing celiac disease in family medicine: the importance of hypothesis generation. Med Decis Making 2009; **29**(3):282-90.
38. Kostopoulou O, Rosen A, Round T, et al. Early diagnostic suggestions improve accuracy of GPs: a randomised controlled trial using computer-simulated patients. Br J Gen Pract 2015;**65**(630):e49-e54.
39. Kostopoulou O, Mousoulis C, Delaney BC. Information search and information distortion in the diagnosis of an ambiguous presentation. Judgm Decis Mak 2009;**4**(5):408-18.

Table 1: Mean diagnostic accuracy [95% CIs] per thinking mode and case.

	Immediate	UT	CT	Total
Case 1	0.19 [0.08 – 0.35]	0.23 [0.12 – 0.38]	0.11 [0.03 – 0.25]	0.17 [0.11 – 0.26]
Case 2	0.21 [0.10 – 0.37]	0.25 [0.13 – 0.41]	0.19 [0.08 – 0.35]	0.22 [0.15 – 0.30]
Case 3	0.46 [0.30 – 0.63]	0.40 [0.24 – 0.57]	0.37 [0.22 – 0.54]	0.41 [0.32 – 0.50]
Total	0.29 [0.20 - 0.38]	0.29 [0.21 – 0.38]	0.22 [0.15 – 0.31]	0.27 [0.22 – 0.32]

Table 2: Frequencies (%) of incorrect, partially correct and correct responses per case and thinking mode.

		Incorrect	Partially correct	Correct	Total responses
Case 1	Immediate	23 (62.2%)	7 (18.9%)	7 (18.9%)	37
	UT	23 (57.5%)	8 (20%)	9 (22.5%)	40
	CT	27 (71%)	7 (18.4%)	4 (10.5%)	38
Case 2	Immediate	14 (36.8%)	16 (42%)	8 (21%)	38
	UT	18 (45%)	12 (30%)	10 (25%)	40
	CT	18 (48.7%)	12 (32.4%)	7 (18.9%)	37
Case 3	Immediate	2 (5.4%)	18 (48.7%%)	17 (46%)	37
	UT	0	23 (60.5%)	15 (39.5%)	38
	CT	1 (2.6%)	23 (60.5%)	14 (36.8%)	38

APPENDIX: The 3 vignettes broken down into individual cues. Each bulleted cue was presented sequentially on screen for 4 seconds. The potentially diagnostic cues are underlined and the differentiating cues are underlined and bolded.

Vignette 1	Vignette 2	Vignette 3
• 31-year old male. • Previous history of irritable bowel syndrome. • Investigated for IBS, including a colonoscopy. • *"When I open my bowels there's blood. There's blood on the toilet paper, and in the toilet."* • *"You know, quite a bit of blood. I mean, it's pretty alarming!"* • *"It's not really painful, but it's uncomfortable."* • *"I've been a little bit constipated over the past few weeks I suppose, but nothing too bad."* • *"I've never had piles before."* • *"My dad also gets these kinds of problems and it looks like it's IBS too."* • *"I'm really worried about this."* • *"My wife's due with our first baby in two weeks."* • *"I'm worried it's not going away before the baby comes."* • Around the edge of the anus there are several hemorrhoids and evidence of bleeding. • The patient is treated with an Anusol pessary and ointment. • Results from full blood count, ESR and ferritin are within normal ranges. • The results are given over the phone. • *"The stuff you gave me seems to have worked, the bleeding's stopped."* • One month later the patient requests a telephone consultation. • ***"I woke up this morning in a pool of blood."*** • *"I'm really worried about this."* • The patient attends the surgery the next day. • The patient looks tired. • *"I've not slept properly for over a month. The baby's up every few hours."* • *"I'm not that constipated."* • *"I didn't notice I was bleeding*	• 52-year old male. • **He had a coronary artery graft 7 years ago**. • He appears obese. • The skin on his face is like that of a smoker. • His fingers are stained yellow from tar. • His voice is very deep. • *"I just don't have any energy."* • *"Even after a short walk I need to sit down."* • *"I don't feel breathless, just exhausted."* • *"A month ago I felt fine."* • *"Then suddenly I feel like this."* • *"It came on over a few days."* • *"I'm finding it hard to do my job."* • *"Usually I'm working really hard all day, but I can't do that, feeling like this."* • *"I'm a chef."* • *"It's really busy and quite strenuous."* • ***"I've had angina for years*** • *but it hasn't got any worse."* • *"I've had no other pains in my chest."* • *"I used to smoke a packet a day."* • *"I stopped smoking after my bypass."* • Pulse is regular. • Heart sounds are normal. • Chest sounds are normal. • Results from a full blood count are within normal range. • Glucose is normal. • Thyroid function is normal. • Liver function is normal. • Kidney function is normal.	• 42-year old female. • She is a GP colleague at your practice. • She is very fit and in good health. • She complains of a severe headache. • *"Can you cover my patients?"* • *"It's just come on pretty quickly."* • *"I've taken some paracetamol."* • *"I don't think I can drive home like this."* • *"I'm just going to rest in my office for now."* • *"I'm sure I'll feel better soon."* • *"I'll get my husband to pick me up later if I still feel ill."* • You check on your colleague two hours later. • *"I feel terrible."* • *"It's got much worse."* • *"The headache is awful."* • *"I feel quite nauseous."* • *"I don't know what's wrong."* • *"I've never had a headache like this before."* • *"I never really get headaches."* • *"The light is really bothering me."* • *"The noise is bothering me too."* • *"It's just getting worse."* • She is alert. • She is able to walk. • She has normal sensation in her extremities. • Reflexes are normal. • **Examination of the eyes is unremarkable.** • **Her temperature is 36.7 °C.** • Her pulse is regular and

until I woke up." • *"There was a lot of blood."* • The hemorrhoids previously examined appear to have amalgamated. • **The affected area is larger and more obvious.**		normal.

IBS: Irritable Bowel Syndrome
ESR: Erythrocyte Sedimentation Rate, an inflammatory marker
COPD: Chronic Obstructive Pulmonary Disease
GP: General Practitioner (UK family physician)
Paracetamol: Pain relief medication
A temperature of 36.7 °C is normal