Detecting fast, on-line reasoning processes in clinical decision making

Amanda Flores, Pedro L. Cobos, Francisco J. López, and Antonio Godoy (1)

University of Málaga and Instituto de Investigación Biomédica de Málaga (IBIMA), Málaga (Spain)

(1) University of Málaga

Authors' Note

This research was supported by grants 2008-SEJ-03586 from Junta de Andalucía, 2007-63691/PSIC, and PSI2011-24662 from the Spanish Ministry of Science and Innovation. Amanda Flores had a FPI PhD scholarship that was awarded by Junta de Andalucía. We would like to thank to Colegio Oficial de Psicólogos de Andalucia Oriental, and José Díaz Ruiz for their invaluable help to conduct the experiment.

Correspondence concerning this article should be addressed to Amanda Flores, Departamento de Psicología Básica. Facultad de Psicología. Universidad de Málaga Campus de Teatinos, s/n (29072), Málaga, Spain. E-mail: amandafm@uma.es

Abstract

In an experiment that used the inconsistency paradigm, experienced clinical psychologists and psychology students performed a reading task using clinical reports and a diagnostic judgment task. The clinical reports provided information about the symptoms of hypothetical clients who had been previously diagnosed with a specific mental disorder. Reading times of inconsistent target sentences were slower than that of control sentences, demonstrating an inconsistency effect. The results also showed that experienced clinicians gave different weights to different symptoms according to their relevance when fluently reading the clinical reports provided, despite the fact that all the symptoms were of equal diagnostic value according to the DSM-IV. The diagnostic judgment task yielded a similar pattern of results. In contrast to previous findings, the results of the reading task may be taken as a direct evidence of the intervention of reasoning processes that occur very early, rapidly, and online. We suggest that these processes are based on the representation of mental disorders and that these representations are particularly suited to fast retrieval from memory and to making inferences. They may also be related to the clinician's causal reasoning. The implications of these results for clinician training are also discussed.

Key words: diagnostic criteria, clinical reasoning, inconsistency paradigm, causal reasoning.

The DSM-IV generally assumes that all the diagnostic criteria for a mental disorder are equivalent (Kim & Ahn, 2002). Despite this, some studies have shown that clinical psychologists and psychology students do not give the same weight to the different diagnostic criteria for a mental disorder when making a diagnosis. Specifically, Kim and Ahn (2002) showed that clinicians and psychology students were more likely to apply certain diagnostic categories when a hypothetical client presented a given set of symptoms than when she/he presented a different set. This result is consistent with other findings related to previous versions of the DSM (see Davis, Blashfield, and McElroy, 1993; Garb, 1996; Rubinson, Asnis, and Friedman, 1988). Kim and Ahn suggested that clinicians' and students' diagnostic judgments were affected by their idiosyncratic theories, which prevented them from giving the same weight to all the diagnostic criteria specified in the DSM-IV.

The standard approach to studying this situation has been based on tasks in which participants have to make diagnostic judgments with sufficient time to reflect on permanently available information about clients' symptoms. Thus, the tendency to give more weight to some symptoms rather than to others may be the effect of slow, effortful, and deliberative reasoning processes that take place when the participants are asked to make a diagnostic judgment. This raises the question of whether other reasoning processes that take place very early, rapidly, and in a partially unconscious manner may also be responsible for the differential weighting of symptoms. Specifically, these fast reasoning processes could take place online as part of the reasoners' comprehension processes as they receive relevant information about clinical cases. Online processes refer to a wide variety of well-timed processes triggered by a stream of incoming information on a clinical case. If this information were provided in written format, these processes would range from visual perception or lexical access to inference and integration. These online processes are essential to the clinician to obtain a global understanding of a clinical case. Importantly, these processes are

thought to be automatic or semiautomatic because they must occur very rapidly and simultaneously as readers proceed from one piece of information to the next. In theory, some of these fast online reasoning processes would rely on the reasoners' previous domain-specific knowledge and theories, which would underlie the different weights given to the different diagnostic criteria. This hypothesis may be especially pertinent in the case of expert clinicians according to some theories on how expert clinicians' represent and use their knowledge and theories (for example, see Charlin, Boshuizen, Custers, & Feltovich, 2007; Charlin, Tardif, & Boshuizen, 2000; Schmidt, Norman, & Boshuizen, 1990; Smith, 1989).

Previous studies have already shown that clinicians take a few minutes to make diagnostic decisions, with only slight variations in the resulting diagnosis if more time is taken (Kendell, 1973; Sandifer, Hordern, & Green, 1970). In this period of time, judgement-based, slow, reflective and resource-demanding processes can take place. However, this study focuses on processes that take place in a few tenths of a second and that demand very few cognitive resources. Kahneman's (2011) distinction between System 1 and System 2 processes provides an appropriate framework to differentiate between online, semiautomatic, and slow, deliberate reasoning processes. The online reasoning processes that are the focus of this study correspond to System 1 processes, which encompass numerous fast processes and heuristics that have been thought to underlie a huge number of biases and errors in reasoning and decision making. Therefore, the aim of this study was to show how System 1 processes affect the differential weighting of the DSM-IV diagnostic criteria, as demonstrated in experiments such as those conducted by Kim and Ahn (2002).

Reading clinical reports is an activity in which early, fast, online clinical reasoning may be found. All the major text comprehension studies have assumed that readers make online inferences during reading (Graesser, Singer, &Trabasso, 1994; McKoon & Ratcliff, 1992; van Dijk & Kintsch, 1983; Zwaan & Radvansky, 1998). One of the functions of online

inference-making is to maintain the coherence of a text at the global and local levels. Maintaining text coherence occasionally requires the search and discovery of links that connect different portions of the text (Black & Bower, 1980; Kendeou, Smith, & O'Brien, 2012; Schank, 1975; Trabasso & Sperry, 1985; Trabasso & van den Broek, 1985). It has been claimed that processes related to the search for coherence belong to System 1 processes (Kahneman, 2011). This viewpoint has been strongly supported by experiments using the so-called inconsistency paradigm in reading comprehension studies (Albrecht & O'Brien, 1993; Long & Chong, 2001; Peracchi & O'Brien, 2004). The results obtained suggest that reading an inconsistent sentence (i.e., a sentence inconsistent with preliminary information in a text) takes longer than reading a consistent or a neutral text. If readers encounter a contextually inconsistent sentence while attempting to maintain text coherence, more time and cognitive resources are needed to resolve the conflict. For example, if a preliminary sentence states that a given client has been previously diagnosed with avoidant personality disorder and, some sentences later, a target sentence states that she/he is convinced of being interesting, competent and appealing, the reader may detect an inconsistency provided he/she possesses the appropriate knowledge. Thus, reading times for an inconsistent target sentence should be longer than for a consistent or neutral one. Consequently, access to previous knowledge and online inference-making during reading can be directly detected under this paradigm. Inconsistency detection during fluent reading entails the following processes: a) fast access to domain-specific knowledge and theories; b) rapid inference-making from the target sentence based on prior knowledge and/or theories; and c) the detection of a contradiction between the inference drawn and the preliminary information (Long, Seely & Oppy, 1996).

Regarding the objectives previously described, the inconsistency paradigm can be specifically used to study fast, semiautomatic, online reasoning processes in relation to the different effect of each diagnostic criterion on diagnostic judgments. The rationale is quite

simple. A hypothetical clinical report begins with a series of sentences stating that a client has been diagnosed with, for example, avoidant personality disorder and that she/he presents some symptoms that form part of the DSM-IV diagnostic criteria for this disorder. Subsequently, a target sentence is encountered stating the absence of one of two possible symptoms (e.g., either the absence of "views self as socially inept, personally unappealing, or inferior to others", or the absence of "is unusually reluctant to take personal risk or to engage in any new activities"), both of which are considered to be diagnostic criteria for the disorder. According to the reader's previous theories about avoidant personality disorder, if the first symptom is more relevant to them than the second one, then the target sentence referring to the absence of the first symptom should cause more conflict than a target sentence referring to the absence of the second symptom. Thus, reading the more inconsistent target sentence should slow down the reading process more than reading a less inconsistent target sentence.

The other part of the strategy used in the experiment involved the selection of symptoms of varying degrees of relevance in order to create different degrees of inconsistency. This issue was addressed by drawing on Kim and Ahn's (2002) study. In their experiments, the participants (most of them expert clinicians) had to draw a causal map for each disorder, indicating the relationship between symptoms by the use of arrows. For each map and participant, the causal centrality score of each symptom was calculated according to a specific algorithm. Causally central symptoms were those that, according to the causal map, were responsible for the occurrence of many other symptoms which, in turn, might cause further symptoms. Peripheral symptoms were the effect of other symptoms and did not cause further symptoms. Finally, isolated symptoms were those that did not have any causal relationship with the other symptoms. After calculating the causal centrality score for each symptom per disorder per participant, an average score across participants was calculated, on the basis of which symptoms were ordered from the most central to the most peripheral (or

isolated) for each disorder. Kim and Ahn found that causally central symptoms had a greater impact on the participants' diagnostic judgments than peripheral and isolated symptoms despite the fact that, in all cases, the hypothetical clients presented symptoms that formed part of the diagnostic criteria for the different disorders according to DSM-IV. Thus, under the assumption that more central symptoms were more relevant than less central symptoms, the two symptoms with the highest mean centrality score and lowest mean centrality score within each disorder were selected to maximize the difference in relevance. Thus, it was predicted that the use of the central symptom in the target sentence would be associated with a greater inconsistency effect than the use of the peripheral symptom. It was also predicted that the central symptom would have a greater impact on diagnostic judgments than the peripheral symptom.

Method

Participants and apparatus. A total of 34 participants took part in the experiment on a voluntary basis. The sample consisted of psychology students (n=17; 50%) from Malaga University (Spain) and experienced clinicians (n=17; 50%) who worked in independent practice in the Malaga area. Their experience as clinicians ranged from three to 28 years (average 17 years).

Materials and design. A total of 24 clinical reports divided into two groups of 12 were created for the inconsistent and control conditions, respectively. The reports for the inconsistent condition included six different DSM-IV disorders, as in Kim and Ahn's study: major depressive disorder, specific phobia, antisocial personality disorder, schizophrenia, borderline personality disorder, and avoidant personality disorder; the disorders in the control condition included cannabis dependence, sleepwalking disorder, pathological gambling, orgasmic disorder, gender identity disorder, and hypochondria. In the inconsistent condition,

texts included a target sentence that stated the absence of a symptom regarded as a diagnostic criterion for the disorder that had been previously mentioned, whereas in the control condition the same target sentence appeared in a clinical report in which the diagnosed disorder bore no relationship to this absent symptom. In addition, the target sentences in six of the reports in the inconsistent condition related to highly relevant symptoms, whereas in the other six reports they related to symptoms of low relevance. As mentioned, in line with Kim and Ahn (2002), it was assumed that more causally central symptoms would have greater relevance than peripheral symptoms. Consequently, symptoms with the highest and lowest mean centrality scores were selected. Therefore, it was predicted that the participants would detect an inconsistency in the inconsistent condition alone, and hence reading times (RTs) for the target sentence would be longer in the inconsistent condition. It was also predicted that there would be a greater inconsistency effect (i.e., a greater difference in reading times between the target inconsistent and control sentences) associated with the absence of a highly relevant symptom than with a symptom of low relevance.

All the clinical reports were created using the same structure as the texts used in inconsistency paradigm experiments (Albrecht & O'Brian, 1993). Each clinical report consisted of 16 sentences of comparable length and semantics as well as syntactic complexity. The introductory sentence was followed by the DSM-IV diagnosis that the hypothetical client had received. The next six sentences included three sentences reporting the presence of three symptoms (one in each sentence) consistent with the disorder, intermixed with three more sentences including unrelated information. These symptoms had intermediate causal centrality scores in Kim and Ahn's (2002) study. In addition, two very frequent symptoms (i.e., present in numerous DSM-IV disorders) were also included. Immediately before the target sentence, there were four filler sentences related to non-clinical information. The filler information would make the previous information on clinical

symptoms unavailable from the participant's working memory by the time the target information was read. The last two sentences in the text were the target and the post-target sentences. In a text using avoidant personality disorder as an example, the participants in the inconsistent/highly relevant symptom condition read the following sentence: "...she is convinced of being interesting, competent and appealing...", which contradicts the criterion "...views self as socially inept, personally unappealing, or inferior to others..." (DSM-IV-TR). In the inconsistent/low relevance condition, the sentence was "...she becomes easily involved in new activities...", which contradicts "...avoids personal risk or new activities..." (DSM-IV-TR). The post-target sentence described clinically irrelevant information and was introduced to detect any carryover effect that could have been produced by reading the target sentence.

Procedure. The participants read the instructions on a computer screen and any questions were answered before the experimental task began. The instructions emphasized that attention should be paid to the task because after reading each clinical report they would be asked to what extent they agreed with a clinician's diagnosis. Thus, the use of clinical reasoning was encouraged during the reading task.

Participants were also instructed to carefully and fluently read the different reports. The whole text was initially unreadable as it was masked with slashes, one per written character. The reading task was self-paced: the participants were required to press the space bar — at which point the sentence became readable — in order to proceed from one sentence to the next; returning to previous text was not permitted. Once the space-bar was pressed the previous sentence became unreadable again. The participants were presented with a sample text to familiar themselves with the reading procedure.

Immediately after the text had been read, the diagnostic judgment task was administered in which the participants had to rate on a continuous 0 to 100 (i.e., from complete disagreement to complete agreement) rating scale the extent to which they agreed

with the diagnosis provided in the text. Once the participant had rated the diagnosis they rested for a few minutes before proceeding to the next clinical report.

The experimental task took place in two sessions, separated by at least 1 week. The participants read 12 different clinical reports in each session. Assigning different texts per session ensured that the participants could not read the same target and post-target sentences twice within the same session. The reading order of the different texts within each session was randomized. The procedure followed ensured that in each session, six of the clinical reports were from the inconsistent condition and six were from the control condition. Orthogonally to this, half of the clinical reports were from the highly relevant condition, whereas the other half were from the low relevance condition. Each session took from 20 to 30 minutes to complete.

Results

Reading times. The RTs for the target and post-target sentences were analyzed. An α of .05 was used in all the statistical analyses. The RTs were filtered by removing outliers that were 3 standard deviations from the mean. Following the filtering process, a single mean RT per experimental condition and participant was calculated, yielding four averaged measures for the target sentences and another four for the post-target sentences. In total, 12 target sentence RTs and 10 post-target sentence RTs were eliminated.

Table 1 shows the mean RTs for the target and post-target sentences in each condition within each sample. As shown, the students' and the clinicians' RTs for the target sentence were longer in the inconsistent condition than in the control condition, this being consistent with an inconsistency effect. In addition, in the case of clinicians, the difference in RTs between the inconsistent condition and the control condition appeared to be greater in the condition including highly relevant symptoms than in the condition including symptoms of low relevance; this was not observed in the sample of students. This finding was confirmed

by conducting separate analyses for each sample. A repeated measures ANOVA 2 (Inconsistency: Inconsistent vs Control) x 2 (Relevance of the symptoms: High vs Low) on the students' RTs yielded a significant main effect of inconsistency [F (1, 16) = 24.091, MSE = 271235.080; $p < .001$; $\eta^2 = .56$]. None of the other effects were significant (all F values < 2.92). The same trend, although much smaller, was observed for the post-target sentence. However, an identical ANOVA 2 on the RTs for the post-target sentence yielded no significant effect (all F values < 1.27). Regarding the sample of clinicians, the same ANOVA 2 on RTs for the target sentences yielded a significant effect of inconsistency [F (1, 16) = 12.801, MSE = 1036503.4, $p = .003$, $\eta^2 = .44$], relevance of symptoms [F (1, 16) = 9.043, MSE = 186673.901, $p = .008$, $\eta^2 = .36$], and inconsistency x relevance of symptoms [F (1, 16) = 6.505, MSE = 286012.319, $p = .021$, $\eta^2 = .289$]. Simple effects analyses revealed a significant inconsistency effect in both conditions of the relevance of symptoms factor, [F (1, 16) = 14.203, MSE = 882461.693, $p = .002$, $\eta^2 = .47$; F (1, 16) = 5.899, MSE = 440053.978, $p = .027$, $\eta^2 = .27$], for the high relevance and low relevance conditions, respectively. Table 1 also reveals similar results for the RTs of post-target sentences, i.e., a greater effect of inconsistency in the high relevance condition than in the low relevance condition. This was confirmed by the same ANOVA 2, which yielded a significant main effect of inconsistency [F (1, 16) = 5.565, MSE = 297226.601, $p = .031$, $\eta^2 = .26$], and a marginally significant effect of the inconsistency x relevance of symptoms interaction [F (1, 16) = 3.691, MSE = 125731.452, $p = .073$, $\eta^2 = .19$]. The main effect of relevance of symptoms was not significant ($F < 0.66$). Planned tests for simple effects yielded an inconsistency effect within the high relevance condition [F (1, 16) = 5.873, MSE = 329524.764, $p = .028$, $\eta^2 = .27$], but not within the low relevance condition [F (1, 16) = 1.958].

These results indicate that, during the reading task, the students and clinicians both engaged in some form of fast online clinical reasoning that entailed the retrieval and use of

DSM-IV diagnostic criteria for the mental disorders used. In the case of the students, the online reasoning processes were not modulated by the relevance of symptoms, as no inconsistency x relevance of the symptoms interaction was found. However, in the case of the clinicians, the online reasoning processes were not completely in accordance with the DSM-IV prescriptions as the inconsistency effect was modulated by the relevance of symptoms despite all the symptoms being of equivalent diagnostic value.

Diagnostic judgments. A single mean diagnostic judgment (i.e., the degree of agreement with the diagnosis provided) per participant was calculated for each experimental condition within each sample (see Table 1). In general, there was more agreement among the participants on the diagnosis stated in the preliminary information in the control condition than in the inconsistent condition, indicating familiarity with the diagnostic criteria of DSM-IV. However, the difference between means was greater in the high relevance condition than in the low relevance condition within the sample of clinicians, but not within the sample of students. This was confirmed by statistical analyses. A repeated measures ANOVA 2 (Inconsistency: Inconsistent vs Control) x 2 (Relevance of symptoms: High vs Low) on the students' judgments yielded a significant main effect of inconsistency [F (1, 16) = 39.754, MSE = 189.997; $p < .001$; $\eta^2 = .71$]. Neither the effect of relevance of symptoms nor the interaction between the two factors were statistically significant (all F values < 3.16). Regarding the sample of clinicians, the same ANOVA 2 yielded a significant effect of inconsistency [F (1, 16) = 46.896, MSE = 247.928, $p < .001$, $\eta^2 = .75$] and inconsistency x relevance of symptoms [F (1, 16) = 5.586, MSE = 128.387, $p = .031$; $\eta^2 = .26$]. The effect of relevance of symptoms did not reach significance [F (1, 16) = 1.431]. The inconsistency effect was greater in the high relevance condition than in the low relevance condition. Nevertheless, simple effects analyses revealed that the effect was significant in both conditions, [F (1, 16) = 31.44, MSE = 288.161, $p < .001$; $\eta^2 = .66$; F (1, 16) = 37.257, MSE =

88.154, $p < .001$; $\eta^2 = .7$] for the high relevance condition and low relevance condition, respectively.

****

Please insert Table 1 about here

****

These results are consistent with those identified in the RT analysis and suggest that the reasoning processes that occurred during the reading task could also be responsible for the effects that were observed in the diagnostic judgment task. Specifically, the greater impact of the highly relevant symptoms on the clinicians' diagnostic judgments could have been determined by the greater impact of these symptoms on the online and fast reasoning processes during the reading task. The results also show that the symptoms that were given different weights by the participants in Kim and Ahn's (2002) study were also given different weights by the sample of experienced clinicians in this study.

In addition, the results of the psychology students and the experienced clinicians were analyzed for differences between samples. Although planned comparisons allowed the results from the two samples to be analyzed independently, an omnibus analysis was performed with the type of sample as a factor. The ANOVA 2 (Inconsistency: Inconsistent vs Control) x 2 (Relevance of symptoms: High vs Low) x Type of sample (Psychology students vs Experienced clinicians) revealed that the target inconsistency x relevance of symptoms x type of sample second-order interaction was not significant regarding the RTs for the target sentence [F (1,32) = .703], only marginally significant for the post-target sentence [F (1,32) = 4.02, MSE = 125526.813, p = .053, η^2 = .11], and not significant for the diagnostic judgment task [F (1,32) = 1.37].

Discussion

The results obtained regarding RTs showed that the participants were able to activate fast online reasoning processes to detect inconsistencies during their fluent reading of clinical reports. Detecting the inconsistencies involved the fast retrieval from memory of knowledge concerning the diagnostic criteria for the different disorders used. In addition, experienced clinicians gave different weights to different symptoms when detecting these inconsistencies during the reading task. The differential weighting of diagnostic criteria may be taken as a departure from the DSM-IV prescriptions, which is consistent with Kim and Ahn's (2002) results. However, in contrast to their study, we suggest that this differential weighting must have originated from fast online reasoning processes given that slow, effortful, and deliberative reasoning processes could not be occurring during fluent reading. This theoretical interpretation is supported by the results of the RTs, which showed a greater inconsistency effect when the target sentence referred to the absence of a highly relevant symptom than when it referred to the absence of a low relevance symptom. Two conclusions can be drawn from these results in relation to how experienced clinicians represent their knowledge concerning mental disorders (see also, Charlin et al., 2007; Schmidt, Norman, &Boshuizen, 1990): a) the status of the DSM-IV diagnostic criteria varies in the clinicians' representations of mental disorders; and b) mental disorders are represented in a way that allows for both the fast and efficient retrieval from memory of domain-specific knowledge and the fast and efficient use of retrieved information for fast inference-making and integration processes.

The question of why experienced clinicians, but not students, gave different weights to the symptoms used in the high-relevance condition than to those used in the low-relevance condition remains open. A plausible explanation is that the symptoms in the high-relevance condition may have a closer correspondence to the clinicians' prototypical representations of mental diseases than the symptoms in the low-relevance condition. This explanation would

be consistent with studies that have shown that clinicians rely on representational heuristics in diagnostic judgment tasks (Maj, 2011; Westen, 2012; Westen & Shedler, 2000). However, in line with Kim and Ahn's (2002) study, it could also be argued that the clinicians' prototypical representations of mental disorders could be based on their causal theories. Thus, symptoms with a high causal role (high-relevance symptoms) would be more prototypical than symptoms with a low causal role (low-relevance symptoms). In this sense, it is noteworthy that the high- and low-relevance symptoms were those that, on average, had the highest and lowest mean centrality scores, respectively, based on causal maps for the different disorders drawn by the participants in Kim and Ahn's (2002) study. In contrast, the students may have relied on less refined prototypical representations of mental disorders, which would have been based on text descriptions (such as DSM-IV), rather than on causal theories or on any real-life exemplars. Reasoning processes based on this type of prototypical representation of mental disorders would make students less likely to demonstrate the differential weighting effect found in clinicians.

Finally, it may be argued that clinicians do not usually make a diagnosis by reading a clinical report alone. Thus, the generalizability of the results may be open to question. In fact, clinicians do far more than simply read clinical reports. However, at the end of an assessment process, they have to review all the material that has usually been compiled in a clinical report or in another written document. Moreover, the reading of clinical reports sometimes precedes the assessment process. In these cases, the assessment process could be guided by the hypotheses that clinicians may generate after reading a clinical report. Finally, there is no reason to think that the comprehension and reasoning processes studied here are not at work when listening to a client in an interview instead of reading a report. Therefore, the same online and semiautomatic processes demonstrated in this experiment are very likely to be present in real clinical contexts.

References

Albrecht, J. E. & O'Brien, E. J. (1993). Updating a mental model: Maintaining both local and global coherence. *Journal of Experimental Psychology: Learning, Memory and Cognition, 19,* 1061-1070. doi: 10.1037//0278-7393.19.5.1061

American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision (DSM-IV-TR)*. Washington, D.C.: American Psychiatric Association.

Black, J. & Bower, G. H. (1980). Story understanding as problem-solving. *Poetics, 9,* 223–50. doi: 10.1016/0304-422X(80)90021-2

Charlin, B., Tardif, J., & Boshuizen, H. P. A. (2000). Scripts and medical diagnostic knowledge: Theory and applications for clinical reasoning instruction and research. *Academic Medicine, 75*, 182–90. doi: 10.1097/00001888-200002000-00020

Charlin, B., Boshuizen, H. P. A., Custers, E. J., & Feltovich, P. J. (2007). Scripts and clinical reasoning. *Medical Education, 41*, 1178-1184. doi: 10.1111/j.1365-2923.2007.02924.x

Davis, R.T., Blashfield, R. K., McElroy, R. A. Jr. (1993). Weighting criteria in the diagnosis of a personality disorder: a demonstration. *Journal of Abnormal Psychology, 102*, 319-22.

Garb, H. N. (1996). The representativeness and past-behavior heuristics in clinical judgment. *Professional Psychology: Research and Practice, 27*, 272-277, doi: 10.1037/0735-7028.27.3.272.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review, 101,* 371-395. doi: 10.1037//0033-295X.101.3.371.

Kahneman, D. (2011). *Thinking, Fast and Slow*. New York: Farrar, Strauss, and Giroux.

Kendell, R. E (1973). Psychiatric diagnoses: A study of how they are made. *British Journal of Psychiatry, 122*, 437-445.

Kendeou, P., Smith, E. R., & O'Brien, E. J. (2012). Updating During Reading Comprehension: Why Causality Matters. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Advance online publication. doi: 10.1037/a0029468

Kim, N. S. & Ahn, W. (2002). Clinical psychologists' theory-based representations of mental disorders predict their diagnostic reasoning and memory. *Journal of Experimental Psychology: General, 131*, 451-476. doi: 10.1037//0096-3445.131.4.451

Long, D. L. & Chong, J. L. (2001). Comprehension skill and global coherence: A paradoxical picture of poor comprehenders' abilities. *Journal of Experimental Psychology: Learning, Memory and Cognition, 27,* 1424-1429. doi: 10.1037//0278-7393.27.6.1424

Long, D., L., Seely, M., R., & Oppy, B., J. (1996). The availability of causal information during reading. *Discourse processes, 22,* 145-170. doi: 10.1080/01638539609544970

Maj, M. (2011). Psychiatric diagnosis: pros and cons of prototypes vs. operational criteria. *World Psychiatry, 10,* 81-82.

McKoon, G. & Ratcliff, R. (1992). Inference during reading. *Psychological Review, 99,* 440-466. doi: 10.1037/0033-295X.99.3.440

Peracchi, K. A. & O'Brien, E. J. (2004). Character profiles and the activation of prototypes vs. operational criteria. *Memory & Cognition, 32,* 1044–1052. doi: 10.3758/BF03196880

Rubinson, E.P., Asnis, G.M., & Friedman, J.M.H. (1988). Knowledge of diagnostic criteria for major depression: A survey of mental health professionals. *Journal of Nervous and Mena Disease, 176*, 480–484.

Sandifer, M.G., Hordern, A., & Green, L.M. (1970). The psychiatric interview: The impact of the first three minutes. *American Journal of Psychiatry, 126*, 968-973.

Schank, R. (1975). *Conceptual Information Processing*. Amsterdam: North-Holland.

Schmidt, H. G., Norman, G. R., & Boshuizen H. P. A. (1990). Cognitive perspective on medical expertise: Theory and implication. *Academic Medicine, 65*, 611–21. doi: 10.1097/00001888-199010000-00001

Smith, E. E. (1989). Concepts and induction. In: M. I. Posner (Ed). *Foundations of Cognitive Science* (pp. 501-526). Cambridge, MA: MIT Press.

Trabasso, T. & Sperry, L. (1985). Casual relatedness and importance of story events. *Journal of Memory and Language, 24*, 595-611.

Trabasso, T. & van den Broek, R. (1985). Casual thinking and the representation of narrative events. *Journal of Memory and Language, 24,* 612-630.

Van Dijk, T, A. & Kintsch, W. (1983). *Strategies of discourse comprehension.* New York: Academic Press.

Westen, D. (2012). Prototype diagnosis of psychiatric syndromes. *World Psychiatry, 11,* 16-21.

Westen D., & Shedler, J. (2000). A prototype matching approach to diagnosing personality disorders toward DSM–V. *Journal of Personality Disorders, 14*, 109–126. doi: 10.1521/pedi.2000.14.2.109

Zwaan, R. A. & Radvansky, G. A. (1998). Situation Models in Language Comprehension and Memory. *Psychological Bulletin, 123,* 162-185. doi: 10.1037/0033-2909.123.2.162

Table 1.

Mean reading times (in milliseconds)and standard deviations for target and post-target sentences, as well as mean diagnostic judgments in the sample of students and experienced clinicians.

	High relevance symptom				Low relevance symptom			
	Inconsistent		Control		Inconsistent		Control	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Students								
Target sentences	2992.83	1025.74	2205.50	630.87	2708.16	1007.04	2233.68	592.93
Post-target sentences	2049.21	642.46	1942.93	768.85	2104.15	623.44	1980.50	952.73
Judgments	54.31	10.96	77.60	10.37	53.63	16.81	72.50	11.89
Experienced clinicians								
Target sentences	3472.15	1385.86	2257.85	562.92	2826.20	1084.93	2273.56	706.01
Post-target sentences	2612.94	840.63	1948.70	704.45	2420.38	662.66	2329.13	824.15
Judgments	36.91	19.05	69.56	23.14679	47.01	22.05	67.25	24.67