CAUSAL REASONING IN THE DIAGNOSIS OF MENTAL DISORDERS: EVIDENCE FROM ON-LINE AND OFF-LINE MEASURES

Doctoral Thesis by

Amanda Flores Martín

Universidad de Málaga

Departamento de Psicología Básica

Under the supervision of

Dr. D. Pedro L. Cobos

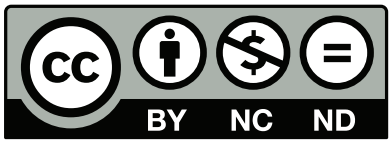
February, 2014

UNIVERSIDAD DE MÁLAGA

Publicaciones y Divulgación Científica

AUTOR: Amanda Flores Martín

EDITA: Publicaciones y Divulgación Científica. Universidad de Málaga



PhD Thesis funded with a FPI scholarship awarded by

JUNTA DE ANDALUCÍA

CONSERJERÍA DE INNOVACIÓN, CIENCIA Y EMPRESA

A mis padres y a Miguel

AGRADECIMIENTOS

En el momento y en el lugar adecuado. Así puedo considerar a día de hoy que comenzó mi vida como investigadora. Llegué en el momento ideal en el que un estupendo grupo de personas podía darme una oportunidad. No puedo dejar pasar una página más sin darles las gracias por confiar en mí. A Paco, por su incalculable ayuda en esta tesis (y en multitud de aspectos). A David y Joaquín, por siempre tener algo que aportar y algo de lo que aprender, casi de cualquier cosa. A Estrella por haberse ofrecido a echarme una mano en más de un momento. Y, muy en especial, a Pedro, un director de tesis que reúne todas las cosas buenas que alguien podría esperar de una persona que te guía. Siempre en una medida perfectamente equilibrada entre lo académico y lo humano. CCG, trabajar con vosotros me ha hecho crecer. Me habéis hecho sentir privilegiada. Un sinfín de gracias por todo (incluidas las risas).

Gracias a Antonio Godoy por haber sido un eslabón esencial en este trabajo, aportándonos en momentos clave la parte más clínica en la que a veces podemos perdernos. A Juan Luís Luque por haber sido un apoyo y una ayuda en diversos momentos. Varias personas del departamento de personalidad aportaron también su granito, destacando a Aurora Gavino, que estuvo ahí siempre que fue necesario. A York Hagmayer, que me acogió en su laboratorio y se esforzó de múltiples y chocolateadas formas en hacerme sentir como en casa. A muchos psicólogos de Málaga y Marbella, a José Díaz, y a muchas otras personas que ofrecieron su tiempo y su interés en participar en mis experimentos, y sin las cuales nada habría ido hacia adelante. A mis compañeros del 118 y a muchos de los que en algún momento tuvieron paso por allí y dejaron su huella, no sólo en forma de *AZ*s. En muchos de vosotros he encontrado amigos. Gracias a todos.

No pocas personas me apoyaron de otras formas y apostaron siempre por mí de manera incondicional. Las personas más importantes de mi vida: mi familia, mis amigos. Me encantaría poder dedicar varias páginas a destacar uno por uno los momentos en los que cada uno de ellos me ha demostrado su cariño y su apoyo de una u otra forma, pero es un deber destacar a mi Andrómeda, grande como la galaxia, que tuvo la absoluta certeza que yo aún no tenía, hace más de 5 años, de que me doctoraría y de que allí estaría ella para verlo. A mis padres que, aunque no tuvieran una idea clara a veces de mi trabajo, siempre estuvieron convencidos de que aquello en lo que empleaba tiempo y esfuerzo en mi vida debía ser indiscutiblemente importante. Mamá, papá, siempre fuisteis, y seréis, mi punto de referencia. Mi faro. Gracias, de corazón.

Dejo para el final a la persona que me recuerda cada día quién soy y cuánto tengo por alcanzar. La persona que se escondió entre los alumnos el primer día de mi vida que impartí una clase, la que estuvo el día que presenté mi tesina, la que escuchó cada ensayo de cada presentación que tenía pendiente e incluso aprendió y me ayudó a mejorarla, y la que hará lo imposible por estar sentado cerca de mí el día que lea esta tesis. En definitiva, la que ha estado ahí insufriblemente. Desde que entraste en mi vida, a todo en ella le has otorgado más sentido. Gracias, Miguelón.

You cannot trace how you came to the belief that there is a lamp on the desk in front of you, or how you detected a hint of irritation in your spouse's voice on the telephone, or how you managed to avoid a threat on the road before you became consciously aware of it. The mental work that produces impressions, intuitions, and many decisions goes on in silence in our mind.

(DANIEL KAHNEMAN)

INDEX

INTRODUCTION

STARTING POINT: THE CAUSAL KNOWLEDGE

Causal knowledge constitutes a highly efficient tool for adaptive behavior. Knowing the causal relationships in which target events are embedded provides a basis from which to infer how those target events will unfold over time in the environment and thus how behavior should adjust accordingly. Causal knowledge entails a reduction in the uncertainty that characterizes the environment in which behavior occurs, allowing for accurate predictions concerning what will happen, why a specific behavior occurred or what the consequences of this behavior will be. The present work further attempted to evaluate the incidence of causal reasoning processes in a specific area: the diagnosis of mental disorders. Specifically, the main objectives were to find compelling evidence of the use of causal reasoning in the diagnosis of mental disorders, and to analyse the nature of these reasoning processes. In order to achieve these objectives, we conducted four main experiments. The first two were conducted with clinicians and students and, the second two, only with students.

CAUSAL REASONING IN THE DIAGNOSIS OF MENTAL DISORDERS

Causal knowledge and reasoning are crucial to make accurate inferences, to choose effective interventions, and to understand the world around us (cf. Waldmann & Hagmayer, 2013). There is substantial amount of evidence for the influence of causal knowledge in many different areas of cognition such as categorization (e.g., Heit, 2000; Rehder, 2010; Waldmann, Holyoak & Fratianne, 1995), conceptual representation (e.g.,

Ahn, Kim, Lassaline & Dennis, 2000; Rehder, 2003; Sloman, Love & Ahn, 1998), decision making (e.g., Hagmayer & Sloman, 2009; García-Retamero & Hoffrage, 2006), text comprehension (e.g., Graesser, Singer & Trabasso, 1994; Trabasso & Sperry, 1985; Trabasso & van den Broek, 1985), interventions (e.g., Sloman & Lagnado, 2005; Steyvers, Tenenbaum, Wagenmakers & Blum, 2003; Waldmann & Hagmayer, 2005) and inference making (e.g., Rehder & Burnett, 2005; Waldmann, 1996; Waldmann & Holyoak, 1992). Although much of the theories explaining causal reasoning have come up from experiments using artificial laboratory settings (see, e.g., Penn & Povinelli, 2007 or Shanks, 2010 for recent reviews of causal learning), there is also evidence that demonstrates its influence in more natural, applied domains, such as in clinical tasks (see, e.g., de Kwaadsteniet, Hagmayer, Krol & Witteman, 2010; de Kwaadsteniet, Kim & Yopchick, 2013, Einhorn, 1986 or Haynes & Williams, 2003; Kim & Keil, 2003; Kim & LoSavio, 2009; Rehder & Kim, 2006; Yopchick and Kim, 2009). However, more applied domain studies are needed in order to test whether the theories built up in artificial laboratory settings are valid in the former domains as well.

In this study, we focused on the diagnosis of mental disorders. On the one hand, one may think that the use of causal reasoning in the diagnosis of mental disorders should not be surprising, as such tasks generally demand cognitive processes that are related to comprehension, categorization, and inference making, for which the influence of causal reasoning has been previously established. For example, the study by Patel and colleagues (Patel & Groen, 1986 and Patel, Evans & Groen, 1989), demonstrates the incidence of causal reasoning in medical diagnostic decisions and how the differences in the nature of the alleged causes of these decisions depend on whether the participants are professional clinicians or researchers. Nevertheless, on the other hand, if we take into account the conception of the classification of mental disorders in the

Diagnostic and Statistical Manual of Mental Disorders (DSM-IV, American Psychiatric Association, 2000), the influence of causal reasoning may seem rather surprising. The taxonomy of mental disorders in this manual is not based on causal considerations. Its classification system is intended to be atheoretical or, at least, neutral with respect to the different theoretical approaches clinicians may adhere to. Classifications are based on diagnostic criteria, most of which are neither necessary nor sufficient. For example, a diagnosis of borderline personality disorder only requires the presence of five out of nine defining features. All possible combinations of these diagnostic features are considered as equivalent for diagnostic purposes. In many cases, the DSM-IV assigns the same weight to all symptoms that are part of the diagnostic criteria for a mental disorder. Therefore, if causal reasoning were shown to be involved in the diagnosis of mental disorders in spite of the atheoretical prescriptions of the DSM-IV, a cognitive bias would be detected. We refer to this cognitive bias as *causal bias*.

Kim & Ahn's (2002) study

Despite the atheoretical nature of the DSM-IV, Kim and Ahn (2002) found that clinicians and Psychology students relied on their idiosyncratic causal theories when asked to make diagnostic judgements. They showed that both clinicians and Psychology students were more likely to apply some diagnostic categories when a hypothetical patient presented some specific symptoms than when she/he presented some others. This result was consistent with previous demonstrations showing that clinicians give a different consideration to different diagnostic criteria of the same disorder (see Davis, Blashfield, and McElroy, 1993; Garb, 1996; Rubinson, Asnis, and Friedman, 1988), leading, in some cases, to a lack of adherence to the DSM.

In their experiments, Kim and Ahn (2002) requested their participants to draw causal maps relating with arrows different symptoms considered as diagnostic criteria for particular disorders according to the DSM-IV. The participants were also allowed to arrange these symptoms in groups if they thought that this was a better method to characterise the relationships between them. The participants were also asked to assign a causal strength to each arrow on a numerical rating scale and then to rate their confidence in their drawings of the causal relationships that were established for each disorder. The objective of this drawing task was to obtain the participants' causal representation about the disorders and a causal rating of each diagnostic criterion. In a second session, 14 days following the drawing task, the participants were presented with hypothetical clinical cases concerning patients who had three causally *central* symptoms, three causally *peripheral* symptoms or three *isolated* symptoms. Central symptoms referred to symptoms that were able to either generate or cause a high number of other symptoms. Peripheral symptoms referred to symptoms caused by other symptoms of the disorder and that did not cause any further symptoms according to the specific clinician's theory of a particular disorder, whereas isolated symptoms were unrelated to any other symptom or diagnostic criteria. These labels (central, peripheral, and isolated) were assigned according to the causal status of symptoms in the participants' own causal drawings. It should be noted that the story cases of all of these hypothetical patients satisfied the diagnostic criteria of the DSM-IV to the same extent, regardless of whether the symptoms were causally central, peripheral or isolated.

After receiving information concerning the symptoms present in hypothetical patients, the participants were required to make a diagnostic judgement indicating the disorder that best fit that set of symptoms. As a result, the participants were more likely to diagnose a hypothetical patient with a specific disorder if that patient had causally

central rather than causally peripheral or isolated symptoms. That is, the participants' causal theories for each mental disorder biased their performance in the diagnostic task. Furthermore, the memory of the symptoms used in the experimental task was also biased by their causal status, such that causally central symptoms were better remembered than peripheral and isolated symptoms. These results are especially relevant considering that clinicians are trained to use the DSM-IV diagnostic criteria without incorporating any additional notions they may have regarding how symptoms relate to each other. And yet, in the experimental series, clinicians' diagnostic and memory task performance appeared to have been biased by such notions.

So far, Kim and Ahn's study is the only evidence of a causal bias in the diagnosis of mental disorders that we are aware of. Nevertheless, there are alternative explanations of their results that cannot be ruled out. In Kim and Ahn's study, the symptoms differing in causal status may also have differed regarding other relevant features, such as their statistical distribution, their conceptual centrality or their diagnostic value. For example, the statistical frequencies of central symptoms may be greater than those of peripheral symptoms in different disorders. In fact, this specific problem was acknowledged by the authors. In our study, an important objective was to find more compelling evidence demonstrating the implication of causal reasoning processes in the diagnosis of mental disorders.

Kim & Ahn's study shows how the diagnosis of mental disorders may be subjected to a causal biases even though clinicians are well trained in the use of the DSM-IV. A possible reason for this bias could be the engagement of less deliberate and controlled processes. If this were the case, any attempt to avoid or control the use of the causal bias should start by detecting such less controlled processes, explaining them and

understanding how they work. This idea is in line with typical dual-processes theories, which have strongly focused on the explanation of biases in reasoning and judgement tasks (Evans, 2008; Tversky & Kahneman, 1977; Sloman, 1996; Stanovich, 1999).

DUAL-PROCESS THEORIES IN REASONING: SYSTEM 1 AND SYSTEM 2

High cognitive processes encompass different processes such as thinking, reasoning, decision making and judgement. These cognitive processes can be partitioned into two main families – traditionally called *intuition* and *reason* – which are now widely embraced under the general framework of *dual-process theories* (Chaiken & Trope, 1999; Hammond, 1996; Sloman, 1996). These theories establish a distinction between processes that are unconscious, rapid, automatic, and high capacity, and those that are conscious, slow, deliberative and capacity-limited (Evans, 2008), or between cognitive operations that are associative and quick and those that are rule-governed and slow (Gilbert, 1999). Kahneman & Frederick (2002) and Stanovich (1999) started to use neutral terms for these two different modes of processing: *System 1* and *System 2* process.

Kahneman & Frederick (2002) used System 1 and System 2 as labels for collections of processes that are distinguished by their speed, controllability, and the contents on which they operate (see Table 1). According to them, System 1 quickly proposes impressions and intuitive answers to judgement problems as they arise, and System 2 monitors the quality of these proposals, which it may endorse, correct, or override. Although System 1 is more primitive than System 2, it is not necessarily less capable. On the contrary, complex cognitive operations eventually migrate from System

2 to System 1 as proficiency and skill are acquired (Kahneman & Frederick, 2002). The roles of the two systems in determining stated judgements depend on the features of the task and on the individual, including the time available for deliberation (Finucane et al., 2000), the respondent's mood (Isen, Nygren, & Ashby, 1988; Bless et al., 1996), intelligence (Stanovich & West, Chapter 24), and exposure to statistical thinking (Nisbett et al., 1983; Agnoli & Krantz, 1989; Agnoli, 1991). Kahneman & Frederick (2002) assume that System 1 and System 2 can be active concurrently, that automatic and controlled cognitive operations compete for the control of overt responses, and that deliberate judgements are likely to remain anchored on initial impressions.

System 1 (Intuitive)	System 2 (Reflective)
Process Characteristics	
Automatic	Controlled
Effortless	Effortful
Associative	Deductive
Rapid, parallel	Slow, serial
Process opaque	Self-aware
Skilled action	Rule application
Content on Which Processes Act	
Affective	Neutral
Causal propensities	Statistics
Concrete, specific	Abstract
Prototypes	Sets

Table 1. Two Cognitive Systems, from Kahneman & Frederick (2002).

According to Kahneman (2011), System 1 is informed by natural drives and instincts but is also capable of learning, which it does by connecting up novel stimuli with known stimuli according to shared characteristics, contiguity in time and place, or causality. System 1 has been shaped by evolution to provide a continuous assessment of the main problems that an organism must solve to survive as quickly as possible, thus allowing us to respond to it immediately. In order to do so, System 1 relies on general rules and guidelines called *heuristics*. These heuristics are primarily geared to help us in the moment and are tilted towards protecting us from danger, and in this respect they are mostly very useful. Still, heuristics can be misleading. For example, the conjunction rule is the most basic qualitative law of probability: the probability of a conjunction cannot exceed the probabilities of its constituents. However, the representativeness and availability heuristics can make a conjunction appears more probable than one of its constituents. Interestingly, this fallacy has also been shown to be the result of the attribution of causal relations between elements (Tversky & Kahneman, 1983).

The impressions that System 1 forms are also fed up to System 2. Indeed, whenever System 1 senses something out of the ordinary or dangerous, System 2 is automatically mobilized to help out with the situation. And even when System 2 is not mobilized specifically out of danger, it is constantly being fed suggestions by System 1. The impressions of System 1 are fairly effective in protecting us from moment to moment. Nevertheless, they are much less effective in long-term planning than System 2. Of course, System 2 is capable of overriding the impressions of System 1, and of avoiding the errors. However, System 2 is often completely unaware that it is being influenced or misled by System 1; and therefore, is not naturally well-equipped to catch the errors (Kahneman, 2011).

SYSTEM 1 AND SYSTEM 2 IN CAUSAL REASONING

Therefore, errors in judgements can be attributed to System 1 and to System 2. System 1 can generate a faulty intuition, which the controlled operations of System 2 fail to detect and correct (Morewedge & Kahneman, 2010) or deliberate judgements are directly anchored on initial impressions (Kahneman & Frederick, 2002). These errors in judgements are frequently based on the use of causal reasoning. Therefore, causal reasoning is not restricted to System 2 processes, but some causal reasoning processes may take place immediately in a fast and partially inadvertent manner (Kahneman, 2011). In fact, it has been shown that the automatic activation and processing of causal information may lead to judgement biases (e.g., Tversky & Kahneman, 1977, 1983; Fugelsang & Thompson, 2003). Thus, causal reasoning may be related not only to System 2 but, also, to System 1 processes.

According to Kahneman (2011), looking for a cause that explains the events that are unfolding over time is a strategy that System 1 uses in order to make sense of the information received. This proclivity is not something that is learned, but is rather innate. The reason why this causal radar has evolved is fairly easy to see. To begin with, cause and effect adheres in nature; as such, it is a good general strategy to assume that a specific cause underlies any given event, and also to seek out and identify it to be better prepared to react. However, many phenomena are better explained in terms of randomness, statistics, or blind luck; and therefore, the assumption of causality that System 1 makes can sometimes lead us into errors.

Regarding the clinical domain, some researchers argue that expert clinicians' knowledge (including causal knowledge) is represented via structurally organized units (scripts) that allow for automatic and efficient access from memory through fast activation processes (cf., Charlin, Boshuizen, Custers, & Feltovich, 2007; Charlin, Tardif, & Boshuizen, 2000; Schmidt, Norman, & Boshuizen, 1990; Smith, 1989). An activation of these representations enables fast inferences and the effective and efficient integration of new incoming information. Previous studies have already provided evidence that clinicians' diagnostic decisions can take a few minutes, with only slight variations in the resulting diagnosis if some more time is spent (Kendell, 1973; Sandifer, Hordern, & Green, 1970).

According to the considerations above, it is tempting to think that the causal bias in the diagnosis of mental disorders may be the result of System 1 causal reasoning processes. But how can we be sure that the bias found by Kim and Ahn is the result of System 1 rather than System 2 processes? Participants in Kim and Ahn's (2002) study had plenty of time to reflect on the permanently available information on diagnostic symptoms, to consider various potential hypotheses and to systematically derive an explicit diagnostic judgement. Hence, participants had ample opportunity to engage in slow, deliberative and resource-demanding processes of thinking. Dual-process models classify such processes as System 2 processes (cf. Evans, 2008; Kahneman, 2011). Nevertheless, causal reasoning may not only be expected to be observed during active decision making or in tasks that involve an explicit, deliberate effort to make clinical decisions or judgements. If System 1 is involved in causal biases, causal reasoning should be expected to operate through rapid and efficient processes in a more automatic manner and without time exclusively dedicated to deliberative thinking in a judgement task. Therefore, to show the implication of System 1 processes in the production of

causal biases in the diagnosis of mental disorders, we should be able to provide evidence of fast and on-line causal reasoning as part of the comprehension processes that take place at the very same time in which reasoners receive information about symptoms.

To sum up, we have some evidence that 1) causal reasoning may bias the diagnosis and mental disorders (Kim & Ahn, 2002), 2) System 1 processes based on causal reasoning can produce judgements biases (Tversky & Kahneman, 1977, 1983; Fugelsang & Thompson, 2003), and 3) fast processes that demand very few cognitive resources to operate can also take place in diagnostic decisions (Kendell, 1973; Sandifer, Hordern, & Green, 1970). Therefore, two main objectives were addressed in this study: 1) to provide more evidence demonstrating that causal reasoning can bias the diagnosis of mental disorders and 2) to provide evidence showing the implication of System 1 in such causal bias. However, we are not aware of previous studies well suited to study the implication of very fast, on-line reasoning processes in diagnostic reasoning. Thus, our aims should imply a different methodological approach.

As said above, if the implication of causal reasoning in diagnostic judgements is mediated by System 1 processes, these processes should be activated very fast, at the right moment in which the information about symptoms is being received. In other words, causal reasoning should take place in an on-line manner, i.e., at the very moment in which relevant information is gathered. Such a demonstration of the rapid and efficient involvement of causal reasoning would serve to further deep into the nature of the processes involved in diagnostic performance. According to some dual-process theories, as mentioned previously, System 2 may endorse the impressions and intuitive answers to judgement problems of System 1, after monitoring the quality of these

proposals. Our methodological approach should demonstrate the involvement of System 1 in these specific processes. Thus, this methodology should allow us to detect fast activation of causal features and inferences to make sense of clinical cases within a coherent mental model. As will be shown, the on-line techniques and procedures used in text comprehension are especially well suited to this aim.

SYSTEM 1 IN TEXT COMPREHENSION

Fast, automatic causal reasoning processes attributable to System 1 also seem to underlie text comprehension. All major accounts of text comprehension assume that readers make on-line inferences during reading (Graesser et al., 1994; McKoon & Ratcliff, 1992; van Dijk & Kintsch, 1983; Zwaan & Radvansky, 1998). One of the purposes of on-line inference making is to create and maintain a coherent representation of a text on both global and local levels. At least some of these inferences are causal (Black & Bower, 1980; Kendeou, Smith, & O'Brien, 2012; Schank, 1975; Trabasso & Sperry, 1985; Trabasso & van den Broek, 1985). Therefore, one may expect to find System 1 causal reasoning when clinicians read clinical reports for later diagnostic judgements. Based on the information provided, relevant domain specific theories or beliefs may be activated and expectations about additional symptoms may be generated. If the subsequent information is coherent with the already received information and the activated domain specific causal information, it can easily be integrated into a mental model and it could be perceived as plausible. By contrast, if it is incoherent, cognitive effort is required to solve the inconsistency and to integrate the new information into a unified structure (Ericsson & Kintsch, 1995; Kintsch, Patel, & Ericsson, 1999). If no solution is provided, the information would be considered implausible. Note that these

inference processes have to operate in a fast and on-line manner whenever reasoners are not provided with extra time to stop and deliberate about the given information.

Morewedge and Kahneman (2010) identified System 1 with the automatic operations of associative memory and claimed that the associative coherence may activate and trace its role in intuitive judgements. Therefore, the computation of causal coherence underlying reading comprehension may rely upon mechanisms as those that have been modeled with dynamic or attractor neural networks. It is well known that the automatic activation of representations in such neural networks tends to produce a comprehensive and internally consistent interpretation of the information provided. It is not a simple coincidence that Hinton (1990) characterised intuitive inferences as the settling into stable states of dynamic neural networks.

Ultimately, as causal reasoning processes attributable to System 1 seem to underlie text comprehension and on-line inferences are made during reading, we proposed a methodology based on reading comprehension. This methodology would be useful to register on-line reasoning processes that depend on System 1.

MAIN OBJECTIVES AND RESEARCH STRATEGY

The objectives of this work were 1) to find stronger evidence of biases due to causal reasoning processes in the diagnosis of mental disorders in experienced clinicians and in people without experience and 2) to assess the implication of System 1 processes in the production of such biases. The second objective reduces to assessing whether causal reasoning takes place through fast and on-line activation and inference processes

as reasoners receive relevant information, and without time specifically dedicated to deliberate thinking.

Research on text comprehension has led researchers to the development of specific experimental paradigms to detect fast, on-line reasoning processes in a non-intrusive way. A particularly interesting experimental paradigm for this purpose is the so-called *inconsistency paradigm*, which has been used within the reading comprehension research field (Albrecht & O'Brien, 1993; Long & Chong, 2001; Peracchi & O'Brien, 2004). According to previous results, reading an inconsistent text (i.e., a text in which the content of sentences is incoherent with each other at first sight) takes longer than reading a consistent (i.e., a text in which coherence is facilitated) or a neutral text. As readers attempt to maintain a coherent representation of the text, finding an inconsistent sentence demands time and cognitive resources to resolve the conflict. Note that the detection of an inconsistency during fluent reading entails the following processes: a) rapid access to domain specific knowledge or theories; b) rapid inference making from the target sentence based on the retrieved knowledge and/or theories; and c) detection of a contradiction between an inference and the information conveyed by a sentence (Long, Seely & Oppy, 1996).

The texts used in this experimental paradigm follow a characteristic structure (Albrecht & O'Brien, 1993; Long & Chong, 2001). First, certain preliminary information regarding a character or an event is provided. Occasionally, this preliminary information is followed by a filler paragraph to clear the content of working memory. This filler paragraph is generally longer than the paragraph where the preliminary information is presented. Next, a target sentence follows, which is inconsistent with the preliminary information; reading times (RTs) for this sentence are expected to be longer

compared with a control condition in which the preliminary sentence is neutral or consistent with the target sentence. The text finishes with a post-target sentence, which is useful to detect possible carryover effects due to the previous inconsistent target sentence (Albrecht & O'Brien, 1993; Long & Chong, 2001; McKoon & Ratcliff, 1992).

The inconsistency paradigm can be used in different ways to detect fast, semiautomatic, and on-line causal reasoning processes during reading of clinical reports. The rationale is very simple. Imagine that a clinical report starts with a series of sentences stating that a hypothetical client has been diagnosed with avoidant personality disorder, and that she/he presents some symptoms that form part of the DSM-IV diagnostic criteria for such disorder. Then, 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 them considered as diagnostic criteria for the disorder. If, according to the reader's previous causal theories about the avoidant personality disorder, the first symptom has a higher causal status than the second one, the sentence stating the absence of the former should raise more conflict than the sentence stating the absence of the latter. Thus, the reading of the more inconsistent target sentence should slow down the reading process more than the reading of the less inconsistent target sentence. Consequently, this paradigm allows for a direct detection of fast and semiautomatic activation of causal features and inference making in an on-line manner during reading. Therefore, the detection of this sort of inconsistency would allow us to infer the implication of System 1 processes on biases due to causal reasoning.

OVERVIEW OF THE EXPERIMENTS

Experiment 1

The specific objectives of Experiment 1 were 1) to replicate Kim and Ahn's (2002) causal status effect and 2) to demonstrate that such causal bias could be the consequence of causal reasoning processes attributable to System 1.

As mentioned before, participants in Kim and Ahn's (2002) experiments (most of them expert clinicians) had to draw a causal map for each disorder, relating the symptoms with arrows. For each map and participant, the causal centrality score of each symptom was calculated according to a specific algorithm. After calculating the causal centrality score for each symptom per disorder per participant, an average score across participants was computed, on the basis of which symptoms were ordered from the most central to the most peripheral (or isolated) one for each disorder. What Kim and Ahn found was that causally central symptoms had a greater impact on participants' diagnostic judgements than peripheral and isolated symptoms despite that, in all cases, the hypothetical patients presented symptoms of equal importance as diagnostic criteria according to the DSM-IV. According to this, we selected the symptom with the highest mean centrality score and the symptom with the lowest mean centrality score within each disorder to increase the difference in causal status. We used these selected symptoms to create clinical reports.

We tried to detect the causal bias in two different ways: through an on-line technique based on the inconsistency paradigm (see the previous section), and through an off-line measure based on diagnostic judgements. We expected to find a greater inconsistency effect associated with the use of the central, compared with the use of the

peripheral symptom, in the target sentence. At the same time, we expected to find a greater impact of the central symptom, compared with the peripheral one, on diagnostic judgements.

Experiment 2

In Experiment 2, we assessed the implication of causal reasoning by manipulating the temporal order of the symptoms, which is a fundamental defining feature of causal relationships, as well as by providing explicit information about causal connections between symptoms. This manipulation allowed us to overcome some limitations of Experiment 1 and to provide stronger evidence of causal reasoning. Also, as in Experiment 1, we tested whether our manipulation could bias participants' diagnostic reasoning, and assessed the implication of causal reasoning processes attributable to System 1 on such bias.

We created clinical reports that could be either consistent or inconsistent with participants' causal theories of different disorders. The clinical reports included target sentences providing information about the temporal sequence of the symptoms and a final sentence explicitly informing about causal links between them. In the consistent condition, both the temporal order of the reported symptoms and the causal links between them were consistent with a causal theory for the disorder that the participants were supposed to entertain. In the inconsistent condition, the temporal order was reversed and no causal link between the symptoms was said to have been found. Thus, the consistent condition was

causally plausible according to the theory, whereas the inconsistent condition was not.

As in Experiment 1, RTs for target sentences in both conditions were compared, and so were the diagnostic judgements. We expected to observe longer RTs in the inconsistent than in the consistent condition. At the same time, we expected both the information regarding the temporal sequence of symptoms and the information regarding causal connections between the symptoms to influence participants' diagnostic judgements.

Experiment 3

The specific objective of Experiment 3 was to find further evidence that the computation of causal coherence is at the core of comprehension processes dependent on System 1 during reading as well as of diagnostic judgements. This objective was addressed by assessing the impact of sentences stating the absence of one of three possible symptoms that formed part of a causal chain on RTs and on diagnostic judgements.

We presented participants with clinical reports about hypothetical patients based on disorders described in the DSM-IV. Each clinical report began by providing information about the diagnosis received by the patient. Later on, participants read three sentences providing information regarding the presence or absence of three symptoms that are considered diagnostic criteria for the disorder according to the DSM-IV. Such symptoms formed part of a causal chain (S1→S2→S3) that was not explicitly stated in the clinical report but was supposed to form part of the participants' causal theory for

the disorder. We created inconsistencies by explicitly stating the absence of one of the three symptoms in one of the sentences. With this manipulation, two types of inconsistencies were created: *categorical*, between the diagnosis and the absent symptoms, and *causal,* between the presence of some symptoms and the absence of their causal antecedents in the causal chain connecting the symptoms. From these two types of inconsistency, we could derive some predictions according to different causal theories related to coherence and causal status.

As in Experiment 1 and 2, RTs for target sentences were compared, and so were the diagnostic judgements. Specific predictions concerning RTs and diagnostic judgements can be derived from different reasoning theories. We will elaborate more on these predictions later.

Experiment 4

The specific objectives of Experiment 4 were 1) to test whether diagnostic reasoning is sensitive to the manipulation of causal mechanisms connecting the symptoms, and 2) to assess whether such sensitivity to causal mechanisms can be traced back to System 1 causal reasoning during reading.

Experiment 4 focused on fast and on-line reasoning processes based on the participants' assumptions regarding causal mechanisms. We assessed whether informing participants explicitly about the mechanisms connecting the symptoms of a disorder affected on-line causal reasoning and later diagnostic judgements. In this experiment, all of the clinical reports informed that a given patient presented with three symptoms being diagnostic criteria of a disorder. According to clinical background

knowledge, the three symptoms formed a causal chain (S1→S2→S3). The only difference concerned the causal mechanisms leading from one symptom to the next. In one condition, the causal mechanisms were consistent with the causal theory for the disorder mentioned at the beginning of the clinical report. In the other condition, alternative plausible causal mechanisms were described, which were inconsistent with the causal theory of the disorder.

As in all previous experiments, RTs for target sentences were compared, and so were the diagnostic judgements. We expected participants to be sensitive to the causal mechanisms and, therefore, to detect mechanistic inconsistencies. Hence longer RTs were expected in those cases in which the mechanisms did not conform to the causal theory of the disorder. We also expected the information about mechanisms to affect final diagnostic judgements despite keeping the diagnostic criteria unaltered.

EXPERIMENTS

EXPERIMENT 1

The objective of this experiment was twofold. On the one hand, we tried to replicate Kim and Ahn's (2002) causal bias in diagnostic judgements based on the causal status effect. On the other hand, we searched for evidence consistent with the idea that such bias could be the consequence of causal reasoning processes reliant on System 1. This second aim was addressed by demonstrating participants' engagement in fast, on-line causal reasoning processes during reading of clinical reports. We relied on the inconsistency paradigm to address this goal. Although our primary interest was to assess clinicians' causal reasoning processes, we were also interested in examining students'.

The strategy used relied upon the manipulation of two variables. First, we manipulated the consistency of the clinical reports by building up inconsistent and control clinical reports. The clinical reports consisted of sentences that provided information regarding a hypothetical patient. At the beginning of each clinical report, there was a sentence providing information about the diagnosis received by the patient from a clinical psychologist. Such a diagnosis was a specific DSM-IV disorder. This preliminary information was potentially inconsistent with a target sentence that was located near the end of the clinical report. In the inconsistent condition, the target sentence made a statement that contradicted one of the symptoms considered as a diagnostic criterion (according to DSM-IV) for the disorder that was mentioned in the preliminary information. For example, if the diagnosis stated in the preliminary information was avoidant personality disorder, the target sentence could state that the patient was convinced of being interesting, competent and appealing, which is just the

opposite of "*she/he is convinced of being inferior, unappealing, or inept*", the latter being a DSM-IV diagnostic criterion for the disorder. In the control condition, the same target sentence appeared in a clinical report in which the diagnosed disorder had no relationship with the contradicted symptom. For example, the target sentence of the example could appear in a clinical report for a patient who was diagnosed with a sleepwalking disorder. Only in the inconsistent condition are participants expected to detect an inconsistency between the target sentence and the preliminary information. Therefore, the RT for the target sentence should be longer in the inconsistent than in the control condition.

Second, we manipulated the causal status of the contradicted symptom by using a causally central and a causally peripheral symptom. In the Procedure section, we describe the procedure that was followed to determine the causal status of the symptoms. In the central-cause condition, the target sentence contradicted a causally central symptom; in the peripheral-cause condition, the sentence contradicted a causally peripheral symptom. If participants cannot help using their causal theories of the diagnosed disorder, they should perceive a stronger inconsistency in the central-cause than in the peripheral-cause condition. Thus, the difference in RT for the target sentence between the inconsistent and the control conditions should be greater in the central-cause than in the peripheral-cause condition.

Method

Participants and apparatus

A total of 34 participants took part in the experiment on a voluntary basis. Half of them were undergraduate Psychology students from Malaga University (Spain) and the other half were experienced clinicians who worked in independent practice in Málaga area. Their experience as clinicians ranged from three to 28 years (average 17 years). Two of the clinicians used a cognitive approach, one used a systemic approach, one used a psychoanalytic approach, three used multiple approaches, and the remaining participants used a cognitive-behavioral approach.

Materials and design

Overall, a total of 24 clinical reports were created: Twelve of them for the inconsistent condition and twelve for the control condition. The reports from the inconsistent condition were referred to six different DSM-IV disorders, namely, major depressive disorder, specific phobia, antisocial personality disorder, schizophrenia, borderline personality disorder, and avoidant personality disorder (as in Kim and Ahn' study); whereas those from the control condition were referred to 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 could be read, though in a clinical report in which the diagnosed disorder bore no relationship with this absent symptom. Therefore, only in the

inconsistent but not in the control condition were participants expected to detect an inconsistency and hence, RTs for the target sentence should be longer in the inconsistent than in the control condition. Additionally, the target sentences from half of the reports in the inconsistent condition were referred to central-cause symptoms whereas the other half referred to peripheral-cause symptoms. We expected a greater inconsistency effect (i.e., a greater difference in RTs between the target inconsistent and control sentences) associated to the absence of a central-cause than of a peripheral-cause symptom.

To manipulate the causal status of the symptoms, we selected those with the highest and lowest mean centrality scores in Kim and Ahn's (2002) study. Of course, it is sensible to expect individual differences regarding the causal theories entertained by Kim and Ahn's participants and by our participants. As a consequence, symptoms that were considered as causally central by some of their participants might be considered as causally peripheral by some of ours. However, according to Kim and Ahn's results, there are reasons not to expect huge divergence between idiosyncratic causal theories. In fact, Kim and Ahn found important similarities between causal maps even when comparing expert clinicians and students. These similarities were especially apparent when comparing causal maps for familiar disorders. Taking into account that we used familiar disorders in our study, the causal theories entertained by our participants and by those in Kim and Ahn (2002) are expected to have remarkable commonalities. Thus, by selecting the symptoms with the highest and the lowest average score in causal centrality for each disorder in Kim and Ahn's study, we maximized the possibility that the former would be consider as more central than the latter by our participants. Another alternative would be to have our participants drawing causal maps as in Kim and Ahn's experiments. However, we preferred to avoid this alternative procedure because any

evidence of causal reasoning processes could be attributed to having artificially made the participants make their causal theories explicit. In other words, performing the causal-map drawing task would entail a strong effort and a good amount of time allotted, which might artificially induce the use of causal reasoning.

All clinical reports were made up following the structure of 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 across the different reports. After an introductory sentence, participants could read the DSM-IV diagnosis that the hypothetical patient 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 irrelevant information. The three symptoms were selected from those that received intermediate average ratings of causal centrality in Kim and Ahn's (2002) study. Then, participants could read two sentences stating the presence of two highly frequent symptoms (i.e., present in numerous DSM-IV disorders). Right before including the target sentence, four filler sentences referring to non-clinical information could be read. Note that this filler information would make previous information regarding clinical symptoms unavailable from participant's working memory by the time the target information is read. And finally, the last two sentences appeared in the text: The target and the post-target sentence. An example of target sentence (referred to the avoidant personality disorder) in the central-cause condition could be: "*she is convinced of being interesting, competent and appealing*", which contradicts the criterion "*views self as socially inept, personally unappealing, or inferior to others*" (from DSM-IV-TR) (the most central symptom in Kim & Ahn, 2002). In the peripheral-cause condition, the sentence that could be read was "*she becomes easily involved in*

new activities", which contradicts "*avoids personal risk or new activities*" (from DSM-IV-TR) (the most peripheral symptom in Kim & Ahn, 2002). The additional post-target sentence provided clinically irrelevant information and was included to detect any possible carryover effect that could have been produced by the reading of the target sentence (see Appendix A and B).

Procedure

The experimental task was performed on PCs in a laboratory that was equipped with ten semi-isolated cubicles to prevent participants from visual contact. Nevertheless, although the task was performed individually, participants were assembled in groups that could range from 5 to 10 individuals. The sample of clinicians ran the experiment individually in their consulting rooms. The experimenter went to some length to ensure that the participants were not interrupted by phone calls or by individuals entering the office or knocking on the door. All PCs were equipped with home-built software programmed in Visual Basic 2005 (Microsoft, USA).

In the first session, the participants started reading the instructions on the computer screen and all participants' doubts were solved before the experimental task began. The students were asked to imagine that they were clinical psychologists and had to make some decisions about several patients. All participants were instructed to read the material attentively and, at the same time, fluently. After reading each clinical report, they would be required to judge the extent to which they agreed with the diagnosis received by the patient. Note that this way, the use of clinical reasoning was promoted during the reading task.

After reading the instructions, all participants were presented with an example of a clinical report based on a disorder (generalized anxiety disorder) that was different from those that were used in the actual experimental task. The example text consisted of 16 sentences with a structure similar to the experimental clinical reports. All of the texts were displayed within a 14 x 14 cm text box using a Courier New 12 point font. The whole text appeared initially unreadable and the reading task was self-paced. Initially, every letter of the text was substituted by a mask that consisted of a forward slash. Each bar press made all of the letters of a sentence visible while hiding the slashes. A second bar press had the reverse effect on the previously read sentence and turned the following sentence visible. The RT for each sentence was the time that elapsed between the two consecutive bar presses. As usual in self-pace reading tasks, the readers were not allowed to go back during the reading. Pressing the space bar after reading the final sentence made the text disappear, and a rectangular box at the center of the screen was displayed that contained a scale below the message, "*The diagnosis received by the client was generalized anxiety disorder. Please rate the extent to which you agree with the diagnosis using the scale below*" (translated from Spanish). The participants made their ratings using a horizontal scroll bar that was displayed below the message. Within a small text box on the right of the scroll bar, the participants could see a numeric representation of the location of the scroll-bar face. The ratings could range from 0 to 100, indicating complete disagreement and complete agreement, respectively. No feedback was provided. After completing the example of clinical report, the participants could ask questions to solve any doubt regarding the task.

The experimental task took place in two sessions, separated at least by one week. Participants read 12 different clinical reports in each session. The assignment of the different texts per session ensured that participants could not read twice the same target

and post-target sentences within the same session. The reading order of the different texts within each session was randomized. A counterbalanced procedure ensured that for each session, half of the clinical reports were from the inconsistent and half from the control condition. Orthogonally to this, half of the clinical reports were from the central-cause and half from the peripheral-cause condition. Each session took between 20 to 30 minutes to complete.

Results

Our aim was to evaluate whether the participants engaged in fast, on-line causal reasoning processes during reading of clinical reports, and whether such reasoning processes were consistent with diagnostic judgements in a later diagnostic task. Specifically, we assessed whether RTs for target sentences in clinical reports and later diagnostic judgements varied depending on whether these reports offered either consistent or inconsistent information regarding causal theories of the disorders involved. Finally, we addressed these objectives in a sample of undergraduate Psychology students and in a sample of experienced clinicians.

Reading times. The analyses were carried out on RTs for both, the target and post-target sentences. All statistical analyses reported in this study used an α of .05. These measures 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 computed, giving four averaged measures for the target and another four for the post-target sentences. Overall, only 12 and 10 RT measures were withdrawn from the target and post-target sentences, respectively.

Table 2 shows mean RTs for the target and post-target sentences in each condition within each sample. As observed, the students' and the clinicians' RTs for the target sentence were longer in the inconsistent than in the control condition, which is consistent with an inconsistency effect. Additionally, the difference between the inconsistent and control condition appeared to be greater in the central-cause than in the peripheral-cause condition in the case of clinicians, but not in the case of students. Separate analyses were performed for each sample to confirm these impressions. A repeated measures ANOVA 2 (Inconsistency: Inconsistent vs. control) x 2 (Causal Status: Central-cause vs. peripheral-cause), on the students' RTs yielded a significant main effect of Inconsistency [F (1, 16) = 24.091, MSE = 271235.080; $p < .001$; η^2 = .56]. None of the remaining 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 on RTs for the post-target sentence yielded no significant effect (all F values < 1.27). Regarding the clinicians' sample, the same ANOVA on RTs for the target sentences yielded a significant effect of Inconsistency [F (1, 16) = 12.801, MSE = 1036503.4, p = .003, η^2 = .44], Causal Status [F (1, 16) = 9.043, MSE = 186673.901, p = .008, η^2 = .36], and Inconsistency x Causal Status [F (1, 16) = 6.505, MSE = 286012.319, p = .021, η^2 = .289]. Simple effects analyses revealed a significant inconsistency effect in both conditions of the Causal Status factor, F (1, 16) = 14.203, MSE = 882461.693, p = .002, η^2 = .47; F (1, 16) = 5.899, MSE = 440053.978, p = .027, η^2 = .27 for the central-cause and the peripheral-cause condition, respectively. Table 2 also reveals similar results for RTs for post-target sentences, i.e., a greater effect of inconsistency in the central-cause than in the peripheral-cause condition. This impression was confirmed by the same ANOVA, which yielded a significant main effect of Inconsistency [F (1, 16) = 5.565, MSE = 297226.601, p = .031, η^2 = .26], and a

marginally significant effect of the Inconsistency x Causal Status interaction [F (1, 16) = 3.691, MSE = 125731.452, p = .073, η^2 = .19]. The main effect of Causal Status was not significant (F < 0.66). Planned tests for simple effects yielded an inconsistency effect within the central-cause condition [F (1, 16) = 5.873, MSE = 329524.764, p = .028, η^2 = .27], but not within the peripheral-cause condition [F (1, 16) = 1.958].

These results indicate that, during reading, both students and clinicians engaged in some form of fast and on-line clinical reasoning that entailed the retrieval and use of DSM-IV diagnostic criteria for the mental disorders used, as the inconsistency effect was found in both. However, in the case of students, we could not find any sort of modulation by the causal status of symptoms in on-line reasoning processes, as no Inconsistency x Causal Status interaction was found. Conversely, in the case of clinicians, the on-line reasoning processes seemed to depart from the DSM-IV's prescriptions, as the inconsistency effect was modulated by the causal status of symptoms despite that all symptoms were equivalent diagnostic criteria. This result suggests that clinicians engaged in on-line, fast, and semiautomatic causal reasoning. Therefore, clinicians' causal reasoning during reading seemed to depend on System 1 processes.

Diagnostic judgements. A single mean diagnostic judgement (i.e., judgement of agreement with the diagnosis provided) per participant was calculated for each experimental condition within each sample (see Table 2). In general, participants agreed more on the diagnosis stated in the preliminary information in the control than in the inconsistent condition, indicating previous familiarity with diagnostic criteria from the DSM-IV. However, the difference between the means was greater in the central-cause than in the peripheral-cause condition in the clinicians' but not in the students' sample.

These impressions were confirmed by the statistical analyses. A repeated measures ANOVA 2 (Inconsistency: Inconsistent vs. control) x 2 (Causal Status: Central-cause vs. peripheral-cause) on the students' judgements yielded a significant main effect of Inconsistency [F (1, 16) = 39.754, MSE = 189.997; $p < .001$; $\eta^2 = .71$]. Neither the effect of Causal Status nor the interaction between the two factors, were statistically significant (all F values < 3.16). Regarding the clinicians' sample, the same ANOVA yielded a significant effect of Inconsistency [F (1, 16) = 46.896, MSE = 247.928, $p <$ .001, $\eta^2 = .75$], and of Inconsistency x Causal Status [F (1, 16) = 5.586, MSE = 128.387, $p = .031$; $\eta^2 = .26$]. The effect of Causal Status did not reach significance [F (1, 16) = 1.431]. The inconsistency effect was greater in the central-cause than in the peripheral-cause condition. Simple effects analyses revealed that the effect was nevertheless 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 central-cause and the peripheral-cause conditions, respectively.

		Central-cause symptom				Peripheral-cause symptom			
		Inconsistent		Control		Inconsistent		Control	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Psychology 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
	Diagnostic judgements	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
	Diagnostic judgements	36.91	19.05	69.56	23.15	47.01	22.05	67.25	24.67

Table 2. Means and standard deviations of reading times (in milliseconds) for target and post-target sentences, as well as means and standard deviations of diagnostic judgements, in the sample of students and experienced clinicians.

These results were consistent with those that were identified in the RT analysis and suggest that the reasoning processes that occurred during reading could also be responsible for the effects that were observed in diagnostic judgements. Hence we have at least indirect evidence that System 1 causal reasoning processes may have affected diagnostic judgements. Specifically, the greater impact of central-cause symptoms on clinicians' diagnostic judgements could have been determined by a greater impact of such symptoms during on-line and fast reasoning processes during reading. The results also revealed that those symptoms that were considered as central-cause and peripheral-

cause by Kim and Ahn's (2002) participants were also considered differently by our sample of experienced clinicians.

Finally, we also explored whether statistically different results were obtained in the two samples of participants, Psychology students and experienced clinicians. Though planned comparisons allowed us to analyse the results in the two samples independently, we nevertheless considered an omnibus analysis with the type of sample as a factor. The ANOVAS 2 (Inconsistency: Inconsistent vs. control) x 2 (Causal Status: Central-cause vs. peripheral-cause) x Type of sample (Psychology students vs. experienced clinicians) revealed that the target Inconsistency x Causal Status x Type of sample second order interaction was not significant in the RTs for the target sentence [F (1,32) = .703], marginally significant for the post-target sentence [F (1,32) = 4.02, MSE = 125526.813, p = .053, η^2 = .11] and not significant for diagnostic judgements [F (1,32) = 1.37].

Discussion

The pattern of results obtained regarding RTs showed that our participants engaged in fast and on-line reasoning processes during their fluent reading of clinical reports. This detection involved fast knowledge retrieval from memory concerning the diagnostic criteria for the different disorders, and fast inferential and integration processes. In addition, experienced clinicians' detection of inconsistencies was affected by the causal status of the symptoms. This causal status effect may be taken as a departure from DSM-IV's prescriptions, and suggests that clinicians engaged in causal reasoning processes, which is consistent with Kim and Ahn's (2002) results. At

variance, though, with Kim and Ahn's study, we have shown that these causal reasoning processes have taken place in a very fast and on-line manner, as slow, effortful, deliberate reasoning processes could not be at work during fluent reading. This theoretical interpretation is supported by RT results that showed a greater inconsistency effect when the target sentence referred to the absence of a central-cause symptom than when it referred to the absence of a peripheral-cause symptom. The fact that clinicians' fast and on-line reasoning processes were biased by their causal theories suggest that the causal reasoning processes that have been at work have the properties that have been attributed to System 1 processes.

At odds with experienced clinicians, Psychology students did not show any differential weighting effect, either during the reading task or the diagnostic judgement task. In other words, they did not ponder differently symptoms varying in causal status in any of the tasks. This result, however, may have, at least, two different interpretations. On the one hand, students may have been more inclined to follow DSM-IV's prescriptions, treating all diagnostic criteria in a similar way. On the other hand, the absence of a differential weighting effect may have been related to the actual symptoms used as central-cause and peripheral-cause symptoms so that the former were not effectively perceived as more causally central than the latter. Should other central-cause and peripheral-cause symptoms be used, the modulating effect might be obtained even with a sample of students. This, in turn, may also be a consequence of the students' lack of clinical experience. Unfortunately, the present experiment does not allow us to distinguish between these two alternative interpretations.

One limitation of our experiment is that it does not allow us to discard alternative interpretations of the results from the clinicians. For example, one may claim

that symptoms in the central-cause condition were also more frequent than symptoms in the peripheral-cause condition from the point of view of the clinicians' professional experience. Also, the clinicians might consider that central-cause symptoms are more defining features of the disorders (or more conceptually central) than peripheral-cause symptoms. This idea would be consistent with studies showing clinicians' reliance on the representational heuristic in diagnostic judgements (Maj, 2011; Westen, 2012; Westen & Shedler, 2000). This same limitation has also been acknowledged by Kim and Ahn regarding their own study (see the General Discussion section in Kim & Ahn, 2002). However, in our case, this limitation may raise further concern as we did not directly tested the clinicians' idiosyncratic causal theories for the different mental disorders in an independent task.

EXPERIMENT 2

As in Experiment 1, our general aim was to show the implication of System 1 processes in causal reasoning when reasoners are provided with information about mental disorders and diagnostic criteria from the DSM-IV. One concern regarding previous demonstrations of causal reasoning in the diagnosis of mental disorders is that they have been based on manipulations of the presence or absence of symptoms (Kim & Ahn, 2002). However, causal reasoning should also be tapped by providing information with clear implications about causal connections between symptoms without altering the presence or absence of such symptoms. After all, clinical reports are much more than mere lists of symptoms from the DSM-IV. They frequently include additional information such as the temporal order in which symptoms develop, or statements making explicit the causal connections between symptoms inferred from the clinical assessment process. Imagine, for example, that a clinician is provided with information about a client who has been previously diagnosed with a specific disorder. If the clinician has a causal theory of the disorder, she/he would expect some symptoms to have occurred according to a specific temporal sequence as a consequence of the specific causal connections between such symptoms. Consequently, additional information consistent or inconsistent with the expected temporal order and causal connections should have an impact both on on-line causal reasoning processes, and on the extent to which the clinician agrees on the diagnosis received by the client, despite that neither the temporal order of symptoms nor the causal connections between them form part of the diagnostic criteria established by the

DSM-IV. The manipulation of the causal information provided through clinical reports without altering the symptoms suffered by the hypothetical clients has several interesting advantages. First, as this manipulation is not based on variations in the symptoms, we can avoid any confound between the causal role of the symptoms and their weights in the diagnostic process. Note that such weights may not necessarily (or exclusively) be based on causal theories. Second, this approach allows us to know the impact of causal information that goes beyond the diagnostic criteria. A strict application of the DSM-IV criteria and prescriptions should lead clinicians to ignore those aspects that are not considered as diagnostic criteria. Therefore, an effect of the inclusion of causal information on on-line and off-line causal reasoning while holding the symptoms constant may contribute to find compelling evidence of the use of causal theories in the diagnosis of mental disorders.

Therefore, our main objective in Experiment 2 was twofold. On the one hand, our aim was to test whether diagnostic reasoning and diagnostic judgements are biased by aspects that go beyond the causal status of present or absent symptoms. Specifically, we assessed whether diagnosticians are influenced by the temporal order of the symptoms, which is a fundamental defining feature of causal relationships, as well as by explicit information about causal connections between symptoms. On the other hand, we assessed whether such bias could be due to fast, on-line causal reasoning processes attributable to System 1.

Then, we created clinical reports that could be either consistent or inconsistent with participants' causal theories of the disorder mentioned in a preliminary sentence (see below). The target sentences provided information about the temporal sequence of three

symptoms, which could be either consistent or inconsistent with participants' causal theories. Assume, for example, that, according to a clinician's causal theory of Disorder X, Symptom S1 causes S2, which in turn causes S3. A clinical report stating that a client who is diagnosed with Disorder X developed S1 followed by S2, and then S3 would be consistent with the clinician's expectations based on her/his causal theory. Conversely, a clinical report stating that the client first developed S3, then S2, and then S1 would be inconsistent with the clinician's causal theory. In such a case, clinicians should spend more time reading the clinical report to solve the inconsistency. Therefore, we expected to observe longer RTs in the inconsistent than in the consistent condition. At the same time, we expected both the information regarding the temporal sequence of symptoms and the information regarding causal connections between the symptoms to influence participants' judgements of agreement with the diagnosis. Thus, these results would provide converging evidence for the use of causal theories both from the RT measures and from participants' diagnostic judgements, as in Experiment 1.

After the reading and the diagnostic judgement tasks, participants carried out a treatment efficacy judgement task. Judgements in this task were analysed to check whether participants' causal beliefs were in accordance with the causal chain model (i.e., S1→S2→S3) on which we based our manipulation and predictions. For every clinical report, the participants were required to judge the efficacy of three different treatments for removing each symptom. Treatments T1, T2, and T3 were thought to have a direct removal effect on symptoms S1, S2, and S3, respectively. Interventions have an interesting consequence that only holds when the variables are linked within a causal structure and provided that people reason according to a rational approach to causal reasoning (Hagmayer et al., 2007; Meder, Hagmayer & Waldmann, 2008; Pearl, 2000; Sloman & Lagnado, 2005). For example, a direct intervention on a specific symptom not

only has its effects on this target symptom (i.e., a direct effect) but also on other symptoms that are causally connected with it (i.e., an indirect effect). In an S1→S2→S3 causal chain, for instance, a direct intervention on S2 would be equivalent to removing the S1→S2 causal link. This effect occurs because intervening on a variable renders it independent of its causes but not of its effects. Thus, a direct intervention removing S2 would have no consequences on the probability of S1 (a so-called backward effect) but would still vary the probability of S3 (a so-called forward effect). In other words, if participants assume the S1→S2→S3 causal chain, they would conclude that the removal of a symptom would also have consequences *down* the causal chain (i.e., removing the effects of the intervened symptom) but not *up* the chain (i.e., not altering the cause of the intervened symptom). This asymmetry should tend to disappear if no clear causal model links the different symptoms (Meder et al., 2008; Sloman & Lagnado, 2005). Thus, a distinctive pattern of treatment efficacy judgements should be found if participants' causal knowledge of the disorders was in agreement with the causal theory underlying our experimental manipulation of the consistency factor.

First, if participants understood the task correctly and had basic knowledge about the treatments, T1, T2, and T3 should receive the highest effectiveness ratings for the removal of symptoms S1, S2, and S3, respectively. That is, direct effects (e.g., for example, the effect of T1 on S1) should have a greater impact than indirect effects (e.g., for example, the effect of T1 on S2).

Second, indirect forward effects should receive higher ratings than indirect backward effects. For example, the efficacy of T1 to remove S3 should receive higher ratings than the efficacy of T3 to remove S1.

Third, differences between forward and backward effects should be modulated by the causal information provided by the clinical reports. Specifically, the difference between forward and backward effects should tend to disappear in the inconsistent condition in which no causal connection exists among the symptoms. Note that, for example, if S1 is thought to be a causal antecedent of S3, then removing the former with treatment T1 should contribute to the removal of the latter. In contrast, T3 would not be effective in removing S1 because, in general, removing an effect (S3) of a given symptom leaves its causes unaltered. This logic would no longer apply if no causal connection exists between S1 and S3. In such a case, the effectiveness of T1 and T3 to remove S3 and S1, respectively, should tend to be similarly viewed.

Fourth, the effect of causal information should only be evidenced in forward effects. Efficacy judgements for forward effects – i.e., T1 on S2 (T1-S2 hereafter), T2-S3, etc. should be higher in the consistent causal than in the inconsistent non-causal condition. This is because indirect forward effects are expected only to the extent that the symptoms are causally connected. On the other hand, no effects of causal information should be observed in backward indirect effects –i.e., T3-S2, T2-S1, etc. In this case, for the reasons just explained, low efficacy ratings were expected in both, the consistent causal and the inconsistent non-causal conditions.

To sum up, an important objective of this experiment was to evaluate whether the influence of causal reasoning processes may also be evidenced by altering the temporal order in which a set of symptoms (i.e., diagnostic criteria) of a DSM-IV mental disorder is expected to occur according to a causal theory for such disorder. Another objective was a) to evaluate whether participants' performance in a judgement diagnostic task was equivalent to that obtained in a reading task, and b) to evaluate participants' causal theory

through a treatment efficacy judgement task. And finally, we were interested in addressing these objectives in a sample of undergraduate Psychology students and in a sample of experienced clinicians.

Method

Participants and apparatus

A total of 101 participants took part in the experiment. Seventy-one Psychology undergraduate students from the University of Malaga volunteered to take part in the experiment in exchange for course credits. The sample of experienced clinicians included thirty clinical psychologists from private and public institutions who worked in Malaga and volunteered to participate in the experiment. The main theoretical orientation in their professional practice was: 16 cognitive-behavioral clinicians, three psychoanalysts, one humanist, one gestaltist, and nine who used multiple approaches. Their experience as clinicians ranged from three to 30 years and averaged 10 years.

Materials and design

A total of 12 clinical reports were created to manipulate the causal consistency of the clinical reports. Accordingly, there were six consistent causal and six inconsistent non-causal clinical reports. We tested the effect of this manipulation on the participants' RTs for the target sentences and on the participants' judgements of agreement – i.e., judgements of the extent to which they agreed with the diagnosis that was stated in the preliminary information of the reports. Each clinical report included three different

symptoms considered as diagnostic criteria (according to the DSM-IV) for the diagnosed disorder.

Clinical reports were referred to patients who were diagnosed with one of six possible mental disorders: anorexia nervosa, major depression, specific phobia, obsessive-compulsive disorder, posttraumatic stress disorder, and generalized anxiety disorder. There were two clinical reports per disorder: One for the consistent causal condition, and one for the inconsistent non-causal condition. These specific disorders were selected as they had a relatively high prevalence in the general population according to the DSM-IV (ranging from 0.5% of anorexia nervosa to 8% of posttraumatic stress disorder) and, additionally, because there are specific psychological theories that establish causal connections among the symptoms considered as diagnostic criteria in the DSM-IV (Beck, 1967, 1985; Crisp, 1980; Ladouceur, 1998, Mowrer, 1947, Salkovskis, 1985). For example, according to Crisp's (1980) model of anorexia nervosa, an important cause of the development of the different symptoms is a feared situation, such as a strong fear of gaining weight (i.e., S1). This fear causes a refusal to maintain a minimal body weight (i.e., S2), an effect that is potentially evident in several overt behaviors, such as a strict diet, vomiting, and laxative abuse. These behaviors in turn cause weight loss and eventually a deterioration that may alter menstruation in women, producing amenorrhea (i.e., S3). In the example, symptoms S1, S2, and S3 are DSM-IV diagnostic criteria. Importantly, according to this theory, these diagnostic criteria should appear in a specific temporal order: First S1, then S2, and finally S3.

All clinical reports followed the same structure (see Appendix C and D). After a first introductory sentence, participants could read the diagnosis made by a clinician. The next sentences introduced the symptoms as they were verbalized by the patients. Each

verbalization suggested the development of one symptom and provided additional information regarding the moment in which the symptom appeared. For the consistent causal condition, the temporal order of the symptoms was consistent with the causal theory of the disorder (i.e., S1, S2, and S3). For the inconsistent non-causal condition, the temporal order of the symptoms was reversed (i.e., S3, S2, and S1). These were the target sentences for which RTs were registered. Following these target sentences, a final sentence was included that provided explicit information regarding the causal connections between the symptoms. In the consistent causal condition, the sentence stated that the three symptoms were causally related whereas in the inconsistent non-causal condition it stated that no relationship could be established among them.

We used patients' verbalizations to open up the possibility that the interpretation of these verbalizations as symptoms could be guided by causal theories. This way, if a verbalization suggesting a symptom appears in an inconsistent non-causal clinical report, participants may be more cautious against inferring this symptom. Consequently, the manipulation of causal information may have an effect on participants' tendency to infer the presence of symptoms, thereby increasing the effect of causal information on diagnostic performance.

In order to evaluate whether causal reasoning modulated participants' diagnostic performance, it is crucial to find independent evidence showing that our participants actually had causal beliefs about the disorders involved in the clinical reports consistent with the causal theories that served to define our experimental manipulation. This evidence may be taken as a manipulation check. For this, we set up a task in which participants had to judge the efficacy of three different treatments, T1, T2, and T3 for the

removal of each of three different symptoms, S1, S2, and S3, that is, a total of nine efficacy judgements per disorder (see Appendix E).

Procedure

The experimental task was performed in the same laboratory than Experiment 1 and the sample of clinicians ran the experiment in their consulting rooms. Again, the experiment took place in two sessions, separated by at least a week. All participants read the instructions on the computer screen. The instructions included an example of the different tasks participants had to carry out. First, they made a careful and fluent reading of a clinical report referred to a hypothetical patient. After that, participants were required to make a diagnostic judgement task and finally, they also had to rate the efficacy of three different treatments. Additionally, we also registered how long participants took to make each of these different judgements. Overall, we expected that judgements in the inconsistent non-causal condition would take longer than in the consistent causal condition.

The reading task was self-paced, as in Experiment 1. Once the whole text had been read, the diagnostic judgement task started. The clinical report was again displayed at the top of the screen, so that the participants could re-read it at any time. At the center of the screen, a message prompted our participants to judge the extent to which they agreed with the diagnosis stated in the report. Below this message, the participants could see a horizontal scrollbar that could be manipulated to make their estimations from 0 (i.e., "*Completely sure that the correct diagnosis is different*") to 100 (i.e., "*Completely sure that the clinician indicated the correct diagnosis*"). A small text box just below the

scrollbar allowed participants to observe a numerical translation of the different positions of the scrollbar.

Once the diagnostic judgement had been made, the treatment efficacy judgement task started. Again, the clinical report was displayed at the top of the screen. The participants had to rate the efficacy of three different treatments for each of the three symptoms referred to in the report. For example: "*To what extent do you think that a progesterone-based hormonal treatment will resolve, in the short-, medium-, or long-term, the following problems?*" Note that in the example, the treatment mentioned is thought to be aimed at the amenorrhea symptom. The order in which the efficacy of treatments T1, T2, and T3 had to be judged was counterbalanced across participants. Then, a list including the three statements made by the patient, each suggesting the presence of one symptom, appeared below this message in a random order. At the right of each statement, a scrollbar with a small text box below were shown. The participants could use the scrollbars to estimate how efficient the treatment was, whereas the text box automatically provided the participants with a numeric translation (from 0 to 100) of their estimation. Once the judgement was made, the participants proceeded to evaluate the efficacy of the other two treatments. The participants could revise their ratings before ending the task.

Once the participants made their judgements, they could read the next clinical report and then carry out the corresponding diagnostic and treatment efficacy judgement tasks. For each of the two experimental sessions programmed, one set of six clinical reports was set up based on the six possible mental disorders that were described above. Half of the clinical reports in each set were assigned to the consistent causal and half to the inconsistent non-causal condition. Each set of clinical reports was used in a different

session, the order of the sets being counterbalanced across participants. This way, two clinical reports based on the same disorder were never read during the same session. The order of clinical reports within each session was randomised across participants.

We also measured how long participants took to make each of the judgements requested (i.e., the diagnostic and the treatment efficacy judgements). We expected judgements in the inconsistent non-causal condition to take longer than in the consistent causal condition. In the former case, participants would be expected to spend more time and resources attempting to make sense of a clinical report that was inconsistent with the entertained causal theories. This effect should only occur as far as the participants relied upon causal reasoning processes to make their judgements and decisions.

Results

Our aim was to evaluate whether the participants engaged in fast, on-line reasoning processes during reading of clinical reports, and whether such reasoning processes were consistent with diagnostic judgements in a later diagnostic task. Specifically, we assessed whether RTs for target sentences in clinical reports and later diagnostic judgements varied depending on whether these reports offered either consistent causal or inconsistent non-causal information regarding causal theories of the disorders involved. Finally, we addressed these objectives in a sample of undergraduate Psychology students and in a sample of experienced clinicians.

Sample of Psychology students

Reading times. The same filtering method of Experiment 1 was used. There were only two outlier RTs, each from a different participant. Given that the target sentences were of different lengths, the RTs were normalized to the number of sentence letters (see Table 3). A repeated measures ANOVA 2 (Degree of Consistency: Consistent vs. inconsistent) x 3 [Causal Hierarchy: S1 (high), S2 (medium), S3 (low)] was performed on the normalized RTs, yielding a significant effect of Causal Hierarchy [F (2, 140) = 12.81, MSE = 40.33; $p < .001$; η^2=.16] and a significant Degree of Consistency x Causal Hierarchy interaction [F (2, 140) = 6.32, MSE= 50.93; p = .002; η^2= .08]. The main effect of Degree of Consistency was only marginally significant [F (1, 70) = 3.35, p = .072; η^2= .05].

Due to the significant interaction, simple effects were analysed. The Degree of Consistency effect was only significant within S3, F (1, 70) = 10.78, MSE= 67.92; p = .002; η^2= .13 (remaining F values < 1.65). Thus, the RTs for the target sentence that suggested the presence of S3 (the symptom with the lowest causal status) were significantly longer in the inconsistent non-causal than in the consistent causal condition. Recall that in the inconsistent non-causal condition, S3 was the first reported symptom as well as the first symptom that was experienced by the client. In the consistent causal condition, S3 was the final reported symptom and the final symptom that was experienced by the client. The observed effect could be interpreted as an effect of temporal order, which would be consistent with what would be expected if the participants had been reasoning according to causal theories during reading. However, the absence of a significant effect on RTs for the target sentences that suggested the presence of Symptoms S1 and S2 suggests an alternative interpretation. As the

participants were not clinicians and had not taken any course on psychopathology or psychological treatment, they may have been unaware that the S3 symptoms were diagnostic criteria for the disorders that were mentioned in the preliminary information. Examples of symptoms playing the role of S3 were amenorrhea (for anorexia nervosa), weight alterations (for major depression), or tiring easily (for generalized anxiety disorder). When these symptoms are stated in the first sentence, naïve participants may spend more time reading given that such symptoms are rather unexpected. However, reading the same sentences after the sentences that reported symptoms S1 and S2, as in the consistent causal condition, did not have the same impact given that such symptoms may have been much more expected and consequently may have provided sufficient information to make sense of the reported case. Consequently, given that no main effect of Degree of Consistency nor any simple Degree of Consistency effect were found within S1 or S2, the analyses of RTs do not provide convincing evidence of causal reasoning during reading in the sample of students.

Psychology students												
	Consistent causal condition						Inconsistent non-causal condition					
	S1 (High)		S2 (Medium)		S3 (Low)		S1 (High)		S2 (Medium)		S3 (Low)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
RT target sentences	42.54	11.94	44.71	13.83	42.95	14.61	41.05	13.88	47.45	16.73	46.83	13.86
	M				*SD*			*M*			*SD*	
Diagnostic Judgements	75.00				11.56			58.97			14.96	
Decision time for diagnostic judgements	11801.62				5004.51			14266.16			5841.43	
Decision time for treatment efficacy judgements	112747.90				44533.17			102393.40			33604.03	
Experienced clinicians												
	Consistent causal condition						Inconsistent non-causal condition					
	S1 (High)		S2 (Medium)		S3 (Low)		S1 (High)		S2 (Medium)		S3 (Low)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
RT for Target Sentences	54.69	20.77	56.92	19.33	53.57	16.97	62.89	25.63	62.92	22.86	61.28	17.82
	M				*SD*			*M*			*SD*	
Diagnostic Judgements	70.61				14.35			60.47			11.01	
Decision time for diagnostic judgements	23690.92				10680.38			31423.11			15466.74	
Decision time for treatment efficacy judgements	137554.50				49794.77			173191.10			84050.94	

Table 3. Means and standard deviations of reading times for the target sentences (in milliseconds, normalized per number of letters) across the different causal hierarchy conditions; means and standard deviations of times (in milliseconds) spent in diagnostic judgements; and means and standard deviations of diagnostic judgements (in a 0 to 100 rating scale) in the sample of students and in the sample of clinicians.

Diagnostic judgements. If students rely on causal theories to make diagnostic judgements, we should observe higher ratings of agreement in the consistent causal than in the inconsistent non-causal condition. That is, students should agree to a greater extent with the diagnosis provided in the report if the temporal order of symptoms and the information regarding causal connectivity are consistent rather than inconsistent with the supposedly entertained causal theory of the diagnosed disorder.

Students' judgements were collapsed across clinical reports into a single judgement per condition per participant (see Table 3 for mean judgements in each condition). The results revealed that students' ratings in the consistent causal condition were higher than in the inconsistent non-causal condition. This impression was confirmed using a paired t-test, which yielded a robust significant effect: t (70) = 8.96, p < .001, η^2= .53. Thus, students agreed to a greater extent with the diagnosis provided when the causal information was consistent rather than inconsistent with the causal theory.

We also analysed the time spent making the diagnostic judgements (see Table 3). Two outlier cases (more than 3 *Sds* away from the mean), each from a different participant, were excluded from the analysis. Consistently with the results obtained in the diagnostic judgements, participants took more time in the inconsistent non-causal than in the consistent causal condition. This impression was confirmed using a paired t-test: t (70) = -4.76, p < .001, η^2 = .24.

Treatment efficacy judgement task: Evaluating causal theories. The results from the treatment efficacy judgement task served us to have independent evidence regarding the causal theory that participants held concerning the disorders involved in the clinical reports. Specifically, we were interested in assessing whether the participants' causal theories were consistent with the S1→S2→S3 causal chain model on which we

based our manipulation of causal information. For this, the participants' efficacy judgements were considered in five different conditions: 1) The direct effect condition (e.g., the efficacy of T2 to remove S2), 2) the forward short distance effect condition (e.g., the efficacy of T1 to remove S2), 3) the forward long distance effect condition (the efficacy of T1 to remove S3), 4) the backward short distance effect condition (e.g., the efficacy of T3 to remove S2) and 5) the backward long distance effect condition (the efficacy of T3 to remove S1). Prior to the analyses, the judgements were averaged within these different five conditions and then averaged across the six clinical reports. For example, in the direct effect condition, an average rating was calculated from ratings for T1-S1 (i.e., the efficacy of T1 for removing S1), T2-S2, and T3-S3. This mean was calculated for each clinical report, and the means from the six clinical reports were collapsed into a single average. Figure 1A shows the participants' mean ratings in each condition. The time spent making these treatment efficacy judgements was also analysed. Below, we report the results of the analyses that were conducted to test the different hypotheses.

The first analyses were conducted to test whether the participants attributed to Treatments T1, T2, and T3 specific effects to remove Symptoms S1, S2, and S3. As can be seen in Figure 1A, ratings in the direct condition were, in general, higher than in the remaining conditions. Prior to performing the specific planned comparisons, a global ANOVA was performed with Degree of Consistency (consistent vs. inconsistent) and Treatment Effect (direct vs. forward short distance vs. forward long distance vs. backward short distance vs. backward long distance) as within-subjects factors and the participants' average ratings as the dependent variable. The analysis yielded the significant effect of Degree of Consistency, F (1, 70) = 9.93, MSE = 104.05; p = .002; η^2= .12, Treatment Effect, F (4, 280) = 228.01, MSE = 192.11; p < .001; η^2= .77, and

Degree of Consistency x Treatment Effect, F (4, 280) = 8.62, MSE = 41.24; $p < .001$; $\eta^2 = .11$. Despite the observed interaction effect, to test our first hypothesis, we performed the analyses after collapsing the ratings across the levels of Degree of Consistency. This was done given that the differences between the ratings in the direct level and each of the remaining levels were significant in both of the Degree of Consistency conditions. Unsurprisingly, the participants' ratings in the direct condition were higher than in the forward short distance condition, t (70) = 11.776, $p < .001$, $\eta^2 = .66$, the forward long distance condition, t (70) = 13.197, $p < .001$, $\eta^2 = .71$, the backward short distance condition, t (70) = 16.411, $p < .001$, $\eta^2 = .79$, and the backward long distance condition, t (70) = 20.917, $p < .001$, $\eta^2 = .86$. This result indicates that the participants clearly perceived that the treatments were especially efficient to remove the target symptom for which they had been devised.

According to our second prediction, if the participants' causal theories conformed to the S1→S2→S3 causal chain model, indirect forward effects should receive higher ratings than indirect backward effects. This should be specially the case in the consistent condition. As can be seen in Figure 1A, ratings in the forward conditions were higher than in the backward conditions. To simplify the analyses, we collapsed the means from the forward short distance and the forward long distance condition into a single mean, and so we did with the backward short and the backward long distance conditions. Then, we conducted a repeated measures ANOVA 2 (Degree of Consistency: Consistent vs. inconsistent) x 2 (Treatment Effect: Forward vs. backward) on the participants' mean ratings, which yielded the significant effect of Degree of Consistency, F (1, 70) = 9.69, MSE = 41.92; $p = .003$; $\eta^2 = .12$, Treatment Effect, F (1, 70) = 53.78, MSE = 58.42; $p < .001$; $\eta^2 = .43$, and Degree of Consistency x Treatment Effect, F (1, 70) = 5.67, MSE =

34.91; $p = .020$; $\eta^2 = .08$. The main effect of Treatment Effect confirms the impressions suggested by Figure 1A and provide evidence supporting the second prediction.

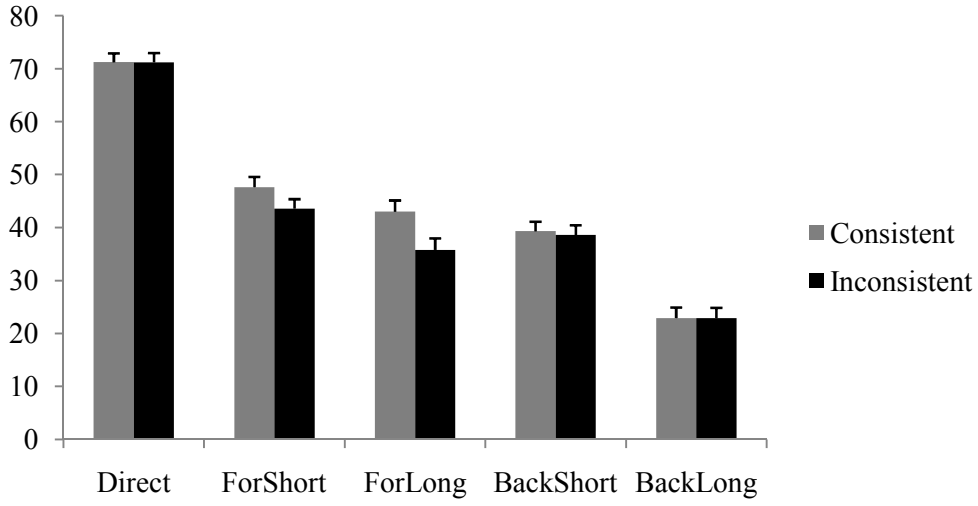
Our third prediction stated that the difference between the forward and the backward effect conditions should be greater in the consistent causal than in the inconsistent non-causal condition. In fact, this is just the impression suggested by Figure 1A. This impression, in turn, is supported by the significant Degree of Consistency x Treatment Effect interaction reported above. However, to strengthen the case for the third prediction, we directly compared the mean difference between the forward and the backward effects in the consistent causal condition against the corresponding mean difference in the inconsistent non-causal condition. As expected, the difference in the former case was significantly greater than in the latter case, $t\ (70) = 3.62$, $p = .001$, $\eta^2 = .16$. It is worth mentioning, however, that ratings for the forward indirect effects were significantly higher than ratings for the backward indirect effects even in the inconsistent non-causal condition, $t\ (70) = 4.40$, $p < .001$, $\eta^2 = .21$. This result suggests that the participants tended to adhere to an S1→S2→S3 causal model despite having received disconfirming information. This tendency may have been induced by two factors. First, the inconsistent non-causal clinical reports did not provide information that allowed the participants to build an alternative causal model with which to make sense of the clinical case. Second, the clinical reports were very brief and could make the participants believe that they lacked a good amount of information. These factors together may have led participants to discredit the clinical report to some extent. As a result, in many cases, the participants may have preferred to rely on their causal theory for the disorder to solve the efficacy judgement task.

Finally, we tested our fourth prediction namely that the consistency effect should only be evidenced in forward effects. Specifically, ratings in the consistent causal condition should be higher than in the inconsistent non-causal condition when considering the forward effect conditions. No difference was expected in the remaining Treatment Effect conditions. The impression suggested by Figure 1A is quite consistent with this prediction. The analyses of simple effects of Degree of Consistency within the different levels of Treatment Effects corroborated this impression. The Degree of Consistency factor was significant within the forward short distance condition, t (70) = 3.973, $p < .001$, $\eta^2 = .18$, and within the forward long distance condition, t (70) = 3.858, $p < .001$, $\eta^2 = .17$. None of the remaining simple effects were significant (all t's < 0.682, and all p's > .49).

We also analysed the time the participants spent in the treatment efficacy judgement task. In this case, if the participants had been engaged in causal reasoning, they might have taken longer to give their efficacy judgements in the inconsistent non-causal than in the consistent causal condition. In the former case, the absence of information regarding causal links between the symptoms may have made participants uncertain regarding the indirect effects of treatments, and thus, participants should have taken longer to make their judgements. Table 3 gives the mean time spent in the consistent causal and in the inconsistent non-causal conditions. As can be seen, a paired t-test yielded no significant effect [t (70) = 1.152, p = .253]. Thus, the time that the participants took to make their treatment efficacy judgements did not appear to reflect the use of causal reasoning. Nevertheless, given the effects found in the treatment efficacy judgements, the absence of significant effects on the time spent very likely reflects a lack of sensitivity of the dependent measure rather than the absence of causal reasoning.

The results found in students provide partial evidence of causal reasoning processes. The participants' diagnostic and treatment judgements were consistent with causal reasoning based on both the causal information that was provided in the clinical reports and on previous causal theories, which were shown to be consistent with the S1→ S2 → S3 causal chain model. However, when considering on-line dependent measures of reasoning processes, we did not observe consistent evidence of causal reasoning. Specifically, the RTs were not consistently affected by the temporal sequence of symptoms. This result cannot be explained by claiming that students do not possess causal theories of the disorders used in the clinical reports. The pattern of results found in the treatment efficacy judgement task clearly suggest that students do possess causal theories of the disorders and that such causal theories are quite consistent with the causal chain model on which we based our manipulation of the temporal sequence of symptoms. There are several possible explanations for the absence of effects on RTs. One explanation is that RTs may not be sensitive enough to reflect the effects of the manipulation of the temporal sequence of symptoms. Another explanation is that students may have performed the reading task in a rather passive manner, which, in turn, could have lead them not to spend enough time and resources to solve the inconsistencies detected. Finally, students may not be familiar enough with the causal theories on which we based our manipulation. As a consequence, the information provided about symptoms may not have produced fast, on-line activation of their causal features, which is an important requisite for System 1 processes to get involved in causal reasoning. Unfortunately, our data do not allow us to distinguish which is the correct explanation.

Panel A



Panel B

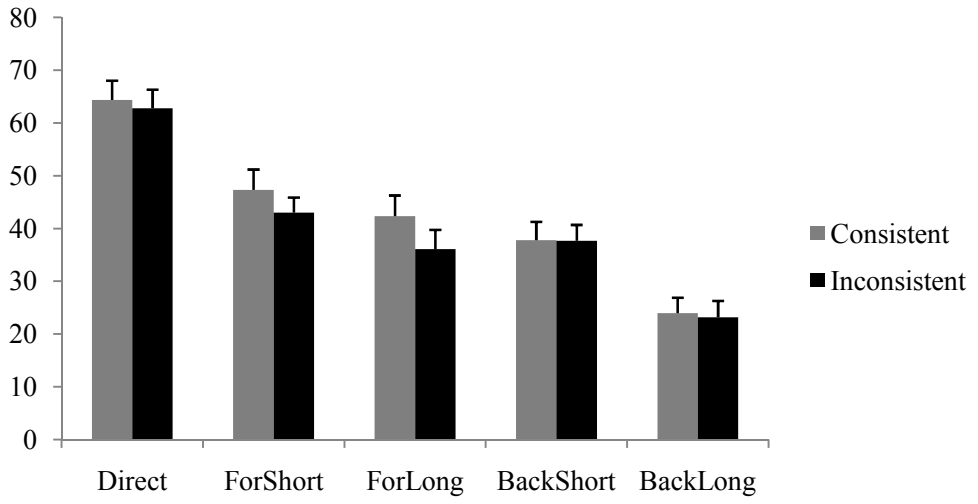


Figure 1. Mean treatment efficacy judgements across different conditions (Direct effects, Forward Long Distance effects, Forward Short Distance effects, Backward Long Distance effects and Backward Short Distance effects), in consistent causal and inconsistent non-causal reports, in the sample of students (Panel A) and in the sample of experienced clinicians (Panel B).

Sample of experienced clinicians

Reading times. As with students, RTs were filtered and normalized to the number of letters per sentence. Only two RTs from one participant were withdrawn from the analysis. Table 3 gives the mean normalized RTs per condition. As expected, the mean RTs were consistently longer in the inconsistent non-causal than in the consistent causal condition. This result was confirmed by a repeated measures ANOVA 2 (Degree of Consistency: Consistent vs. inconsistent) x 3 [Causal Hierarchy: S1 (high), S2 (medium), S3 (low)] performed on the normalized RTs, which revealed a significant effect of Degree of Consistency, F (1, 29) = 10.60, MSE =226.71, p = .003, η^2=.27. Neither the effect of Causal Hierarchy nor the Degree of Consistency x Causal Hierarchy interaction were significant (all F values < .546). Consequently, the clinicians' RTs were altered by on-line reasoning processes as a consequence of the manipulation of the temporal sequence of symptoms. This result suggests that clinicians engaged in on-line causal reasoning processes during reading of clinical reports. In other words, clinicians could engage in fast retrieval of causal theories consistent with the S1→S2→S3 causal model as well as in fast inference processes from such causal model. This result is quite consistent with the idea that causal reasoning processes are based on System 1 when clinical information is processed during reading.

Diagnostic judgements. As with the students, clinicians' judgements were collapsed across clinical reports into a single judgement. An inspection of the mean judgements of agreement shown in Table 3 reveals that, as expected, the clinicians' ratings in the consistent causal condition were higher than in the inconsistent non-causal condition. This impression was confirmed by a paired t-test: t (29) = 4.27, p < .001, η^2=

.39. Consistent with this result, the clinicians took longer to make their diagnostic judgements in the inconsistent non-causal than in the consistent causal condition (see Table 3): t (29) = -3.83, p = .001, η^2=.33.

The consistency between the results from RTs and from diagnostic judgements provides indirect evidence that System 1 causal reasoning processes may have affected diagnostic judgements. However, we cannot discard that, besides the System 1 processes detected during the reading task, the clinicians may also have engaged in System 2 processes during the diagnostic task that could bias their diagnostic judgements.

Treatment efficacy judgement task: Evaluating causal theories. Ratings in this task were analysed to assess whether the clinicians relied on causal theories consistent with the S1→S2→S3 causal chain model. As in the case of the students' ratings, a single mean rating was calculated for each Treatment Effect condition within each level of Causal Information. Figure 1B shows the mean ratings in each condition.

We first started by assessing whether treatments T1, T2, and T3 were attributed some specificity regarding symptoms S1, S2, and S3, respectively. An inspection of Figure 1B clearly suggests that ratings for direct effects were higher than for the remaining treatment effects. A repeated measures ANOVA with 2 (Degree of Consistency: Consistent vs. inconsistent) x 5 (Treatment Effect: Direct vs. forward short distance vs. forward long distance vs. backward short distance vs. backward long distance) as within-subjects factors performed on the participants' average ratings yielded the significant effect of Degree of Consistency, F (1, 29) = 7.172, MSE = 71.95, p = .012, η^2= .20, Treatment Effect, F (4, 116) = 77.28, MSE = 163.36, p < .001, η^2= .72, and a Degree of Consistency x Treatment Effect interaction, F (4, 116) = 4.73, MSE = 21.35, p = .001, η^2= .14. To assess the differences between ratings in the direct condition and the

remaining conditions of Treatment Effects, we first collapsed the ratings across both consistency conditions. Again, the participants' ratings in the direct condition were higher than in the forward short distance condition, t (29) = 6.146, $p < .001$, $\eta^2 = .56$, the forward long distance condition, t (29) = 7.176, $p < .001$, $\eta^2 = .64$, the backward short distance condition, t (29) = 9.767, $p < .001$, $\eta^2 = .77$, and the backward long distance condition, t (29) = 12.138, $p < .001$, $\eta^2 = .83$. This result indicates that clinicians perceived that the treatments were especially efficient to remove the target symptom for which they had been devised.

To test the second prediction, we assessed whether ratings in the forward conditions were higher than in the backward conditions. An inspection to Figure 1B suggests that the data are consistent with the prediction. As explained above, we collapsed the means from the forward short distance and the forward long distance conditions into a single mean, and so we did with the backward short and the backward long distance conditions. Then, we conducted a repeated measures ANOVA 2 (Degree of Consistency: Consistent vs. inconsistent) x 2 (Treatment Effect: Forward vs. backward) on participants' mean ratings, which yielded the significant effect of Degree of Consistency, F (1, 29) = 6.58, MSE = 32.42, p = .016, $\eta^2 = .19$, Treatment Effect, F (1, 29) = 65.28, MSE = 47.67, $p < .001$, $\eta^2 = .69$, and a Degree of Consistency x Treatment Effect interaction, F (1, 29) = 12.52, MSE = 12.84, p = .001, $\eta^2 = .30$. The significant main effect of Treatment Effect provides evidence consistent with the second prediction.

An inspection of Figure 1B also reveals that the difference between the forward and the backward effect conditions was greater in the consistent causal than in the inconsistent non-causal condition, which is in agreement with our third prediction. This impression was supported by the significant Degree of Consistency x Treatment Effect

interaction reported above. However, as we did for the sample of students, we directly compared the mean difference between the forward and the backward effects in the consistent causal condition against the corresponding mean difference in the inconsistent non-causal condition. The resulting difference was greater in the consistent causal than in the inconsistent non-causal condition, t (29) = 3.55, p = .001, η^2= .30. As with students, the asymmetry observed between the forward and backward inferences indicates that clinicians were reasoning according to an S1→S2→S3 causal model and that their reliance on such causal model was less pronounced in the inconsistent non-causal than in the consistent causal condition. However, the ratings for the forward indirect effects were significantly higher than the ratings for the backward indirect effects even in the inconsistent non-causal condition, t (29) = 7.22, p < .001, η^2= .64. As with students, this result suggests that the clinicians tended to adhere to an S1→S2→S3 causal model despite having received disconfirming information.

Additionally, we assessed whether the effect of Degree of Consistency was significant in the forward indirect effect conditions only. An inspection of Figure 1B confirms this fourth prediction. The greatest difference between the consistent causal and the inconsistent non-causal conditions were observed in the forward indirect effects. The analyses of the simple effects revealed that the Degree of Consistency factor was significant within the forward short distance level, t (29) = 3.619, p = .001, η^2= .31, and within the forward long distance level, t (29) = 3.274, p = .003, η^2= .27. In both cases, the ratings were higher in the consistent causal condition than in the inconsistent non-causal condition. None of the remaining simple effects were significant (all t values < 1.68).

Finally, the analysis of the time spent in the efficacy judgement task revealed that clinicians were significantly slower in the inconsistent non-causal than in the consistent

causal condition (see Table 3), t (29) = -2.394, p = .023, η^2= .16. This result suggests that in the absence of consistent causal information, clinicians appeared to spend time attempting to determine the causal mechanism that explained the symptoms to judge the efficacy of treatments on symptoms.

Overall, the results provide consistent support for the use of causal theories in rapid and efficient reasoning processes during reading of clinical reports referred to DSM-IV disorders, and when making diagnostic and treatment judgements. Furthermore, the results obtained in the treatment efficacy judgement task showed that clinicians' causal theories of the disorders were consistent with the causal chain model on which our manipulation was based. This result supports the hypothesis that the clinicians' use of causal theories is not limited to the processing of information regarding what symptoms are present or absent. These theories also appear to be used to process all the relevant information regarding the causal structure underlying the different symptoms.

Discussion

In this experiment, the results found in students showed that their diagnostic judgements were affected by information about the temporal sequence of symptoms together with information about causal connections between them. Specifically, students agreed on the diagnosis received by the hypothetical patients to a greater extent when the temporal sequence of symptoms and the causal connections between them were consistent with their causal theories of the diagnosed disorders than when the information provided was inconsistent with such theories. Also, students spent more time in the diagnostic judgement in the inconsistent non-causal than in the consistent causal

condition. However, we did not find convincing evidence of fast and on-line causal reasoning processes in this sample, as the time spent reading the target sentences conveying information about the temporal sequence of symptoms seemed to be unaffected by whether the temporal order was consistent or inconsistent with the students' causal theory of the disorder mentioned in the preliminary information of the clinical report.

The results found with experienced clinicians also showed a greater agreement on the diagnosis received by the hypothetical patient in the consistent causal condition than in the inconsistent non-causal condition. Clinicians also took longer in the diagnostic judgement task in the latter than in the former condition, suggesting that they were trying to solve the causal inconsistencies found in the clinical report before making the judgement. Additionally, we found evidence of fast and on-line causal reasoning, as evidenced by the clinicians' RTs for the target sentences. Specifically, RTs were significantly longer when the information was inconsistent than when it was consistent with the clinicians' causal theories of the disorder with which the hypothetical patient had been diagnosed.

A key issue of this experiment is to show that the effects found are due to the consistency or inconsistency of the clinical reports with participants' causal theories. In other words, to show that participants' causal theories or believes conformed to the causal chain model (S1→S2→S3) that served as a basis for our manipulation of consistency. The treatment efficacy judgement task developed served this purpose. In this task, participants had to judge the efficacy of three different treatments to remove each of the three symptoms suffered by the patient. Each treatment (T1, T2, and T3) was devise to have a specific direct effect on one of the symptoms (S1, S2, and S3, respectively). The

results found in this treatment efficacy judgement task provided compelling evidence that both students' and clinicians' causal theories were consistent with the causal chain model we assumed for the consistency manipulation. First, we found that judgements for forward indirect effects (indirect effects down the causal chain; e.g., the effect of Treatment 2 on Symptom 3) were higher than judgements for backward indirect effects (indirect effects up the causal chain; e.g., the effect of Treatment 2 on Symptom 1). This difference persisted even in the inconsistent non-causal condition despite the fact that the information provided through the clinical reports was inconsistent with the causal chain model. As expected, however, this difference was larger in the consistent causal than in the inconsistent non-causal condition. Second, the manipulation of the degree of consistency affected participants' judgements of treatment efficacy only within the forward indirect conditions. Judgements in the direct and backward effect conditions did not differ as a function of the consistency condition. This pattern of results is what should be expected if participants assumed a causal chain model of the form S1→S2→S3, and provided that their causal reasoning followed a rational approach. Consequently, the results suggest that both the students' and the clinicians' causal theories of the disorders used in our clinical reports were consistent with this specific causal model, even when the information provided through the clinical reports was inconsistent with such model and discouraged participants to engage in causal reasoning.

It is important to note that, when a specific domain is concerned, the detection of causal inconsistencies through fluent reading not only requires the possession of causal theories relevant within such domain but, also, that such causal theories are represented so as to allow for a fast and efficient access and use. The acquisition of these special representations, in turn, is not likely to occur if domain-specific causal theories are not predominantly used to understand and make sense of events within the domain.

Therefore, the results corresponding to the sample of clinicians suggest that the use of causal theories in the diagnosis of mental disorders is not something rare that only occurs in artificial environments as a result of experimental manipulations. Rather, such use of causal theories seems to be part of reasoning process that takes place in natural environments.

A possible criticism regarding our interpretation of the effect of inconsistency on clinicians' RTs is that, although causal relationships between the symptoms necessarily involve a specific temporal order, the latter does not necessarily imply that the symptoms are causally connected. Longer RTs in the inconsistent non-causal condition could be due to the low frequency of clinical cases in which the different disorders develop according to the inconsistent temporal sequence. Thus, in their experience, clinicians may have encountered more cases in which the disorders develop according to the consistent temporal order than cases in which the development conforms to the inconsistent temporal order. According to this account, although clinicians encounter many cases in which symptoms develop in the consistent order, they would nonetheless remain uncommitted to any interpretation regarding how symptoms are causally related. Therefore, the inconsistency effect would be due to clinicians' previous knowledge of temporal precedence completely free of any causal interpretation. Although this alternative explanation cannot be completely ruled out, it is not very convincing. As said above, the results that were observed in the inconsistent non-causal condition in the treatment efficacy judgement task strongly suggest that the clinicians hold causal theories for the different disorders according to which S1 would be a causal antecedent of S2, and S2 would be a causal antecedent of S3. Moreover, it appears that the clinicians were somewhat reluctant to avoid using such theories despite having received a) information regarding the temporal order of symptoms that contradicted their theories and b) explicit

information regarding causal connections discouraging the participants from engaging in causal reasoning. Given this strong tendency to assume the existence of causal links between symptoms, it is more likely and parsimonious to think that such causal theories played an important role in explaining the impact of the temporal order of symptoms on RTs.

A slightly different explanation of the inconsistency effect on RTs would be that clinicians may prefer to read and write down symptoms of mental disorders in the consistent order (i.e., S1-S2-S3). Thus, the participants may have expected to receive information regarding the symptoms in the consistent rather than in the inconsistent order. According to this hypothesis, the inconsistency effect would be a consequence of the order in which the symptoms are listed in the text rather than the order in which the symptoms appeared in the patient. The problem with this account is that we would lack an explanation of why clinicians' preferences coincide with the causal model that they appear to assume. Given that we used six different disorders to design the clinical reports, it is not very likely that this coincidence is due to randomness. A reasonable explanation would be that clinicians' preference for one symptom order or another is determined by their causal theories. However, if clinicians' preferences for the consistent text order are determined by their causal theories, such causal theories would play an important role in explaining the inconsistency effect on RTs after all.

Our results are in line with previous findings. Specifically, as explained in the Introduction section, Kim and Ahn (2002) also found evidence for the impact of causal theories on diagnostic judgements concerning mental disorders of the DSM-IV. Regarding this previous finding, we have gone some steps further in several respects. First, we have shown that clinicians' causal reasoning can be the result of very fast, on-

line activation and inference processes attributable to System 1. Additionally, the consistency between the reasoning processes detected through on-line measures and the causal-theory-based bias found in clinicians' diagnostic judgements suggests that the former could have a determinant role for the latter. Second, our results show that causal reasoning can be tapped by information that goes beyond the presence or absence of diagnostic criteria. Specifically, the temporal sequence in which symptoms develop seems to play a significant role. Third, by holding constant the symptoms across different conditions, we avoided the confounding influence of other variables such as the conceptual centrality of symptoms or their frequency given each of the disorders used in our experiment. Finally, at variance with Kim and Ahn's study, participants in our experiment did not perform any odd task, such as drawing causal maps, which may be thought as having artificially prompted the use of causal theories. Therefore, the use of causal reasoning (especially, in clinicians) has been shown to occur even when participants are not requested to make their own causal theories explicit through effortful and long, time-consuming processes.

Our results regarding the efficacy judgement task are also in line with previous findings in clinical (Yopchick & Kim, 2009) and in non-clinical contexts (Meder et al., 2008; Sloman & Lagnado, 2005). Specifically, Yopchick and Kim showed that treatment efficacy judgements were determined by the causal status of the symptom more directly affected by the treatment. Thus, if the treatment was aimed at removing the first symptom in a causal chain, it was considered as more effective than if the same treatment was aimed at removing the second symptom in a causal chain. One of the main differences between Yopchick and Kim's study and this experiment is that the causal chain models in their study were created by the experimenters. Also, treatments in their study were not realistic treatments devised to remove specific symptoms. Another important difference is

that the task used by Yopchick and Kim was not focused on mental disorders from the DSM-IV taxonomy. Finally, their study did not include a sample of clinicians. Regarding Meder et al. (2008) and Sloman & Lagnado (2005), participants in their experiments were provided with causal models to link different variables and were requested to make predictive and diagnostic inferences from direct interventions on certain specific variables. As a result, these authors observed the same sort of asymmetry as in our experiment. Specifically, intervening on a variable was judged to have a greater impact on its effects than on its causes. This asymmetry disappeared or tended to decrease when the variables were merely correlated or when the variables were observed rather than acted upon. The interesting aspect of these results is that they indicate the special consideration that interventions have in causal thinking. This consideration is a distinctive feature of causal reasoning that makes intervention tasks a highly relevant experimental tool from which to infer the structure of the causal theory entertained by reasoners.

EXPERIMENT 3

So far, we have used the inconsistency effect as an instrument to detect on-line causal reasoning attributable to System 1. In Experiments 1 and 2, we have found evidence based on the inconsistency effect suggesting that the biases found in the diagnostic judgement task may have occurred as a consequence of causal reasoning processes attributable to System 1. However, our experiments also suggest that, rather than a mere manifestation of causal reasoning, the inconsistency perceived during reading could be understood as a heuristic for diagnostic judgements. According to this idea, as reasoners receive diagnostic information about a patient, they build a coherent and stable mental model to make sense of the case. The more coherent and stable the mental model is, the greater the tendency to agree on the previously established diagnosis. Conversely, the more incoherent and unstable the mental model is, the lesser the tendency to agree on such diagnosis. As said in the Introduction, the formation of mental models based on coherence-driven processes is just what one would expect System 1 to be well suited for. Such processes may well be conceived as the spreading activation processes in dynamic neural networks, which have been shown to be good at producing coherent representations based on fast retrieval and inference. An interesting consequence of these ideas is that the causal coherence of clinical reports can be manipulated to produce a specific pattern of results that can be empirically discriminated from other causal reasoning influences such as the causal status. In the present experiment, we manipulated the coherence of clinical reports to see if such manipulation produced a specific pattern of effects on diagnostic judgements consistent with the effects on RTs. Such consistent effects would provide interesting evidence

supporting the idea that diagnostic judgements are significantly determined by System 1 processes responsible for the computation of causal coherence.

Another concern of Experiment 3 was to assess whether Psychology students could engage in System 1 causal reasoning processes. The effects found in the students' RTs in the previous experiments did not provide any evidence of causal reasoning attributable to System 1. One factor that may have contributed to such results is the use of symptoms whose causal features are not quickly activated in the case of non-experienced clinicians. In Experiment 3 we tried to favour such quick activation by using symptoms whose causal relationships could appear as self-evident, and that could be easily derived from causal theories which students had been previously trained in as part of a course in the academic context. For this reason, we only used students in the present experiment

Participants in Experiment 3 read clinical reports each of which provided information about the diagnosis previously received by a hypothetical patient and about the presence or absence of three symptoms considered as diagnostic criteria according to the DSM-IV (APA, 2000). As in the previous experiment, the symptoms formed part of a causal chain of the sort S1→S2→S3. This causal chain was deduced from a causal theory in which our participants (advanced Psychology students) had recently been trained in a specific course in clinical psychology. By manipulating which of the three symptoms was absent, we created three versions of the clinical report. Information about the diagnosed disorder appeared in the second sentence of the report and was intended to activate participants' representation of the disorder. Later on, participants read a sentence informing about S1 followed by another sentence informing about S2 followed by one more sentence informing about S3. Participants were instructed to read

fluently through the information. Their task was to make a diagnostic judgement after reading all the information. Times spent reading each sentence were recorded unbeknownst to participants. Diagnostic judgements and judgements about causal connections had also to be made by using rating scales without any temporal limitation.

The manipulation allowed us to derive predictions with respect to two types of inconsistencies. The first type of inconsistency is the inconsistency between a diagnosis and the absent symptoms (*categorical inconsistency*). For example, a sentence stating the absence of S1 should conflict with the sentence stating the diagnosis, which should result in longer RTs compared to a sentence stating the presence of S1. Note that this type of inconsistency does not necessarily results from causal inferences, but may only reflect a violation of expectations with respect to symptoms. However, if a causal chain theory connecting the symptoms is activated, then the absence of S1 would be inconsistent with the presence of S2, which is caused by S1 according to the theory (*causal inconsistency*). This should result in longer RTs for the sentence stating the presence of S2 after reading a sentence stating the absence of S1 in comparison to the same sentence on S2 after reading a sentence stating the presence of S1. This specific inconsistency effect is predicted from a simple assumption to compute coherence. According to this assumption, if two elements are positively associated, a good coherence would entail that either both are present, or both are absent. If only one of them is present, the result would be incoherent. This simple assumption is very common in the computation of coherence in some dynamic neural networks. Therefore, a causal inconsistency effect on RTs would suggest that participants engage in System 1 reasoning processes based on the causal chain model.

The combination of symptoms described in the text and participants' assumptions about the causal relations among them should also affect the final diagnostic judgement. According to the *Causal Model Theory of Categorization* (Rehder, 2001, Rehder & Hastie, 2004, Rehder & Kim, 2010), a good member of a category is a member whose features are coherent with the causal laws that form part of the causal model for that category. If we assume that the causal chain S1→S2→S3 is part of the causal model of the disorder, the least coherent exemplar would be the patient lacking symptom S2. This is because the absence of S2 entails the violation of two causal mechanisms. The presence of S1 should produce S2, which is inconsistent with the absence of the latter. S3 should be the effect of S2, but, again, this is contradicted by the absence of the latter. Thus, the patient lacking S2 should receive the lowest ratings in the diagnostic judgement task. The absence of S1 and S3 each violate one causal mechanism. However, according to the *Generative Model*, which is a quantitative model extending Causal Model Theory (cf. Rehder & Kim, 2010), higher diagnostic ratings are predicted for S1 being absent provided that deterministic (or almost deterministic) causal relationships between symptoms are assumed. Otherwise, the Generative Model predicts higher ratings for S3 being absent than S1 being absent. The latter prediction also follows from the *Causal Status Hypothesis* (Ahn, Kim, Lassaline, & Dennis, 2000), according to which the number of features causally affected by a feature should determine its weight in classifications. However, the Causal Status hypothesis could not predict the lowest ratings for the patient lacking S2.

The computation of coherence in Rehder's causal model theory is based on the same assumption referred in the previous paragraph. Basically, if two events are thought to be causally connected, a coherent situation would require the presence or the absence of both of them. The presence of one of them and the absence of the other would lead to

an incoherent situation. Interestingly, Rehder's theory is not committed to a specific algorithm for this computation. Therefore, System 1 processes responsible for the computation of causal coherence may be conceived as an instantiation of Rehder's proposal.

Specific predictions about RTs and diagnostic judgements can be derived from the theoretical models outlined above. Assuming that participants automatically activated a causal chain theory of the disorder, made respective inferences and used them to detect inconsistencies in the text, the predictions shown in Table 4 can be made. Based on the causal model theory of categorization (Rehder & Kim, 2010) predictions with respect to final diagnostic judgements can be derived. They are also presented in Table 4.

	Patient condition (type of clinical report)		
	Patient 1	Patient 2	Patient 3
Sentence about S1	S1 absent (cat inconsistent)	S1 present (consistent)	S1 present (consistent)
RT prediction	Slow reading	Fast reading	Fast reading
Sentence about S2	S2 present (caus inconsistent)	S2 absent (cat and caus inconsistent)	S2 present (consistent)
RT prediction	Slow reading	Slow reading	Fast reading
Sentence about S3	S3 present (consistent)	S3 present (caus consistent)	S3 absent (cat and caus inconsistent)
RT prediction	Fast reading	Slow reading	Slow reading
Diagnostic judgement predictions	Intermediate	Lowest	Highest

Table 4. Predicted pattern of reading times for target sentences drawn from the assumption that participants engage in fast, on-line activation of causal theories and inferences, leading to fast detection of inconsistencies, resulting in longer reading times. The table also shows the predictions derived from the generative model for diagnostic judgements. See text for further explanations.

Note. cat and caus stand for categorically and causally, respectively.

Method

Participants and design

Thirty one undergraduate students from the School of Psychology at the University of Göttingen (Germany) volunteered in our experiment. All of them had been previously trained on the disorders used in our clinical reports in at least two courses. The type of patient factor, defined by the missing diagnostic criterion within a causal chain theory of the disorder (not S1, not S2, or not S3), was manipulated within-subjects. As three types of disorders were used, each participant took part in all nine resulting conditions in random order.

Materials

Depression, obsessive-compulsive disorder of cleaning, and specific phobia to dogs were used to create the clinical reports. The theories on which the causal chain models were based were the following: Beck's (1967) cognitive theory of depression, which proposes that symptoms such as sadness or apathy are the result of an inadequate and biased processing of information; Salkovskis' (1985) cognitive-behavioral model of obsessive-compulsive disorder, according to which the patient tries to reduce her/his anxiety and unease produced by her/his obsessions by doing compulsive rituals; and Mowrer's two-factors model of specific phobia, which states that, initially, an individual acquires an aversion to a stimulus, and then tries to avoid it to reduce the anxiety. The symptoms selected to play the role of S1, S2, and S3 were, in the case of depression, S1: 'To think that bad things always may happen to oneself everywhere', S2: 'Not to feel like going out with friends', S3: 'To be socially isolated and to have a

lot of social problems'; in the case of obsessive-compulsive disorder, S1: 'To feel anxious about getting a bacterial infection', S2: 'To wash hands around 40 times per day', and S3: 'To have strong problems in the workplace because of lack of time'; in the case of specific phobia, S1: 'To have suffered from bad experiences with dogs during childhood', S2:'To feel bad when passing close to dogs' and, S3:'To avoid going to pet shops or parks', respectively.

Every clinical report consisted of six sentences and was structured in the following way. The first sentence was an irrelevant sentence introducing the patient. It was followed by a sentence informing about the diagnosis given by a professional. The third, fourth, and fifth sentences informed about the presence or absence of symptoms S1, S2, and S3, respectively. Every hypothetical patient presented with two of the three symptoms. The absence of a symptom was made explicit by referring to an opposite state or behavior. For example, if the symptom was "*she/he never feels like going out with friends"*, its absence was made explicit by saying that "*she/he always feels like going out with friends"*; or if the symptom was "*she/he washes her/his hands 40 times per day"*, the corresponding sentence for stating its absence was "*she/he washes her/his hands 4 times per day"*. This way, the sentences referring to the presence and the absence of a specific symptom were almost identical regarding length, wording, structure and number of syllables (in German). Finally, the clinical report ended with a final sentence that was held constant across the different clinical reports based on the same disorder (see Appendix F and G).

Procedure

The task was performed in a laboratory with 10 PCs equipped with home-built software programmed in Visual Basic 2005 (Microsoft, USA) and Power Point (Microsoft Office). Participants started by reading the instructions on the computer screen. As in previous experiments, they were informed that they were required to read attentively and fluently a series of clinical reports about hypothetical patients who had been diagnosed with a mental disorder by a clinical psychologist. After reading the instructions, they were presented with an example of a clinical report based on a disorder (anorexia) that was different from those used in the actual experimental task. The example text had the same structure as the experimental clinical reports.

The reading procedure was the same as in Experiment 1 and 2, but instead of using the keyboard to advance, they clicked on the screen by pressing the left mouse button. Each click made all of the letters of a sentence visible while hiding the slashes. A second click had the reverse effect on the read sentence and rendered the following sentence visible. The RTs for each sentence was the time that elapsed between the two consecutive clicks. Clicking after reading the final sentence allowed participants to proceed to the diagnostic judgement task. This time, the ratings could range from 0 to 10, meaning "*Completely sure that the correct diagnosis is another one"*, and "*Completely sure that the diagnosis is correct"*, respectively. After the diagnostic judgement task, participants had to judge the causal relationship for each of all the possible pairs of events mentioned in the clinical report. This task was designed to check whether participants' causal beliefs conformed to the causal chain model on which we based our manipulation. Thus, if a clinical report informed, for example, about the absence of S1 and the presence of S3, the participants had to judge the extent

to which they thought that the absence of S1 could be the cause of the presence of S3, as well as the extent to which S3 could be the cause of the absence of S1. As there were three symptoms per clinical report, and two possible causal directions, participants had to make a total of six causal judgements per clinical report. The order in which participants made their judgements for the different pairs was completely randomized. Before facing this task, the message "*Now, you will have to judge the extent to which you think that each symptom could cause the others. You will use scales from 0 to 10 again"* appeared on the screen. Then, six questions with the wording "*To what extent do you think that* [Symptom X] *affects* [Symptom Y]*?"* prompted participants to make their judgements. Below these messages, the same scrollbar from 0 (meaning *"Completely sure that it does not affect at all"*) to 10 (meaning "*Completely sure that it affects strongly"*) was displayed. After the causal judgement task, participants performed further clinical tasks for a different study. Once these tasks were finished, participants proceeded to the next clinical report. As we used three different disorders to build up the different texts, and there were three different versions per disorder to manipulate the symptom that filled the role of the absent symptom, every participant read a total of nine clinical reports, each one followed by the corresponding diagnostic and causal judgement task. The clinical reports were presented in a random order.

Results and discussion

Manipulation check: Causal-link judgements. The participants' judgements in the causal-link judgement task were analysed to check whether their causal beliefs were in accordance with the causal chain model (i.e., S1→S2→S3) on which we based our predictions. Causal-link judgements were distinguished according to three different

criteria based on the causal chain. If we take into account the direction of the causal link, we have forward (e.g., the extent to which S1 caused S2) and backward causal judgements (e.g., the extent to which S2 caused S1). With respect to the presence of symptoms, we have judgements about the causal relationship between two present symptoms (e.g., the extent to which S3 caused S1) and about a present symptom and an absent symptom (e.g., the extent to which S3 caused the absence of S1). Finally, regarding the contiguity of symptoms in the causal chain, we have causal judgements about the relationship between two contiguous symptoms (e.g., the extent to which S1 caused S2) and between two non-contiguous symptoms (e.g., the extent to which S1 caused S3). These criteria were considered as factors in a 2 (Direction: Forward vs. backward) x 2 (Presence: Two present vs. one present) x 2 (Contiguity: Contiguous symptoms vs. non-contiguous symptoms) repeated measures design. For each participant, judgements were collapsed into a single mean judgement for each of the eight conditions (see Table 5).

	Both symptoms present		*Only one symptom present*	
	Mean	SD	Mean	SD
Contiguous, forward	8.49	0.82	1.10	0.74
Non-contiguous, forward	5.73	2.01	0.96	0.83
Contiguous, backward	4.38	1.78	1.90	1.33
Non-contiguous, backward	3.20	1.71	1.23	1.04

Table 5. Means and standard deviations of causal link judgements (from 0 to 10).

Table 5 shows the participants' mean judgements in each condition. As can be seen, judgements in the two-present condition were much higher than in the one-present condition. Judgements in the latter condition were near 0 and differed between each other by less than one point. This is just what one would expect provided that the participants' causal beliefs conformed to the causal chain model. If we now focus on the two-present conditions, we can also appreciate that, consistent with the causal chain model, judgements in the forward condition were considerably higher than in the backward condition. This occurred, especially, in the contiguous conditions. Finally, and consistent with the causal chain model, ratings were higher in the contiguous than in the non-contiguous symptoms condition. Consistent with the causal chain model, this difference can be more easily appreciated in the forward than in the backward conditions. Thus, those causal links that were consistent with a causal chain model received high ratings, whereas those that were inconsistent with the causal chain model received low ratings. These impressions were confirmed by a repeated measures ANOVA 2 (Direction: Forward vs. backward) x 2 (Presence: Two present vs. one present) x 2 (Contiguity: Contiguous symptoms vs. non-contiguous symptoms) on the participants' judgements, which yielded a significant effect of Direction, $F(1, 30) = 56.44$, $MSE = 2.14$; $p < .001$; $\eta^2 = .65$, Contiguity, $F(1, 30) = 97.51$, $MSE = 0.89$; $p < .001$; $\eta^2 = .77$, Presence, $F(1, 30) = 531.15$, $MSE = 2.02$; $p < .001$; $\eta^2 = .95$, Direction x Presence, $F(1, 30) = 160.56$, $MSE = 1.43$; $p < .001$; $\eta^2 = .84$, Contiguity x Presence, $F(1, 30) = 27.58$, $MSE = 1.37$; $p < .001$; $\eta^2 = .48$, Contiguity x Direction, $F(1, 30) = 9.61$, $MSE = 0.45$; $p = .004$; $\eta^2 = .24$, and the significant three-way interaction Contiguity x Direction x Presence, $F(1, 30) = 22.19$, $MSE = 0.79$; $p = .004$; $\eta^2 = .43$. Because all the differences between ratings tend to disappear when only one of the symptoms is

present, it seems that Factor Presence interacts with the remaining factors and with the Direction x Contiguity interaction by decreasing their effects to make them almost disappear in the one-present condition. Because of this, we only analysed the results in the two-present condition as a follow-up of the global analysis. A repeated measures ANOVA 2 (Direction: Forward vs. backward) x 2 (Contiguity: Contiguous symptoms vs. non-contiguous symptoms) on the participants' judgements in the two-present condition yielded a significant effect of Contiguity, $F(1, 30) = 66.55$, $MSE = 1.8$; $p < .001$; $\eta^2 = .69$, Direction, $F(1, 30) = 124.67$, $MSE = 2.75$; $p < .001$; $\eta^2 = .81$, as well as the significant interaction Contiguity x Direction, $F(1, 30) = 20.94$, $MSE = 0.94$; $p < .001$; $\eta^2 = .41$. For the reasons stated above, the main effects found are quite consistent with the causal chain model. The Contiguity x Direction interaction is also hardly surprising because, if participants were entertaining a causal-chain theory, backward causal links should tend to be viewed as implausible regardless of the contiguity between symptoms. Thus, the difference between the contiguous and non-contiguous conditions should tend to disappear in the backward condition (see Table 5). Overall, the pattern of results found in the participants' causal-link judgements fits quite well the causal chain model on which we based our manipulation and predictions.

Reading times. Statistical analyses were conducted to test whether RTs were longer for the inconsistent sentences than for the consistent sentences. Inconsistent sentences included those sentences stating the absence of a symptom (categorical inconsistency) as well as those sentences stating the presence of a symptom given the absence of the previous symptom in the causal chain (causal inconsistency). As explained above, causal inconsistencies provide direct evidence for fast, on-line causal reasoning processes.

Before the analysis, as in Experiment 2, RTs were normalized by the number of letters of the corresponding target sentence and were subjected to the same filtering process as in the previous experiments. As a consequence, twelve measures were removed. Then, RTs across the different mental disorders were collapsed into a single average RT per patient and symptom condition. Table 6 shows the RTs for every target sentence in each of the different patient conditions collapsed across the different disorders. Thus, the first column shows the RTs for sentences referring to Symptoms S1, S2, and S3 in the patient condition in which the absent symptom was S1. The second and the third columns show the corresponding RTs in those patient conditions in which the absent symptoms were S2 and S3, respectively. As can be seen, the RTs for the inconsistent sentences were, in general, longer than for the consistent sentences. For example, in the case of Symptom 1, the RTs in the first patient condition were longer than in the remaining conditions. In the case of Symptom 2, the RTs in third patient condition (the consistent condition) were shorter than in the remaining inconsistent conditions. Finally, in the case of Symptom 3, the RTs in the first patient condition (the consistent condition) were shorter than in the remaining inconsistent conditions. This pattern of results is consistent with the predictions shown in Table 4. To confirm these impressions, we performed a repeated measures ANOVA 3 (Patient: Patient 1-absence of S1, Patient 2-absence of S2, Patient 3-absence of S3) x 3 (Symptom: Symptom 1, Symptom 2, Symptom 3), which yielded a significant main effect of Patient, F (2, 60) = 5.7, MSE = 124.68; $p < .005$; $\eta^2 = .16$, a significant effect of Symptom, F (2, 60) = 18.85, MSE = 101.26; $p < .001$; $\eta^2 = .39$, and the significant interaction Patient x Symptom, F (2,60) = 13.8, MSE = 100.37; $p < .001$; $\eta^2 = .32$.

	Patient 1 (~~S1~~,S2,S3)		Patient 2 (S1,~~S2~~,S3)		Patient 3 (S1,S2,~~S3~~)	
	M	*SD*	*M*	*SD*	*M*	*SD*
RT Symptom 1	53 ms	14.1	43 ms	9.3	44 ms	11.8
RT Symptom 2	60 ms	14.4	59 ms	12.6	47 ms	9.9
RT Symptom 3	48 ms	11.8	59 ms	19.3	55 ms	12.8
Diagnostic judgements	5.74	1.58	4.41	1.71	7.05	1.35

Table 6. Means and standard deviations of reading times [RT] for the target sentences (in milliseconds per letter), and mean and standard deviations of diagnostic judgements (from 0 to 10).

To test whether participants activated a causal chain model of the disorder and detected causal inconsistencies, we conducted planned comparisons for Symptoms 2 and 3. Note that Symptom 1 is involved in a categorical inconsistency when it is absent in the first patient condition. Thus, it is not possible to disentangle a possible causal inconsistency from a categorical inconsistency in this case. A t-test for paired comparisons yielded a significant difference between Patient 1 and Patient 3 conditions for Symptom 2, $t(30) = 4.63$, $p < .001$, $\eta^2 = .42$. Although Symptom 2 was present in Patient 1 and Patient 3, and its presence was consistent with the diagnosis stated in the clinical report, different RTs resulted. As Symptom 2 was present despite its cause Symptom 1 being absent (Patient 1), the participants seemed to have experienced a

causal inconsistency. This was not the case when both Symptom 1 and Symptom 2 were present (Patient 3), which is in accordance with the causal chain model. A t-test for paired comparisons also revealed a significant difference between Patient 1 and Patient 2 conditions for Symptom 3, $t(30) = -4.04$, $p = .001$, $\eta^2 = .35$. This finding replicates the finding for Symptom 2. Again, Symptom 3 was present for Patient 1 and Patient 2, but it occurred despite Symptom 2 being absent in Patient 2, which created an inconsistency according to the assumed chain model. The participants were apparently sensitive to the causal inconsistencies.

However, one may argue that the differential RTs may not be due to an inconsistency of the observed symptom with the assumed causal chain model, but were due to a carry-over effect of the longer RTs of the previous sentence stating the absence of an expected symptom. To assess the plausibility of this account, we analysed the RTs for the final sentence, which was an irrelevant sentence conveying no information about symptoms. If the reading of an inconsistent sentence stating the absence of a symptom produces a carryover effect on the following sentence, the RTs for the final sentence should be longer in the Patient 3 than in the Patient 1 and Patient 2 conditions. Note that only in the Patient 3 condition, the final sentence was preceded by a sentence stating the absence of a symptom. The mean normalized RTs for the final sentence were 34.7 ms, 33.3 ms, and 32.3 ms, corresponding to the Patient 1, Patient 2, and Patient 3 conditions, respectively. This pattern of results is clearly at odds with a carryover effect. Hence the longer RTs found for Symptom 2 when comparing Patient 1 to Patient 3, and for Symptom 3 when comparing Patient 2 to Patient 1 are most likely due to participants' causal reasoning based on the assumed causal chain model. Thus, the two findings provide strong evidence for the involvement of causal reasoning while reading through a clinical report. As the participants did not stop to deliberate about the given

information, but read fluently through the information, these findings indicate that the participants automatically activated their respective causal theories, engaged in fast inference making, and in fast detection of inconsistencies. This speaks to the presence of System 1 processes of causal reasoning as diagnostic information is being gathered.

Diagnostic judgements. Judgements for the different disorders were collapsed into a single mean per patient condition for each participant. The analyses reported were conducted on these resulting means. Table 6 shows participants' mean judgements in each condition. Participants agreed with the diagnosis stated in the preliminary information to a greater extent in the Patient 3 than in the Patient 1 condition, and, in turn, in the Patient 1 more than in the Patient 2 condition. These impressions are supported by a repeated measures ANOVA (Patient: Patient 1 vs. Patient 2 vs. Patient 3) on the participants' judgements, which yielded the significant main effect of Patient, F (2, 60) = 41.44, MSE = 1.31; $p < .001$; $\eta^2 = .58$. T-tests for paired comparisons revealed significant differences between the Patient 1 and Patient 2 conditions, t (30) = 4.52; $p < .001$; $\eta^2 = .40$, between the Patient 1 and Patient 3 conditions, t (30) = -4.99; $p < .001$; $\eta^2 = .45$, and between the Patient 2 and Patient 3 conditions, t (30) = -8.49; $p < .001$; $\eta^2 = .71$. Note that these differences would still be significant even if we take the Bonferroni approach (i.e., with $\alpha = .016$), which is very conservative, to protect the statistical analyses against an accumulation of Type 1 error.

These findings clearly show that the participants' assumptions about the causal relations among symptoms affected diagnostic judgements. More precisely, the pattern of diagnostic ratings supports the generative causal model of categorization (Rehder & Kim, 2010). As outlined above, this model predicts that the Patient 2 condition should be the least coherent and the Patient 3 condition the most coherent (assuming

probabilistic causal relations). Thus, the lowest judgements of agreement should be found in the Patient 2 condition and the highest in the Patient 3 condition, which was in fact the case. Other models, like the causal status hypothesis (Ahn et al., 2000) cannot predict this finding.

Although we cannot rule out that the participants may have engaged in some form of deliberative, System 2 reasoning processes when facing the diagnostic judgement task, our findings suggest that more intuitive, System 1 processes have been involved as well. The participants' detection of inconsistencies during reading preceded the diagnostic judgements made and they were in accordance with the diagnostic ratings. For Patient 2, two inconsistency effects were found in readings times (for Symptoms 2 and 3) and diagnostic ratings were lowest. For Patient 1, one strong (Symptom 2) and one rather weak (Symptom 1) inconsistency effect resulted and ratings were intermediate. Finally, for Patient 3, only one inconsistency effect in RTs was found (for Symptom 3) and high diagnostic ratings resulted. Hence, perceived inconsistencies between diagnostic categories and observed symptoms (categorical inconsistency) and between causal chain models and observed symptoms (causal inconsistency) were together predictive of diagnostic ratings. These perceived inconsistencies resulted automatically when participants read the clinical information. This strongly suggests that the computation of coherence envisaged by Rehder's Causal Model Theory could have been undertaken by System 1 during reading, which, in turn, could have anchored or significantly determined System 2 at the diagnostic judgement task.

EXPERIMENT 4

The previous experiments showed that reasoners relied on System 1 causal reasoning processes during reading and that the outcomes of such processes were consistent with later diagnostic judgements. Also, the pattern of results found in Experiments 2 and 3 suggests that participants are sensitive to information confirming or contradicting there assumptions about the causal structure underlying the relationships between symptoms. Specifically, the participants in these experiments seemed to have reasoned according to a causal chain model of the type S1→S2→S3. However, several authors have pointed out that causal reasoning is not only based on assumptions about causal relatedness but on assumptions about causal mechanisms through which variables are connected (Ahn & Bailenson, 1996; Ahn & Kalish, 2000; Ahn, Kalish, Medin, & Gelman, 1995). According to this notion, when people say that A causes B, they mean that there is a mechanistic process between A and B, which allows for the transmission of power from the former to the latter and that somehow forces the occurrence of B given A (Ahn & Kalish, 2000). If this is correct, causal reasoning about two causally related symptoms, S1 and S2, should involve assumptions about the mechanism connecting S1 and S2, and explaining how the former causes the latter. These assumptions in turn could result in a new type of perceived causal inconsistency: Mechanistic inconsistency. These inconsistencies should be experienced if a) participants automatically activate causal theories of a disorder including assumptions about the causal mechanisms connecting potential symptoms, and b) they are informed that symptoms indicating a certain disease or disorder were linked through unexpected mechanisms. Note that this type of inconsistency may arise even when all

observed symptoms are coherent with the assumed disorder. In this case, causal inconsistencies should be experienced without categorical inconsistencies.

Experiment 4 tested this hypothesis by manipulating the information concerning the causal mechanism connecting the symptoms within a causal chain while holding constant the causal structure underlying the connections between symptoms. In one condition (the consistent condition) the presented mechanisms were consistent with the participants' causal theory of the disorder mentioned in the clinical report. In the alternative, inconsistent condition, the causal mechanisms were inconsistent with the causal theory. In both conditions, the hypothetical patient presented with the same symptoms, which were connected through the same causal chain (S1→S2→S3). For example, in one condition readers were informed that a patient had been diagnosed with anorexia nervosa. Later in the clinical report, they were informed that the patients' fear of getting fat (Symptom 1) led to food aversion (Mechanism 1), which, in turn, led to weight loss (Symptom 2), which, in turn, led to a strong hormonal change (Mechanism 2), which, finally, led to amenorrhea (Symptom 3). In this case, a causal mechanism is made explicit that is consistent with some theories of anorexia nervosa. Alternatively, participants were informed that the patient's fear of getting fat (Symptom 1) led to stomach bleeding (Mechanism 1b), which, in turn, led to weight loss (Symptom 2), which, in turn, led to the ingestion of prescribed vitamins and medicines (Mechanism 2b), which, finally, led to amenorrhea (Symptom 3). These mechanisms may be seen as less consistent with psychological theories of anorexia nervosa, and more consistent with bio-medical mechanisms related to stomach ulcer. Hence as a first factor we manipulated the type of mechanism connecting the symptoms. As a second factor we manipulated the diagnosis assigned to a patient. For each set of symptoms, the hypothetical patient could have been diagnosed with either a psychological disorder, for

which the symptoms were diagnostic criteria, or a bio-medical disease consistent with the observed symptoms. We expected participants to experience causal inconsistency either when the patient was diagnosed with a psychological disorder but the clinical report informed about a bio-medical mechanism or if the patient was diagnosed with a bio-medical disease but the clinical report informed about a psychological mechanism. Consequently, we predicted longer RTs in the inconsistent than in the consistent condition. As in our previous experiments, we also expected participants' diagnostic judgements to be higher in the consistent than in the inconsistent conditions. Note that in contrast to Experiments 1 and 3 these judgements would be determined only by causal and not categorical inconsistency as the same diagnostic criteria were present in all conditions.

Method

Participants and design

Thirty-four Psychology students from the University of Malaga (Spain) participated. Twenty seven were Psychology students in their last year, whereas the remaining participants were postgraduate students enrolled in a Master's degree program in Health Psychology. All of them had received some training on the diagnostic criteria and theories of the different disorders. All students participated voluntarily. Although we were also interested in clinicians, we preferred to start conducting the experiment with students to see if we could find further evidence of System 1 causal reasoning in students. Two factors were manipulated within-subjects: Type of disease/disorder (psychological vs. biomedical) and type of mechanism

(psychological vs. bio-medical). Three psychological disorders and three bio-medical diseases were used with each participant receiving a set of twelve problems.

Materials

Half of the hypothetical patients suffered from one of three possible mental disorders: Anorexia nervosa, social phobia, and generalized anxiety. The other half suffered from one of the following medical diseases, which could result in the same symptoms: Stomach ulcer, psoriasis, and migraine. The symptoms used in the reports were selected from the diagnostic criteria of the mental disorders established by the DSM-IV-TR. Three symptoms per mental disorder were selected: For anorexia S1: High anxiety about the possibility of gaining weight, S2: Substantial loss of weight and, S3: Amenorrhea; for social phobia S1: Fear of feeling embarrassed in social situations, S2: Avoidance of social situations and, S3: Problems at work; for generalized Anxiety disorder S1: Excessive and uncontrollable worries, S2: Fatigue, muscular tension and restlessness, and S3: Sleep problems. These symptoms formed part of a causal chain (S1→S2→S3) that was made explicit in the clinical reports. Each causal chain could operate according to two different sets of mechanisms: a) mechanisms compatible with causal psychological theories of mental disorders, and b) bio-medical mechanisms compatible with causal theories of medical diseases. The theories that we relied on to derive the psychological mechanisms were: Crisp's (1980) behavioral model of anorexia nervosa, Clark and Wells' (1995) cognitive-behavioral model of social phobia, and Ladouceur's (1998) cognitive model of generalized anxiety disorder. All participants had been previously trained on these theories in different courses. The bio-

medical mechanisms were compatible with familiar notions that people normally acquire from exposure to the media.

Twelve clinical reports were built up to fill up the four cells of a 2 (Diagnosis: Psychological disorder vs. medical disease) x 2 (Mechanism: Psychological mechanism vs. bio-medical mechanism) within-subjects design. For each triplet of symptoms (or causal chain) there were two possible mechanisms, one psychological and one bio-medical. The psychological mechanism of anorexia nervosa was contrasted with the bio-medical mechanism related to stomach ulcer, the psychological mechanism of social phobia with the bio-medical mechanism of psoriasis; and the psychological mechanism of generalized anxiety with the bio-medical mechanism of migraine. Each clinical report comprised ten sentences. The first one was an introductory sentence. The second sentence informed about the diagnosis previously received by the patient. This diagnosis could be either consistent with the mechanism described in the report or inconsistent. The third sentence provided information about a precipitating event consistent with the causal mechanism described later in the text. For example, in the case of anorexia nervosa, participants read the sentence "*She told that, in her childhood, her schoolmates called her 'fat'*". The next sentence introduced the causal mechanism. The next five sentences described the causal mechanisms together with the symptoms. For example, in the case of anorexia nervosa, the participants read the following sentences: "*It all started with high anxiety about the possibility of gaining weight. The high anxiety about the possibility of gaining weight led to food aversion. The aversion to food led to a remarkable loss of weight. The remarkable loss of weight led to a strong hormonal change. The strong hormonal change led to the stopping of the period"* (all the sentences have been translated from Spanish). As can be seen, these sentences describe the causal mechanism step by step from the first through the last link of the

chain. This same structure was used in every clinical report. Following the description of the mechanism, participants read an irrelevant final sentence (see Appendix H).

Procedure

The experimental task was programmed in Visual basic 2005 and was run in the same laboratory than Experiment 1 and Experiment 2. The participants started by reading the instructions on the computer screen, which included an example based on a patient diagnosed with obsessive-compulsive disorder. The reading task of clinical reports proceeded as in previous experiments: The participants moved from one sentence to the next by pressing the space bar. After reading each clinical report, the participants were confronted with a diagnostic judgement task in which they had to rate the extent to which they agreed with the diagnosis received by the patient (in a scale from 0 to 100).

Once the participants had made their diagnostic judgement, they were asked to decide whether a psychological or a medical treatment would be more effective for the patient. For this purpose, a new display was shown on the screen in which the participants could read the following message: "*Now, you will have to make a decision about treatment. For the client described in the report, what type of treatment do you think it would be more efficient: A psychological or a medical treatment? Place the face of the scale below on the position that best represents your decision. Although the face appears at the centre of the scale, you can move it to the left or right depending on your preferred treatment*". Below this message, a horizontal scale was displayed together with a label indicating the psychological treatment at one extreme and another label

indicating the medical treatment at the other extreme. The specific side for each treatment was counterbalanced. The percentage numbers 50%, 75%, and 100% were shown on top of the scale. The former appeared in the middle position, whereas, each of the remaining numbers appeared at both sides of the scale at further positions from the centre. The initial position of the scrollbar face corresponded to the 50% rating, and the participants could move it to the left or right depending on which of the treatments they would choose according to the information provided through the clinical report. Thus, the percentage ratings for the two types of treatment were complementary in such a way that if the participants' rating for the psychological treatment was, for instance, 75%, the rating for the medical treatment was 25%. As people's decisions on interventions depend on their causal model explaining the symptoms, the participants' preference for the psychological treatment should be greater in the psychological mechanism than in the medical mechanism condition. Complementarily, the preference for the medical treatment should be greater in the medical mechanism than in the psychological mechanism condition. However, this result should be found only if the descriptions provided for the psychological and the medical mechanisms through the clinical reports are interpreted as such by the participants. Therefore, this treatment-decision making task could be used to check whether the participants interpreted the descriptions of the causal mechanisms as intended.

After the treatment decision task the participants proceeded to the next clinical report. They had to read a total of twelve different clinical reports (3 causal chains x 2 causal mechanisms x 2 diseases) followed by the corresponding diagnostic judgement and treatment decision tasks. The order of the clinical reports was completely randomized across subjects.

Results and discussion

Reading times. Statistical analyses were conducted to test whether RTs for the target sentences (i.e., the sentences describing the causal mechanism together with the symptoms) were affected by their consistency (or inconsistency) with the preliminary information. As outlined above, we expected to find longer RTs for the bio-medical mechanism than for the psychological mechanism in clinical reports about patients who had been diagnosed with the psychological disorder. The opposite pattern of results was expected in clinical reports about patients who had been diagnosed with a medical disease.

The analyses were performed on the sum of the RTs for the five target sentences normalized by the total number of letters. The same filtering process as in previous experiments was used. As a consequence, six outlier RTs were removed. After the filtering process, a single mean RT per condition and participant was computed, leading to four average RT measures, one per condition.

Diagnosis	Psychological				Medical			
Mechanisms	Psychological		Bio-medical		Psychological		Bio-medical	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Reading times	53 ms	15.6	58 ms	18.5	52 ms	12.5	42 ms	14.0
Diagnostic judgements	76.72	20.52	33.92	22.19	9.95	13.51	35.20	20.03

Table 7. Mean normalized reading times and standard deviations for target sentences (in milliseconds per letter) and mean final diagnostic judgements (from 0 to 100) averaged over three different disorders per condition.

Table 7 shows the descriptive statistics. In the psychological diagnosis condition, longer RTs were found for bio-medical mechanisms than for psychological mechanisms. When the diagnosis was medical, longer RTs were found for psychological mechanisms than for medical mechanisms. This pattern of results shows an interaction effect consistent with our predictions. These impressions were confirmed by a repeated measures ANOVA 2 (Diagnosis: Psychological disorder vs. medical disease) x 2 (Mechanism: Psychological mechanism vs. bio-medical mechanism), which yielded the significant effect of Diagnosis, F (1, 33) = 8.6, MSE = 287.75; p = .006; η^2 = .21, and the significant effect of the interaction Diagnosis x Mechanism, F (1,33) = 33.43, MSE = 61.76; $p < .001$; η^2 = .50. The effect of Mechanism was not significant, F (1, 33) = 2.66, MSE = 58.41; p = .112. The analyses of the simple effects of Mechanism

within each level of Diagnosis revealed the significant effect of the former within both the psychological diagnosis condition, t (33) = -3.59; p = .001; η^2 = .28, and the medical diagnosis condition, t (33) = -4.64; p < .001; η^2 = .39.

As expected, we found an inconsistency effect which was significantly modulated by the preliminary information concerning the diagnosis received by the patient. When the causal mechanisms were consistent with the participants' causal theories for the disorder or disease, the target sentences were read faster than when the causal mechanisms were inconsistent. Note that this result cannot be explained by claiming that the detection of inconsistencies was based on the plausibility of the causal mechanisms regardless of the participants' causal theories of the disorders/diseases. Without an involvement of the participants' causal theories, no interaction effect should have been found. Thus, the best account of the results in RTs is that the participants automatically activated their causal theories of the disorders/diseases mentioned in the preliminary information, and that the theories were rapidly and efficiently used to make inferences and to integrate information from later sentences, which led to a fast detection of inconsistencies. As in previous experiments, this speaks to the presence of System 1 processes of causal reasoning.

A possible criticism concerning our interpretation of the results is that, according to such interpretation, a significant difference between the psychological disorder and the medical disease conditions should have been found within the psychological mechanism condition. Specifically, RTs should be longer in latter than in the former condition. However, such a straightforward comparison may be misleading given the large and robust main effect of Diagnosis on RTs. In general, the participants were slower in the psychological-disorder than in the medical-disease condition. As all

participants were Psychology students, they might have been more concerned for their performance in the former than in the latter condition. After all, Psychology students are expected to perform well on psychological rather than on medical diagnosis. This difference in the participants' concern may well explain why they made a more careful reading in the psychological-disorder than in the medical-disease condition.

Diagnostic judgements. A single mean diagnostic judgement was calculated for each participant in each of the four conditions by averaging across disorders/diseases. Table 7 shows the participants' mean judgements in each condition. As expected, in the psychological diagnosis condition, the participants agreed with the previously given diagnosis to a greater extent when the mechanisms were psychological than when they were bio-medical. Conversely, in the medical diagnosis condition, the participants' ratings were higher when the described mechanisms were bio-medical than when they were psychological. This interaction indicates that the participants agreed with the diagnosis to a greater extent in the consistent than in the inconsistent reports. This pattern was confirmed by a repeated measures ANOVA 2 (Diagnosis: Psychological disorder vs. medical disease) x 2 (Mechanism: Psychological mechanisms vs. bio-medical mechanisms) on the participants' diagnostic judgements, which yielded the significant main effects of Diagnosis, F (1, 33) = 97.25, MSE = 374.84; $p < .001$; η^2 = .75, and Mechanism, F (1, 33) = 9.90, MSE = 264.28; p = .003; η^2 = .23, as well as the significant interaction Diagnosis x Mechanism, F (1, 33) = 149.14, MSE = 263.87; $p <$ .001; η^2 = .82. Further analyses of the simple effects of Mechanism within each level of Diagnosis revealed the significant effect of the former factor in the psychological diagnosis condition, $t(33) = 10.93$; $p < .001$; η^2 = .78, and in the medical diagnosis condition, t (33) = 6.36; $p < .001$; η^2 = .55.

Our results show that the participants' diagnostic judgements were affected by several factors. First, they were affected by the coherence among the observed symptoms and the diagnosis regardless of mechanisms (main effect of diagnosis). As the main symptoms were diagnostic criteria of mental disorders, this effect is not surprising. Diagnostic judgements were also affected by the type of mechanism regardless of diagnosis. Given that main symptoms were diagnostic criteria and that symptoms connecting them were a mix of somatic and psychological symptoms, it seems that they were regarded to be more consistent with psychological than bio-medical mechanisms. Most important, the interaction effect found shows that diagnostic judgements depended on the coherence among diagnosis, the causal theories connected with the diagnosis, which includes assumptions about the underlying mechanisms, and the observed symptoms and mechanisms. Focusing on the diagnosis of psychological disorders, the results show that assumptions about mechanisms strongly influence diagnosis even when the diagnostic criteria established by the DSM-IV are held constant.

As in the previous experiments, the pattern of results found in diagnostic judgements fits quite well the pattern of results found in RTs. This again suggests that the causal reasoning processes that occur automatically when clinical reports are read may play an important role in later diagnostic judgements. It seems that the inconsistencies detected while intuitively processing the given information affect the explicit appraisal of diagnostic hypotheses, and this could be other indirect evidence that System 1 causal reasoning processes may have affected diagnostic judgements. However, this evidence is again indirect.

Treatment-decision judgements. Statistical analyses were conducted to test whether the participants' preference for psychological vs. medical treatments depended on the causal mechanism described in the clinical reports regardless of the diagnosis received by the client. We expected higher ratings for the psychological treatment in the psychological than in the medical mechanism condition. The opposite pattern of results was expected to be found concerning the ratings received by the medical treatment.

As the judgements for each type of treatment were complementary, we analysed the participants' preference for the psychological treatment. Thus, a rating of 30 would imply that the medical treatment received a rating of 70. As in the case of RTs and diagnosis, a single mean decision judgement per condition and participant was calculated. The analyses reported have been conducted on these means. Table 7 shows the participants' mean judgements in each condition. The participants gave higher ratings to the psychological treatment in the psychological mechanism than in the medical mechanism condition. Complementarily, the medical treatment received higher ratings in the medical mechanism than in the psychological mechanism condition. Importantly, the difference between the psychological and the medical mechanism condition did not seem to have been modulated by the diagnosis received by the client. This was confirmed by a repeated measures ANOVA 2 (Diagnosis: Psychological diagnosis vs. medical diagnosis) x 2 (Mechanism: Psychological mechanism vs. medical mechanism) on the participants' judgements, which yielded the significant main effect of Mechanism, F (1, 33) = 140.26, MSE = 210.39; $p < .001$; $\eta^2 = .81$. The effect of Diagnosis was only marginally significant [F (1, 33) = 3.91, MSE = 230.11; $p = .056$], and the interaction was not significant [F (1, 33) = 1.77].

As expected, the participants' decisions about treatment were affected by the causal mechanism that explained how symptoms developed. In the medical mechanism condition, the participants' preference for the psychological treatment decreased, whereas the preference for the medical treatment increased. This suggests that the description provided about the mechanisms through the clinical reports were interpreted as intended.

GENERAL DISCUSSION

The main objective of the present study was to show that not only slow and deliberative reasoning processes, identified as System 2 processes by Kahneman & Frederick (2002), can be at work in diagnostic judgements. Fast and on-line reasoning processes with the properties attributed to System 1 may also be at work. At the same time, we have gone to some length to show that these System 1 processes can be affected by diagnosticians' causal theories. As mentioned in the Introduction, causal reasoning is crucial to make accurate inferences, to choose effective interventions, and plays an important role in comprehension processes in many different situations. However, when it comes to the diagnosis of mental disorders, clinicians are trained to use the DSM diagnostic criteria without taking into account their causal theories or their causal beliefs about how symptoms are linked to each other. Therefore, the diagnosis of mental disorders is an interesting situation to test how strong is our tendency to rely on our causal beliefs, and whether such reliance could bias our diagnostic judgements, thereby causing some departure from the DSM prescriptions.

The objective stated above was divided up into four experimental questions. The first question was whether we could replicate the diagnostic bias based on the causal status effect found by Kim & Ahn (2002), and whether we could find evidence of the involvement of System 1 causal reasoning processes in the commission of such bias. The second question was whether we could find more compelling evidence for the role of causal reasoning processes in biases in the diagnosis of mental disorders by showing the influence of information regarding the temporal order in which symptoms develop and the causal connections between them. Again, we also aimed to find evidence demonstrating the implication of System 1 causal reasoning processes in such bias. The

third question was whether biases in diagnostic judgements were based on System 1 computation of causal coherence, and to see whether, contrary to the previous experiments, we could demonstrate such causal reasoning influence on Psychology students. The fourth question was to what extent information regarding causal mechanisms is also involved in System 1 causal reasoning, and, thus, could also be responsible for the commission of biases in diagnostic judgements. Experiment 1, Experiment 2, Experiment 3 and Experiment 4 were focused on these four questions, respectively. In all of the experiments, the study of System 1 causal reasoning processes was based on the detection of fast, on-line reasoning processes during reading of clinical reports. The on-line technique relied upon RT measures of sentences subjected to the manipulation of consistency according to the inconsistency paradigm in reading comprehension.

In Experiment 1, we created different levels of inconsistencies based on the use of target sentences stating the absence of either central- or peripheral-cause symptoms in the clinical reports. All inconsistencies, though, involved the absence of diagnostic criteria of equal importance according to DSM-IV. The results showed a general inconsistency effect in both samples of participants (i.e., Psychology students and experienced clinicians) in the sense that RTs for the target sentences were longer in the inconsistent condition than in the control (or neutral) condition. Second, only in experienced clinicians, the inconsistency effect was modulated by the causal status of the symptom referred to in the target sentence. Specifically, the inconsistent effect detected was of a greater magnitude when the target sentence referred to the absence a central-cause symptom than when it referred to the absence a peripheral-cause symptom. Third, in all cases, RTs results were congruent with the diagnostic judgements obtained. Both students and clinicians gave lower judgements of agreement

with the diagnosis provided in the inconsistent than in the control condition. Only within the sample of experienced clinicians, these judgements showed an inconsistency effect modulated by the symptom causal status, that is, a greater inconsistency effect in the causal than in the peripheral symptom.

In Experiment 2, we manipulated the information concerning the temporal sequence of symptoms and the causal connections between them. As in Experiment 1, Experiment 2 was carried out both with students and experienced clinicians. The results found with students showed that their diagnostic judgements were affected by information about the temporal sequence of symptoms together with information about causal connections. Specifically, students agreed on the diagnosis received by the hypothetical patients to a greater extent when the temporal sequence of symptoms and the causal connections between them were consistent with the entertained causal theories of the diagnosed disorders than when the information provided was inconsistent with such theories. Also, students spent more time in the diagnostic judgement in the inconsistent non-causal than in the consistent causal condition. However, we did not find convincing evidence of fast and on-line causal reasoning processes in this sample, as the time spent reading the target sentences conveying information about the temporal sequence of symptoms seemed to be unaffected by whether the temporal order was consistent or inconsistent with the causal theory of the disorder mentioned in the preliminary information of the clinical report. The results found in experienced clinicians also showed a greater agreement on the diagnosis received by the hypothetical client in the consistent causal condition than in the inconsistent non-causal condition. Clinicians also took longer in the diagnostic judgement task in the latter than in the former condition, suggesting that they were trying to solve the causal

inconsistencies found in the clinical report before making the judgement. Additionally, we found evidence of fast and on-line causal reasoning, as evidenced by the clinicians' RTs for the target sentences. Specifically, RTs were significantly longer when the information was inconsistent than when it was consistent with causal theories of the disorder with which the hypothetical client had been diagnosed.

In Experiment 3, advanced Psychology students read clinical reports providing information about the presence or absence of three symptoms considered as diagnostic criteria according to the DSM-IV. The students had been previously trained on causal theories of the mental disorders used in our experiment according to which the symptoms formed part of a causal chain of the sort S1→S2→S3. Inconsistencies were created by providing preliminary information about the disorder that the hypothetical client had been diagnosed with and by including a sentence stating the absence of one of the symptoms. The most compelling evidence of on-line causal reasoning was that RTs were slowed down when participants read a sentence stating the presence of a symptom when its causal antecedent was absent. This result indicated that the participants detected a causal inconsistency as, according to the supposedly entertained causal theory, given the absence of the causal antecedent, the following symptom in the causal chain should not have occurred. The pattern of diagnostic judgements conformed to the predictions from the causal model theory (Rehder, 2001, Rehder & Hastie, 2004, Rehder & Kim, 2010), according to which, diagnostic (categorical) judgements are the result of the computation of coherence. Finally, the fact that the inconsistency effects found in RTs were consistent with the participants' diagnostic judgements suggests that the computation of coherence relied upon System 1 processes of causal reasoning,

which took place during reading in an on-line manner. Such processes are also likely to have played an important role in the participants' judgements.

In Experiment 4, we searched for evidence supporting the idea that causal reasoning is intimately related to the notion of causal mechanism (Ahn & Bailenson, 1996; Ahn & Kalish, 2000; Ahn et al., 1995). To achieve this aim, Psychology students previously trained on the disorders used read clinical reports informing about the presence of three symptoms that, as in Experiment 3, formed part of a causal chain that, in this case, was made explicit. Two factors were orthogonally manipulated: the causal mechanism leading from one symptom to the next in the causal chain, and the diagnosis received by the client, which was stated at the beginning of the clinical report. The results found in RTs revealed that the participants relied on causal reasoning processes based on their beliefs about the causal mechanisms underlying the different disorders. This conclusion is supported by the longer RTs found for the sentences informing about bio-medical mechanisms compared with the sentences informing about the psychological mechanisms provided that the client had been diagnosed with a mental disorder. This result cannot be explained by invoking causal beliefs independent from the participants' beliefs about mental disorders because the difference between RTs was reverted when the hypothetical client had been diagnosed with a medical disease. The pattern of results found in the diagnostic judgement task was the same as in the reading task, which suggests again that the on-line and fast causal reasoning processes that took place during reading could have had a determinant impact on diagnostic judgements. Finally, the results from the treatment-decision making task corroborated that the descriptions given for the psychological and the medical mechanism were interpreted as such by the participants. Specifically, the psychological treatment received higher ratings in the psychological than in the medical mechanism condition, whereas the

reverse pattern of result was found regarding the ratings received by the medical treatment.

The pattern of results found in the present experimental series is very consistent in showing that both clinicians and Psychology students engage in very fast, on-line reasoning processes, which can be assimilated to Kahneman and Frederick's (2002) System 1, during fluent reading of clinical reports for a later diagnostic decision. Such reasoning processes involved fast retrieval of causal theories about mental disorders from memory as well as fast inference processes that were shown to be crucial for computing the coherence of the information provided. Such processes are likely to rely upon coherence-driven processes based on activation processes as those that are characteristic of dynamic neural networks. Interestingly, our experimental series provides converging evidence demonstrating that System 1 reasoning processes responsible for the computation of coherence rely on causal theories of mental disorders, which establish the way in which some symptoms give rise to others. The reasoners' reliance on causal theories through fast, on-line reasoning processes has been demonstrated through several manipulations that, taken together, have produced a strongly consistent pattern of results. Additionally, we have found a persistent consistency between the fast, on-line reasoning processes detected through on-line measures and the participants' diagnostic judgements after the reading task, except for the results from students in Experiment 2. This consistency suggests that very early and fast System 1 processes based on causal reasoning may have biased System 2 processes that could be at work during the diagnostic judgement task.

An issue that deserves some attention is that we could only find evidence of System 1 causal reasoning in students in Experiments 3 and 4 but not in Experiments 1

and 2. We could consider two possible explanations. On the one hand, the causal connections between symptoms in the clinical reports of Experiments 3 and 4 could be self-evident compared to the remaining experiments. Thus, the activation of causal features in a fast and on-line manner could have been promoted compared with Experiments 1 and 2. In fact, in the case of Experiments 3 and 4, students had been previously trained on the causal theories on which we based our manipulations. Of course, this is not to say that students and clinicians are equally skilled, but they could be, at least in Experiments 3 and 4, comparable in terms of the involvement of System 1 processes of causal reasoning. On the other hand, students might have been less committed to building a coherent mental model of the clinical reports for comprehension purposes. Thus, they might have detected inconsistencies through System 1 processes but, nonetheless kept on reading at a normal speed without dedicating extra time and resources to solve such inconsistencies. This explanation may well be especially pertinent regarding Experiment 2, where the students were shown to have been engaged in causal reasoning during the diagnostic judgement task.

This study leaves many interesting questions unresolved. One of them is the role of the clinicians' theoretical approach in causal reasoning. Given that the theories on which we based our manipulations come from the cognitive behavioral approach, one may expect to find that the effects of such manipulations are modulated by clinicians' theoretical approaches. Unfortunately, the number of clinicians with non-cognitive-behavioral approaches who participated in Experiment 1 and 2 was very small, precluding our assessment of this modulating role. However, future experiments may be conducted to answer this question. A related question is the source of clinicians' causal theories. Given that our manipulation was based on theories within the cognitive-behavioral framework (especially in Experiments 2 through 4) taught at academic

institutions, it is tempting to conclude that clinicians' and students' causal theories come from this academic instruction. However, it should be acknowledged that the disorders that were used to design the clinical reports are highly prevalent in the population and well known through the media. In this sense, it remains to be determined whether the same results hold for less known and less frequent mental disorders.

One limitation of Experiments 3 and 4 is that they do not allow to generalize the findings to the population of clinicians. It may be argued that if our experiments had been conducted with clinicians, they would have behaved more in accordance with the DSM-IV diagnostic criteria and prescriptions. In other words, clinicians' extended practice in the use of the DSM-IV might make them less prone to biases due to System 1 causal reasoning processes in the diagnosis of mental disorders. However, Experiments 1 and 2 suggest that clinicians are also subjected to biases produced by causal reasoning processes, which are likely to rely upon System 1, in the diagnosis of mental disorders from the DSM-IV. Additionally, previous studies with clinicians have provided evidence for the lack of adherence to previous versions of the DSM (see Davis et al., 1993; Garb, 1996; Rubinson et al., 1988). Furthermore, Kahneman (2011) pointed that, although expert intuition can certainly be very accurate, even experts remain at risk of generating and falling for spurious intuitions. Taking into account these considerations, we think that the results found in Experiment 3 and 4 are very likely to be replicated in a sample of clinicians. A new study aimed to assess this prediction would be very interesting and useful to get to know the processes responsible for clinicians' lack of adherence to the DSM-IV in some circumstances.

Finally, another concern about our study that may be raised is that causal reasoning could have been induced by some procedural aspects of the experiments that

may affect their ecological validity. Specifically, in Experiment 3, participants had to perform a causal judgement task, whereas, in Experiment 4, the clinical reports included information about the causal mechanism responsible for the development of the symptoms. It may be argued that neither of these circumstances is very common in clinicians' daily professional experience, although it is not easy to find evidence supporting this claim. The same claim may be made of many previous studies. Most of the experiments that have been conducted to assess the role of causal reasoning include procedural aspects that may be viewed as even more uncommon than our procedure in real environments. For example, in some studies, participants are provided with information about causal theories through verbal instructions or are asked to make explicit their own causal beliefs to assess the role of causal reasoning in tasks as diverse as diagnostic judgement (Kim and Ahn, 2002), diagnostic reasoning (Kim & Keil, 2003), treatment efficacy judgement (Kwaadsteniet et al., 2010; Yopchick & Kim, 2009), judgement of the need for psychological treatment (Kim & LoSavio, 2009), or information seeking (Kim, Yopchick, & Kwaadsteniet, 2008). It is largely known that the experimental approach to the study of cognitive processes, in general, tend to lack ecological validity. In our view, it is important to overcome this shortcoming in future research on causal reasoning in clinical tasks. Otherwise, it will remain to be unknown whether the effects of System 1 causal reasoning found in our experiments are the result of artificial laboratory preparations or a common tendency in clinical psychologists' environments.

The persistent tendency to use causal reasoning conflicts with both DSM-IV recommendations and with how clinicians are trained to use this resource. It therefore appears that clinicians' initial training regarding DSM-IV prescriptions of not relying on causal theories is not sufficient to prevent them from using causal reasoning when

performing clinical tasks. Our results suggest that one explanation for this difficulty may be that clinicians are not overtly aware of their use of causal theories. The involvement of System 1 causal reasoning that appears to have occurred during the reading task suggests that very rapid, efficient, and semiautomatic processes may have been at work. If such is the case, clinicians' training should be supplemented by training in causal reasoning that is aimed at describing the different and (occasionally) subtle ways in which it can influence judgements and decisions in the clinical context, especially in cases of patients who are potentially suffering from DSM-IV disorders. It is beyond the scope of the present study to assess whether causal reasoning helps clinicians or is a source of errors that should be avoided when dealing with the diagnosis of mental disorders. In any case, training in causal reasoning should help clinicians gain further control of their reasoning and decision making. Furthermore, according to Kahneman (2011), the more we know about the activities and biases of System 1, the more aware we will be of how this system works and how it influences and misleads System 2. In addition, System 2 can be trained to improve (e. g. calculating probabilities and statistics).

Our results also appear to have interesting implications for evidence-based clinical practice, specifically for the application of empirically supported treatments (EST). According to the American Psychological Association, ESTs are currently considered to be the best methods for addressing the treatment of mental disorders and patients' behavioral problems. Although ESTs are quite standardized, there is evidence demonstrating that clinicians have difficulties in following the indications that are prescribed in textbooks (Waller, 2009) and tend to adapt the treatments to either the patients' individual characteristics (McHugh, Murray & Barlow, 2009) or to the clinicians' case formulation, even when such formulations are not explicit or structured (Pain, Chadwick & Abba, 2008; Persons, 2006). Moreover, this tendency has been

considered to be inevitable by other clinicians (Persons, 2005). Our results suggest that causal theories, which appear to be readily available in the clinician's mind and used through System 1 processes, may play an important role in clinical case formulations (Eells, 2007). Such clinical case formulations would in turn be responsible for the difficulties that are experienced by clinicians when attempting to strictly follow the treatment protocol, especially when the theory on which the EST is based differs from the clinician's causal theory (Anderson & Strupp, 1996; Beutler, 1999). Thus, clinicians' application of ESTs may benefit from a certain degree of training in causal reasoning that is aimed to make clinicians aware of the different and subtle ways in which it can affect treatment decision making and treatment application.

REFERENCES

Agnoli, F. (1991). Development of judgemental heuristics and logical reasoning: training counteracts the representativeness heuristic. *Cognitive Development, 6,* 195-217. doi:10.1016/0885-2014(91)90036-D

Agnoli, F., & Krantz, D. H. (1989). Suppressing natural heuristics by formal instruction: The case of the conjunction fallacy. *Cognitive Psychology, 21*, 515–550. doi: 10.1016/0010-0285(89)90017-0

Ahn, W. & Bailenson, J. (1996). Mechanism-based explanations of causal attribution: An explanation of conjunction and discounting effect. *Cognitive Psychology, 31*, 82-123.

Ahn, W. K., & Kalish, C. W. (2000). The role of mechanism beliefs in causal reasoning. In F. C. Keil & R. A. Wilson (Eds.), *Explanation and cognition* (pp. 199–225). Cambridge, MA: MIT Press.

Ahn, W., Kalish, C. W., Medin, D. L., & Gelman, S.A. (1995). The role of covariation versus mechanism information in causal attribution. *Cognition, 54*, 299-352. doi: 10.1016/0010-0277(94)00640-7

Ahn, W., Kim, N. S., Lassaline, M. E., & Dennis, M. J. (2000). Causal status as a determinant of feature centrality. *Cognitive Psychology, 41,* 361-416. doi: 10.1006/cogp.2000.0741

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.

Anderson, T. & Strupp, H. H. (1996). The ecology of psychotherapy research. *Journal of Consulting and Clinical Psychology, 64,* 776–782. doi: 10.1037//0022-006X.64.4.776

Beck, A.T. (1967). *Depression: Clinical, experimental, and theoretical aspects*. New York: Harper & Row.

Beutler, L. E. (1999). Manualizing flexibility: the training of eclectic therapists. *Journal of Clinical Psychology, 55*, 399–404. doi: 10.1002/(SICI)1097-4679(199904)55:4<399::AID-JCLP4>3.3.CO;2-Q

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

Bless, H., Clore, G. L., Schwarz, N., Golisano, V., Rabe, C., & Wolk, M. (1996). Mood and the use of scripts: Does a happy mood really lead to mindlessness? *Journal of Personality and Social Psychology, 71*, 665-679. doi: 10.1037//0022-3514.71.4.665

Chaiken, S. & Trope, Y. (1999). *Dual-process theories in social psychology.* New York: Guilford.

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

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

Clark, D. M. & Wells, A. (1995). A cognitive model of social phobia. In R. Heimberg, M. Liebowitz, D. A. Hope, & F. R. Schneier (Eds.), *Social phobia: Diagnosis, assessment and treatment* (pp. 69–93). New York: Guilford Press.

Crisp, A. H. (1980). *Let Me Be*. London: Academic Press.

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. doi: 10.1037//0021-843X.102.2.319

de Kwaadsteniet, L., Hagmayer, Y., Krol, N. P. C. M., & Witteman, C. L. M. (2010). Causal client models in selecting effective interventions: A cognitive mapping study. *Psychological Assessment, 22,* 581–592. doi: 10.1037/a0019696

de Kwaadsteniet, L., Kim, N.S., & Yopchick, J.E. (2013). How do practising clinicians and students apply newly learned causal information about mental disorders? *Journal of Evaluation in Clinical Practice, 19,* 112-117. doi: 10.1111/j.1365-2753.2011.01781.x

Eells, T. D. (2007). *Handbook of psychotherapy case formulation*. New York: Guilford Press.

Einhorn, H. J. (1986). Accepting error to make less error. *Journal of Personality Assessment, 50*, 387-395. doi: 10.1207/s15327752jpa5003_8

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102,* 211-245. doi: 10.1037//0033-295X.102.2.211

Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgement, and social cognition. *Annual Review of Psychology, 59*, 255–278. doi: 10.1146/annurev.psych.59.103006.093629

Finucane, M. L., Alhakami, A., Slovic, P., & Johnson, S. M. (2000). The affect heuristic in judgements of risks and benefits. *Journal of Behavioral Decision Making, 13*, 1–17. doi: 10.1002/(SICI)1099-0771(200001/03)13:1<1::AID-BDM333>3.0.CO;2-S

Fugelsang, J., & Thompson, V. (2003). A dual-process model of belief and evidence interactions in causal reasoning. *Memory & Cognition, 31*, 800-815. doi: 10.3758/BF03196118

Garb, H. N. (1996). Taxometrics and the revision of diagnostic criteria. *American Psychologist, 51*, 553–554. doi: 10.1037//0003-066X.51.5.553

García-Retamero, R. & Hoffrage, U. (2006). How causal knowledge simplifies decision making. *Minds & Machines. Special Volume on Causality, Uncertainty and Ignorance, 16,* 365-380. doi: 10.1007/s11023-006-9035-1

Gilbert, D. T. (1999). What the mind's not. In S. Chaiken & Y. Trope (Eds.), *Dual Process Theories in Social Psychology* (pp. 3–11). New York: Guilford.

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

Hagmayer, Y. & Sloman, S. A. (2009). Decision makers conceive of their choices as intervention. *Journal of Experimental Psychology: General, 138*, 22-38. doi: 10.1037/a0014585

Hagmayer, Y., Sloman, S. A., Lagnado, D. A., & Waldmann, M. R. (2007). Causal reasoning through intervention. In A. Gopnik, A. & L. E. Schulz (Eds.), *Causal Learning: Psychology, Philosophy, and Computation* (pp. 86-100). Oxford: Oxford University Press.

Hammond, K. R. (1996). *Human Judgement and Social Policy*. New York: Oxford Univ. Press.

Haynes, S. N. & Williams, A. E. (2003). Case formulation and design of behavioral treatment programs: Matching treatment mechanisms to causal variables for behavior problems. *European Journal of Psychological Assessment, 19,* 164 –174. doi: 10.1027//1015-5759.19.3.164

Heit, E. (2000). Properties of inductive reasoning. *Psychonomic Bulletin & Review, 7*, 569-592. doi: 10.3758/BF03212996

Hinton, G. H. (1990). Mapping part-whole hierarchies into connectionist networks. *Artificial Intelligence, 46*, 47-75. doi: 10.1016/0004-3702(90)90004-J

Isen, A. M., Nygren, T. E., & Ashby, F. G. (1988). Influence of positive affect on the subjective utility of gains and losses: It is just not worth the risk. *Journal of*

Personality and Social Psychology, 55, 710-717. doi: 10.1037//0022-3514.55.5.710

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

Kahneman, D. & Frederick, S. (2002). Representativeness revisited: attribute substitution in intuitivem judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and Biases: The Psychology of Intuitive Judgment* (pp. 49–81). Cambridge, UK: Cambridge Univ. Press.

Kendell , R. E. (1973). Psychiatric diagnoses: A study of how they are made. *British Journal of Psychiatry, 122,* 437-445. doi: 10.1192/bjp.122.4.437

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, 39,* 854-865. 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

Kim, N. S. & Keil, F. C. (2003). From symptoms to causes: Diversity effects in diagnostic reasoning. *Memory & Cognition, 31,* 155-165. doi: 10.3758/BF03196090

Kim, N. S. & LoSavio, S. T. (2009). Causal explanations affect judgments of the need for psychological treatment. *Judgment and Decision Making, 4*, 82-91.

Kim, N. S., Yopchick, J. E., & de Kwaadsteniet, L. (2008). Causal diversity effects in information seeking. *Psychonomic Bulletin & Review, 15,* 81-88. doi: 10.3758/PBR.15.1.81

Kitsch, W., Patel, V. L., & Ericsson, K. A. (1999). The role of long-term working-memory in text comprehension. *Psychologia, 42*, 186-198.

Ladoceur, R., Blais, F., Freeston, M.H., & Dugas, M. J. (1998). Problem solving and problem orientation in generalized anxiety disorder. *Journal of Anxiety Disorders, 12,* 139-152. doi: 10.1016/S0887-6185(98)00002-4

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: 10.1037//0278-7393.27.6.1424

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

McHugh, R. K., Murray, H. W., & Barlow, D. H. (2009). Balancing fidelity and adaptation in the dissemination of empirically-supported treatments: The promise of transdiagnostic interventions. *Behavior Research and Therapy, 47,* 946-953. doi: 0.1016/j.brat.2009.07.005

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

Meder, B., Hagmayer, Y., & Waldmann, M. R. (2008). Inferring interventional predictions from observational learning data. *Psychonomic Bulletin & Review, 15,* 75-80. doi: 10.3758/PBR.15.1.75

Morewedge, C. K. & Kahneman, D. (2010). Associative processes in intuitive judgement. *Trends in Cognitive Sciences, 14,* 435-440. doi: 10.1016/j.tics.2010.07.004

Mowrer, O. A. (1947). On the dual nature of learning: A reinterpretation of 'conditioning and problem solving'. *Harvard Educational Review, 17,* 102-148.

Nisbett, R. E., Krantz, D. H., Jepson, C., & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. *Psychological Review, 90,* 339-363.

Pain, C. M., Chadwick, P., & Abba, N. (2008). Client's experience of case formulation in cognitive behavior therapy for psychosis. *British Journal of Clinical Psychology, 47*, 127-138. doi: 10.1348/014466507X235962.

Patel, V. L. & Groen, G. J. (1986). Knowledge-based strategies in medical reasoning. *Cognitive Science, 10*, 91-116. doi: 10.1016/S0364-0213(86)80010-6

Patel, V. L., Evans, D. A., & Groen, G. J. (1989). Biomedical knowledge and clinical reasoning. In D. A. Evans & V. L. Patel (Eds.), *Cognitive science in medicine: Biomedical modeling* (pp. 53-112). Cambridge, Mass: The MIT Press.

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. New York: Cambridge.

Penn, D. C. & Povinelli, D. J. (2007a). Causal cognition in human and nonhuman animals: A comparative, critical Review. *Annual Review of Psychology, 58*, 97–118. doi: 10.1146/annurev.psych.58.110405.085555

Peracchi, K. A. & O'Brien, E. J. (2004). Character profiles and the activation of predictive inferences. *Memory & Cognition, 32*. doi: 10.3758/BF03196880

Persons, J. B. (2005). Empiricism, mechanism, and the practice of cognitive-behavior therapy. *Behavior Therapy, 36*, 107–118. doi: 10.1016/S0005-7894(05)80059-0

Persons, J. B. (2006). Case formulation–driven psychotherapy. *Clinical Psychology: Science and Practice, 13*, 167–170. doi: 10.1111/j.1468-2850.2006.00019.x

Rehder, B. (2001). Interference between cognitive skills. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 451-469. doi: 10.1037//0278-7393.27.2.451

Rehder, B. (2003). Categorization as causal reasoning. *Cognitive Science, 27*, 709–748. doi: 10.1016/S0364-0213(03)00068-5

Rehder, B. (2010). Causal-based classification: A review. In B. Ross, (Ed.), *The Psychology of Learning and Motivation, 52,* 39-116.

Rehder, B. & Burnett, R. (2005). Feature inference and the causal structure of categories. *Cognitive Psychology, 50,* 264-314. doi: 10.1016/j.cogpsych.2004.09.002

Rehder, B., & Hastie, R. (2004). Category coherence and category-based property induction. *Cognition, 91*, 113-153. doi: 10.1016/S0010-0277(03)00167-7

Rehder, B. & Kim, S. W. (2006). How causal knowledge affects classification: A generative theory of categorization. *Journal of Experimental Psychology: Learning, Memory & Cognition, 32,* 659–683. doi: 10.1037/0278-7393.32.4.659

Rehder, B. & Kim, S. (2010). Causal status and coherence in causal-based categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 1171-1206. doi: 10.1037/a0019765

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 Mental Disease, 176*, 480–484. doi: 10.1097/00005053-198808000-00005

Salkovskis, P. M. (1985). Obsessional-compulsive problems: A cognitive-behavioral analysis. *Behavior Research and Therapy, 23*, 571-583. doi: 10.1016/0005-7967(85)90105-6

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

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

Shanks, D. (2010). Learning: From association to cognition. *Annual Review of Psychology, 61,* 273–301. doi: 10.1146/annurev.psych.093008.100519

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*, 3-22. doi: 10.1037//0033-2909.119.1.3

Sloman, S. A. & Lagnado, D. A. (2005). Do we 'do'? *Cognitive Science, 29,* 5–39. doi: 10.1207/s15516709cog2901_2

Sloman, S. A., Love, B. C., & Ahn, W. (1998). Feature centrality and conceptual coherence. *Cognitive Science*, *22,* 189-228. doi: 10.1016/S0364-0213(99)80039-1

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.

Stanovich, K. E. (1999).*Who is Rational? Studies of individual differences in reasoning*. Mahwah, NJ: Elrbaum.

Stanovich, K. E. & West, R. F. (2003). Evolutionary versus instrumental goals: how evolutionary psychology misconceives human rationality. In D. E. Over (Ed.), *Evolution and the Psychology of Thinking* (pp. 171–230). Hove, UK: Psychol. Press.

Steyvers, M., Tenenbaum, J. B., Wagenmakers, E. J., & Blum, B. (2003). Inferring causal networks from observations and interventions. *Cognitive Science, 27,* 453-489. doi: 10.1016/S0364-0213(03)00010-7

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.

Tversky, A. & Kahneman, D. (1977). Causal thinking in judgement under uncertainty. *Basic Problems in Methodology and Linguistics. The University of Western Ontario Series in Philosophy of Science*, 11, 167-190.

Tversky, A. & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgement. *Psychological Review, 90,* 293-315.

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

Waldmann, M. R. (1996). Knowledge-based causal induction. In D. R. Shanks, K. J. Holyoak, & D. L. Medin (Eds.), *The psychology of learning and motivation* (pp. 47–88). San Diego, CA: Academic Press.

Waldmann, M. R. & Hagmayer, Y. (2005). Seeing versus doing: Two modes of accessing causal knowledge. *Journal of Experimental Psychology: Learning, Memory and Cognition, 31,* 216-227. doi: 10.1037/0278-7393.31.2.216

Waldmann, M. R., & Hagmayer, Y. (2013). Causal reasoning. In D. Reisberg (Ed.), *Oxford Handbook of Cognitive Psychology*. New York: Oxford University Press.

Waldmann, M. R. & Holyoak, K. J. (1992). Predictive and diagnostic learning within causal models: Asymmetries in cue competition. *Journal of Experimental Psychology: General, 121*, 222-236. doi: 10.1037//0096-3445.121.2.222

Waldmann, M. R., Holyoak, K. J., & Fratianne, A. (1995). Causal models and the acquisition of category structure. *Journal of Experimental Psychology: General, 124,* 181–206. doi: 10.1037//0096-3445.124.2.181

Waller, G. (2009). Evidence-based treatment and therapist drift. *Behavior Research and Therapy, 47*, 119–27. doi: 10.1016/j.brat.2008.10.018

Westen, D. (2012). Prototype diagnosis of psychiatric syndromes. *World Psychiatry, 11,* 16-21. doi: 10.1016/j.wpsyc.2012.01.004

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

Yopchick, J. E. & Kim, N. S. (2009). The influence of causal information on judgements of treatment efficacy. *Memory & Cognition, 37*, 29-41. doi: 10.3758/MC.37.1.29

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

APPENDIX

Appendix A

Target sentences used to create central and peripheral inconsistencies in the experimental and control clinical reports in Experiment 1.

Diagnosis	Target sentences	
	Absence of the central symptom	Absence of the peripheral symptom
Major Depressive Disorder	"He has been in a good mood lately"	"He has maintained his weight during all this time"
Avoidant Personality Disorder	"He is convinced of being interesting, competent and appealing"	"He gets easily involved in new activities"
Borderline Personality Disorder	"He has an adequate and stable identity sense"	"He does not have auto-destructive behaviors or thoughts"
Schizophrenia	"His sensorial perception is in accordance with the reality"	"He expresses in a clear and coherent way"
Specific Phobia (Claustrophobia)	"He does not have exaggerated or irrational reactions in anticipation of specific situations"	"He thinks his reactions are reasonable and adjusted to the situation that cause them"
Antisocial Personality Disorder	"He is a sincere and honest person"	"He has just been 17 years old"

Appendix B

Example of the four types of clinical reports used in Experiment 1.

Example of inconsistent clinical report whose symptom used to the elaboration of the target sentence is a central symptom of the Major Depressive Disorder:

P. went to see a clinical psychologist to request some help. After an exhaustive assessment process, he was diagnosed with Major Depressive Disorder. He goes to consultation wearing blue jeans and a simple t-shirt. Since some weeks he cannot avoid to have excessive feelings of guilt almost every day. He tells that he has got as pet a pedigree dog that was given to him recently. P. tells not having any pleasure in the activities that he does in his daily life the great part of time. He lives in a semidetached house in a calm place in the suburbs of the city. He has sleep problems, like insomnia, and he wakes up easily several times during night. Some difficulties in focusing attention have exerted a negative influence in his job. His problem affects his relationships with people around him. P. likes computers very much and computers have much to do with his job. He works as a computer programmer in an important company in his city. He has no problem with getting to work as he lives very near his job. It is his office where he spends most of the time when he is outside home. He has been in a good mood lately. The clinician states that giving suitable psychological care to the patient would be important.

Example of control clinical report whose symptom used to the elaboration of the target sentence is a central symptom of the Major Depressive Disorder:

D. decided to see a clinical psychologist because he suffered from some problems. After an exhaustive assessment process, he was diagnosed with Cannabis Dependence Disorder. He goes to consultation wearing informal clothes and back sport shoes. He needs bigger and bigger amounts of the substance to get the wanted effects. He tells he has got a big collection of tin figures painted by hand. D. has to have cannabis to avoid the abstinence symptoms. He has lived in a shared rent flat with a friend for long time. He has been having the substance more time than he expected and in bigger amounts. Some difficulties in focusing attention have exerted a negative influence in his job. His

problem affects his relationships with people around him. D. works part-time in a vehicle repair shop for two weeks ago. He goes to his job by bus every day because it is far away. He does not work long hours at day, but he spends his free time in something useful at least. He has always liked cars and working with them motivates him. He has been in a good mood lately. The clinician states that giving suitable psychological care to the patient would be important.

Example of inconsistent clinical report whose symptom used to the elaboration of the target sentence is a peripheral symptom of the Major Depressive Disorder:

G. went to see a clinical psychologist to request some help. After an exhaustive assessment process, he was diagnosed with Major Depressive Disorder. He goes to consultation wearing blue jeans and a simple t-shirt. Since some weeks he cannot avoid to have excessive feelings of guilt almost every day. He tells he has got a garden where he cultivates his own fruits and vegetables. G. tells not having any pleasure in the activities that he does in his daily life the great part of time. He lives in a semidetached house with his elder brother and his wife. He has sleep problems, like insomnia, and he wakes up easily several times during night. Some difficulties in focusing attention have exerted a negative influence in his job. His problem affects his relationships with people around him. G. has got a small place to sell fruits in the market of his city. Daily he has his family's help when he has more work. This is not a hard work and it let to know and be in contact with neighbors. All his products are varied, of good quality and are sold well. He has maintained his weight during all this time. His psychologist hopes a good collaboration.

Example of control clinical report whose symptom used to the elaboration of the target sentence is a peripheral symptom of the Major Depressive Disorder:

F. decided to see a clinical psychologist because he suffered from some problems. After an exhaustive assessment process, he was diagnosed with Cannabis Dependence Disorder. He goes to consultation wearing baggy trousers and a waistcoat of the same color. He needs bigger and bigger amounts of the substance to get the wanted effects. He tells he has got a German shepherd that was given to him in his birthday. F has to have cannabis to avoid the abstinence symptoms. He lives in a small flat in the suburbs

since he started his studies in the high school. He has been having the substance more time than he expected and in bigger amounts. Some difficulties in focusing attention have exerted a negative influence in his job. His problem affects his relationships with people around him. F. works some hours at week in a carpenter's workshop to make up for his expenses. He shares his job with three partners and the boss. He makes wardrobes, tables, all type of chairs and tapestries. It is an easy work and it has not been very hard to learn the basic things of the job. He has maintained his weight during all this time. His psychologist hopes a good collaboration.

Appendix C

Target sentences of consistent and inconsistent reports in Experiment 2.

Diagnosis	Target sentences	
	Consistent report	Inconsistent report
Anorexia Nervosa	"At first I started to get distressed about the possibility of becoming fat " "Some time later, I started to refuse to eat " "Eventually, my period stopped"	"At first, my period stopped " "Some time later, I started to refuse to eat " "Eventually, I have started to get distressed about the possibility of becoming fat "
Major Depressive Disorder	"At first I started to feel very bad with myself because all things happened in my family" "Some time later, I started to feel apathy for everything" "Eventually, I have started to have problems to sleep"	"At first I started to have problems to sleep" " Some time later, I started to feel apathy for everything " "Eventually, I have started to feel very bad with myself because all things happened in my family "
Anxiety Generalized Disorder	"At first I started to have continues worries that I didn't know how to control" "Some time later, I started to feel breathlessness, unease and palpitations that appeared in any moment" "Eventually, I have started to have sensations of tiredness and lack of energy"	"At first I started to have sensations of tiredness and lack o energy" "Some time later, I started to feel breathlessness, unease and palpitations that appeared in any moment" "Eventually, I have started to have continues worries that I don't know how to control "
Obsessive Compulsive Disorder	"At first I started to worry continually about seeing any thing dirty" "Some time later, I started to not stop cleaning and checking everything is clean" "Eventually, I have started to feel unease because no time to family and work"	"At first I started to feel unease because no time to family a work" "Some time later, I started to not stop cleaning and checkin everything is clean" "Eventually, I have started to worry continually about seein any thing dirty"
Posttraumatic Stress Disorder	"At first I started to be very nervous about the car accident that I had" "Some time later, I started to try everything to not think about the car accident that I had" "Eventually, I have started to feel unease because I feel discouraged and unable to drive a car"	"At first I started to feel unease because I feel discouraged a unable to drive a car" "Some time later, I started to try everything to not think abo the car accident that I had" "Eventually, I have started to be very nervous about the car accident that I had"
Specific Phobia	"At first I started to feel a lot of worries and unease when I meet a dog" "Some time later, I started to try not to walk for places like parks and residential areas" "Eventually, I have started to feel worry and unease about the idea of leaving home"	"At first I started to feel worry and unease about the idea of leaving home" "Some time later, I started to try not to walk for places like parks and residential areas" "Eventually, I have started to feel a lot of worries and uneas when I meet a dog"

Appendix D

Example of the two types of clinical reports used in Experiment 2.

Example of consistent causal clinical report, where the order of the symptoms is coherent to the order proposed by the theory of the Anorexia Nervosa Disorder and the causal connection between symptoms is showed:

P. decided to see a clinical psychologist because she suffered from some problems. After the assessment process, she was diagnosed with Anorexia Nervosa. In what follows, it is provided some of P.'s verbal expressions that the clinician considered relevant. "At first I started to get distressed about the possibility of becoming fat"."Some time later, I started to refuse to eat"."Eventually, my period stopped". After the assessment process, a second clinician found that the events mentioned were strongly related, so that the distress about the possibility of becoming fat originated the refuse to eat which, in turn, made the period stop in the long run.

Example of inconsistent no-causal clinical report, where the order of the symptoms is incoherent to the order proposed by the theory of the Anorexia Nervosa Disorder and non causal connection between symptoms is showed:

S. decided to see a clinical psychologist because she suffered from some problems. After the assessment process, she was diagnosed with Anorexia Nervosa. In what follows, it is provided some of S.'s verbal expressions that the clinician considered relevant. "At first, my period stopped"." Some time later, I started to refuse to eat"."Eventually, I have started to get distressed about the possibility of becoming fat". After the assessment process, a second clinician could not find any relationship between the events mentioned: the distress about the possibility of becoming fat, the refuse to eat, and the interruption of the period.

Appendix E

Questions about treatments used in Experiment 2.

ANOREXIA NERVOSA:

"*To what extent do you think that a progesterone-based hormonal treatment will resolve, in the short-, medium-, or long-term, the following problems?*"

"To what extent do you think that a treatment focused in comparing the perception of the figure drawing with a real and objective vision of it will resolve, in the short-, medium-, or long-term, the following problems?"

"*To what extent do you think that a treatment that face with the possibility of getting weight, teaching to face with consequences, will resolve, in the short-, medium-, or long-term, the following problems?*"

MAJOR DEPRESSIVE DISORDER:

"*To what extent do you think that a treatment benzodiazepines will resolve, in the short-, medium-, or long-term, the following problems?*"

"*To what extent do you think that a treatment focused in looking for motivation for hobbies and daily activities will resolve, in the short-, medium-, or long-term, the following problems?*"

"*To what extent do you think that a treatment focused in the attributive style of the patient and self-esteem will resolve, in the short-, medium-, or long-term, the following problems?*"

ANXIETY GENERALIZED DISORDER:

"*To what extent do you think that a medical treatment with dietary supplement and energetic will resolve, in the short-, medium-, or long-term, the following problems?*"

"*To what extent do you think that a treatment based on control of activation, as relaxing and breathing activities, will resolve, in the short-, medium-, or long-term, the following problems?*"

"*To what extent do you think that a treatment based on techniques of stop and changing of thoughts will resolve, in the short-, medium-, or long-term, the following problems?*"

OBSESSIVE COMPULSIVE DISORDER:

"*To what extent do you think that a treatment focused in learning guidelines of time organization and management will resolve, in the short-, medium-, or long-term, the following problems?*"

"*To what extent do you think that a treatment of exposition with response prevention will resolve, in the short-, medium-, or long-term, the following problems?*"

"*To what extent do you think that a treatment focused in intrusive-thoughts control techniques will resolve, in the short-, medium-, or long-term, the following problems?*"

POSTTRAUMATIC-STRESS DISORDER:

"*To what extent do you think that treatment based in strengthen self-esteem and in objective demonstration of own capacities will resolve, in the short-, medium-, or long-term, the following problems?*"

"*To what extent do you think that a treatment focused in exposition to specific thoughts and in thoughts control techniques will resolve, in the short-, medium-, or long-term, the following problems?*"

"*To what extent do you think that a treatment focused in learning relaxing and breathing techniques will resolve, in the short-, medium-, or long-term, the following problems?*"

SPECIFIC PHOBIA:

"*To what extent do you think that a treatment focused in progressive exposition to outside situations (cinema, bars, shops) will resolve, in the short-, medium-, or long-term, the following problems?*"

"*To what extent do you think that a treatment focused in progressive exposition to situations where there is a high probability of dogs (parks, residential places, veterinary) will resolve, in the short-, medium-, or long-term, the following problems?*"

"*To what extent do you think that a treatment focused in progressive presence and interaction with dogs will resolve, in the short-, medium-, or long-term, the following problems?*"

Appendix F

Target sentences of the reports used in Experiment 3.

Diagnosis	Target sentences		
	Patient 1	Patient 2	Patient 3
Depression	"**She thinks that all good things happen to her wherever**" "She never feels like going out or seeing her friends" "She is socially isolated and has many social problems"	"She thinks that all bad things happen to her wherever" "**She often feels like going out or seeing her friends**" "She is socially isolated and has many social problems"	"She thinks that all bad things happen to her wherever" "She never feels like going out or seeing her friends" "**She is socially integrated and has no social problems**"
Specific phobia	"**He had sweet experiences with dogs when he was a child**" "He feels very anxious when he is close to a dog" "He very rarely spends time near pet shops or parks"	"He had bad experiences with dogs when he was a child" "**He feels very comfortable when he is close to a dog**" "He very rarely spends time near pet shops or parks"	"He had bad experiences with dogs when he was a child" "He feels very anxious when he is close of a dog" "**He very often spends time near pet shops or parks**"
Obsessive-compulsive disorder	"**He is left indifferent to bacterial infections**" "He washes his hands about 40 times per day" "He has strong problems at the workplace because of lack of time"	"He is very anxious to bacterial infections" "**He washes his hands about 4 times per day**" "He has strong problems at the workplace because of lack of time"	"He is very anxious to bacterial infections" "He washes his hands about 40 times per day" "**He can solve problems at the workplace because of disposition of time**"

Note. Translated from German.

Appendix G

Example of the three types of clinical reports used in Experiment 3.

Example of clinical report in the Patient 1 condition, where the absent symptom was S1. The symptoms formed part of a causal chain model that participants were expected to entertain: S1→S2→S3.

Jan goes to therapy because he's worried about his problem. After an evaluation process, Jan receives a diagnosis of specific phobia. He had sweet experiences with dogs when he was a child. He feels very anxious when he is close to dogs. He stops very rarely near to pet shops or parks. His psychologist tells him that needs his collaboration.

Example of clinical report in Patient 2, where the absent symptom is S2 in a causal chain S1→S2→S3:

Sarah needs professional help for her problem. After of evaluation, Sarah receives a diagnosis of specific phobia. She had bad experiences with dogs when was a child. She feels very comfortable when is close to some dog. She stops very rarely near to pet shops or parks. Both decide to begin the treatment the next session.

Example of clinical report in Patient 3, where the absent symptom is S3 in a causal chain S1→S2→S3:

Stefan considers that has a problem and goes to therapy. After of evaluation, Stefan receives a diagnosis of specific phobia. He had bad experiences with dogs when was a child. He feels very anxious when is close of some dog. He stops very frequently near to pet shops or parks. His psychologist says collaboration of both will be necessary.

Appendix H

Target sentences of the four types of clinical reports used in Experiment 4.

	Target sentences	
	Mechanism Psychological	Mechanism Bio-medical
Diagnosis Psychological: ANOREXIA	"The diagnosis made by a professional was anorexia nervosa" "Firstly, the patient showed a **high anxiety about the possibility of gaining weight**" "The high anxiety caused aversion to food" "The aversion to food produced an **important loss of weight**" "The loss of weight caused a strong hormonal change" "The hormonal change produced **amenorrhea**"	"The diagnosis made by a professional was anorexia nervosa" "Firstly, the patient showed a **high anxiety about the possibility of gaining weight**" "The high anxiety caused a stomach bleeding problem" "The stomach problem produced an **important loss of weight**" "The loss of weight led to the use of pharmacological digestive treatment" "The pharmacological treatment produced **amenorrhea**"
Diagnosis Medical: STOMACH ULCER	"The diagnosis made by a professional was stomach ulcer" "Firstly, the patient showed a **high anxiety about the possibility of gaining weight**" "The high anxiety caused aversion to food" "The aversion to food produced an **important loss of weight**" "The loss of weight caused a strong hormonal change" "The hormonal change produced **amenorrhea**"	"The diagnosis made by a professional was stomach ulcer" "Firstly, the patient showed a **high anxiety about the possibility of gaining weight**" "The high anxiety caused a bleeding stomach problem" "The stomach problem produced an **important loss of weight"** "The loss of weight caused the use of pharmacological digestive treatment" "The pharmacological treatment produced **amenorrhea**"
Diagnosis Psychological: GENERALIZED ANXIETY DISORDER	"The diagnosis made by a professional was generalized anxiety disorder" "Firstly, the patient experienced some **difficulty to deal with permanent worries"** "The difficulty to deal with permanent worries caused a permanent alert state" "The permanent alert state produced **easy fatigability, muscular tension, and restlessness**" "The fatigability, muscular tension, and restlessness led to the ingestion of anxiolytics" "The use of anxiolytics produced **sleep alterations**"	"The diagnosis made by a professional was generalized anxiety disorder" "Firstly, the patient experienced some **difficulty to deal with permanent worries"** "The difficulty to deal with permanent worries caused strong headaches" "The strong headaches produced **easy fatigability, muscular tension, and restlessness**" "The fatigability, muscular tension, and restlessness led to use symptomatic treatment for migraine" "The migraine symptoms produced **sleep alterations**"
Diagnosis Medical: MIGRAINE	"The diagnosis made by a professional was migraine" "Firstly, the patient experienced some **difficulty to deal with permanent worries"** "The difficulty to deal with permanent worries caused a permanent alert state" "The permanent alert state produced **easy fatigability, muscular tension, and restlessness**" "The fatigability, muscular tension, and restlessness led to the ingestion of anxiolytics" "The use of anxiolytics produced **sleep alterations**"	"The diagnosis made by a professional was migraine" "Firstly, the patient experienced some **difficulty to deal with permanent worries"** "The difficulty to deal with permanent worries caused strong headaches" "The strong headaches produced **easy fatigability, muscular tension, and restlessness**" "The fatigability, muscular tension, and restlessness led to use symptomatic treatment for migraine" "The migraine symptoms produced **sleep alterations**"

Diagnosis Psychological: SOCIAL PHOBIA	"The diagnosis made by a professional was social phobia" "Firstly, the patient suffered afraid to feel ashamed **in social situations"** "The fear of feeling embarrassing caused inability to control the anxiety" "The inability to control the anxiety produced **avoidance of social situations"** "The avoidance of social situations caused repeated absences from work" "The repeated absences from work produced **working problems**"	"The diagnosis made by a professional was social phobia" "Firstly, the patient suffered afraid to feel ashamed **in social situations "** "The fear of feeling embarrassing caused deterioration of his skin marks" "The deterioration of his skin marks produced **avoidance of social situations"** "The avoidance of social situations caused repeated absences from work for his disease" "The repeated absences from work for his disease produced **working problems"**
Diagnosis Medical: SKIN PROBLEM	"The diagnosis made by a professional was skin problem" "Firstly, the patient suffered afraid to feel ashamed **in social situations** "The fear of feeling embarrassing caused inability to control the anxiety" "The inability to control the anxiety produced **avoidance of social situations"** "The avoidance of social situations caused repeated absences from work" "The repeated absences from work produced **working problems**"	"The diagnosis made by a professional was skin problem" "Firstly, the patient suffered afraid to feel ashamed **in social situations"** "The fear of feeling embarrassing caused deterioration of his skin marks" "The deterioration of his skin marks produced **avoidance of social situations"** "The avoidance of social situations caused repeated absences from work for to his disease" "The repeated absences from work for his disease produced **working problems"**

Abstract in Spanish

Resumen de la Tesis Doctoral titulada:

Razonamiento causal en el diagnóstico de trastornos mentales:

Evidencia de medidas online y offline

El conocimiento causal constituye una herramienta altamente eficaz para adaptarnos al medio en el que vivimos, ya que nos permite realizar predicciones precisas acerca de acontecimientos y conductas, aprender y actuar en consecuencia. El conocimiento causal está presente en multitud de procesos cognitivos, tales como categorización, representación conceptual, toma de decisiones, comprensión de textos, intervención, realización de inferencias, etc. Muchas de las teorías que explican el razonamiento causal se han desarrollado en condiciones de laboratorio artificiales (Penn & Povinelli, 2007; Shanks, 2010). No obstante, algunas de estas teorías se han centrado en dominios más naturales y aplicados, como el de tareas llevadas a cabo en contextos clínicos (de Kwaadsteniet, Hagmayer, Krol y Witteman, 2010; de Kwaadsteniet, Kim y Yopchick, 2013; Einhorn, 1986; Haynes y Williams, 2003; Kim y Keil, 2003; Kim y LoSavio, 2009; Rehder y Kim, 2006; Yopchick y Kim, 2009). En este trabajo, llevamos a cabo cuatro experimentos centrándonos en el papel del razonamiento causal en un contexto clínico específico: el diagnóstico de trastornos mentales.

El uso del razonamiento causal en el contexto clínico específico del diagnóstico de trastornos mentales no debería resultar sorprendente si tenemos en cuenta que los procesos cognitivos donde el conocimiento causal está presente, tales como los comentados previamente, están también evidentemente presentes en tareas de toma de decisión diagnóstica. Sin embargo, si tenemos en cuenta la concepción de la

clasificación de trastornos mentales que propone el Manual Diagnóstico y Estadístico de Trastornos Mentales (DSM-IV, American Psychiatric Association, 2000), la influencia del razonamiento causal en el diagnóstico sí debería resultar sorprendente. La taxonomía de trastornos mentales que este manual propone no está basada en consideraciones causales, sino que intenta ser ateórico o, al menos, neutro con respecto a las diferentes aproximaciones teóricas a las que los clínicos puedan adherirse. De hecho, según el DSM-IV, las clasificaciones están basadas en criterios diagnósticos que no son necesarios ni suficientes. De esta forma, si el razonamiento causal se pone en marcha en el diagnóstico de trastorno mentales a pesar de las prescripciones de este manual, podríamos decir que se está cometiendo un sesgo cognitivo que podríamos definir como *sesgo causal*.

Kim y Ahn (2002) encontraron que tanto estudiantes de Psicología como expertos clínicos dependían de sus teorías causales idiosincráticas cuando se les pedía que realizaran juicios diagnósticos. En estos experimentos, los participantes tenían que dibujar mapas causales de diferentes trastornos relacionando mediante flechas los diferentes síntomas, considerados criterios diagnósticos del DSM-IV, así como asignar una puntuación de la fuerza causal que atribuían a cada relación. El objetivo de esta tarea fue obtener la representación causal de los trastornos de los participantes, así como una puntuación causal de cada criterio diagnóstico. Unos días después, los participantes realizaban una tarea de diagnóstico en la cual se les presentaba una serie de casos clínicos que contenían los síntomas que ellos habían considerado como los más causalmente centrales y los más causalmente periféricos. Los resultados de este experimento mostraron que los participantes estaban más dispuestos a diagnosticar a un paciente hipotético con un determinado trastorno si éste poseía síntomas causalmente centrales que si poseía síntomas causalmente periféricos. Además, su recuerdo de la

información relacionada con síntomas también estuvo influido por su sesgo causal. Pero, aunque ésta es la única evidencia del sesgo causal en el diagnóstico de trastornos mentales de la que tenemos conocimiento hasta ahora, no podemos descartar explicaciones alternativas a estos resultados, como la distribución estadística de los síntomas dentro del trastorno, su centralidad conceptual o su valor estadístico.

Ya que en el estudio de Kim y Ahn (2002) las respuestas de los participantes estaban basadas en juicios, es posible que el sesgo causal fuera fruto de la implicación de procesos de razonamiento controlados y deliberados, pero no podemos estar seguros de que procesos menos controlados y deliberados pudieran haberse puesto también en marcha. Si éste fuera el caso, cualquier intento de evitar o controlar el uso del sesgo causal en el diagnóstico de trastornos mentales debería comenzar por detectar tales procesos menos controlados, explicándolos y entendiendo cómo funcionan. Estas ideas estarían en línea con las típicas *teorías de proceso dual* (Chaiken y Trope, 1999; Hammond, 1996; Sloman, 1996), las cuales se han centrado en gran medida en la explicación de sesgos en razonamiento y tareas de juicios.

Los procesos cognitivos de alto nivel incluyen diferentes procesos, tales como pensamiento, razonamiento, toma de decisiones y juicios. Estos procesos cognitivos pueden ser divididos en lo que tradicionalmente denominamos *intuición* y *razón,* y que recientemente aceptamos bajo el marco de teorías del procesamiento dual. Estas teorías establecen una diferenciación entre procesos que son rápidos, inconscientes, automáticos, y procesos que son lentos, deliberados y que exigen un mayor esfuerzo. Kahneman y Frederick (2002) y Stanovich (1999) empezaron a usar unos términos neutrales para estas dos diferentes formas de procesamiento: *sistema 1* y *sistema 2*, respectivamente. El sistema 1 actúa por impulsos naturales y ofrece una evaluación

continua de los problemas que el individuo debe resolver para sobrevivir de una forma rápida y eficaz, aunque pudiendo llevar a posibles errores. No obstante, frecuentemente el sistema 2 monitoriza al sistema 1 y se moviliza automáticamente cuando este último encuentra alguna dificultad para resolver la situación, pudiendo llegar a evitar esos errores. Sin embargo, el sistema 2 no es consciente de que a menudo está siendo influenciado por el sistema 1, pudiendo conducirle a la comisión de sesgos y errores.

Por tanto, los errores en los juicios pueden ser atribuidos tanto al sistema 1 como al sistema 2. Frecuentemente, estos errores están basados en el uso de razonamiento causal. Por tanto, el razonamiento causal puede estar relacionado no sólo con procesos del sistema 2, sino también con procesos del sistema 1, por lo que el sesgo causal en el diagnóstico de trastornos mentales puede ser el resultado de procesos de razonamiento causal del sistema 1. Pero, ¿cómo podemos estar seguros de que el sesgo encontrado en el estudio de Kim y Ahn (2002) es el resultado de procesos del sistema 1 más que de procesos del sistema 2? Los participantes en este estudio tenían la oportunidad de poner en marcha procesos de pensamiento altamente demandantes, lentos y deliberados, ya que sus respuestas se basaban en la elaboración de juicios diagnósticos explícitos. Sin embargo, el razonamiento causal puede aparecer también en tareas que no sean de toma de decisiones o juicios. Si el sistema 1 está implicado en el sesgo causal, el razonamiento causal debería aparecer en procesos eficientes y rápidos, de una manera más automática y sin tiempo dedicado exclusivamente al pensamiento deliberado. Por tanto, para mostrar la implicación de procesos del sistema 1 en la producción del sesgo causal en el diagnóstico de trastornos mentales, deberíamos ser capaces de ofrecer evidencia de razonamiento causal rápido y online como parte de los procesos de comprensión que tienen lugar en el mismo momento en el que los razonadores reciben

la información sobre los síntomas. Las técnicas y procedimientos online usados en comprensión de textos son especialmente adecuadas para este propósito.

Procesos de razonamiento causal rápidos y automáticos atribuibles al sistema 1 parecen subyacer a la comprensión de textos. Un lector realiza inferencias durante la lectura con el objetivo de crear y mantener representaciones coherentes del texto tanto a un nivel global como local, y al menos algunas de estas inferencias son causales (Black y Bower, 1980; Kendeou, Smith, y O'Brien, 2012; Schank, 1975; Trabasso y Sperry, 1985; Trabasso y van den Broek, 1985). Por tanto, se puede esperar razonamiento causal del sistema 1 cuando los clínicos leen informes para tomar una decisión diagnóstica posterior. En función de la información ofrecida, se pueden activar teorías específicas, creencias o expectativas sobre síntomas adicionales. Si la información siguiente es coherente con la información ya recibida y la información causal activada, puede ser fácilmente integrada en un modelo mental y podría ser percibido como plausible. Por el contrario, si es incoherente, se requeriría un esfuerzo cognitivo para resolver la inconsistencia y para integrar la nueva información en una estructura unificada (Ericsson y Kintsch, 1995; Kintsch, Patel, y Ericsson, 1999). Si no se encontrara solución, la información sería considerada implausible. Algo que se debe destacar de todo ello es que todas estas inferencias tendrían que operar de una manera rápida y online.

Ya que los procesos de razonamiento causal atribuibles al sistema 1 parecen subyacer a la comprensión de textos y a la realización de inferencias online durante la lectura, usamos una metodología basada en comprensión lectora. En concreto, empleamos el *paradigma de la inconsistencia* (Albrecht y O'Brien, 1993; Long y Chong, 2001; Peracchi y O'Brien, 2004). Según este paradigma, si en un texto

presentamos información que resulta inconsistente con respecto a una información preliminar, el tiempo de lectura de esas frases será mayor que el tiempo de lectura de esas mismas frases en un texto donde no resultara inconsistente con diferente información preliminar. Por tanto, la detección de inconsistencia durante la lectura de un texto quedaría reflejada en el incremento en el tiempo de lectura, el cual tiende a emplearse en tratar de resolver la inconsistencia detectada. Nosotros usamos este paradigma para detectar procesos de razonamiento causal online, rápidos y semiautomáticos, durante la lectura de informes clínicos. Las inconsistencias en los informes clínicos las creamos incluyendo información sobre los síntomas de un paciente hipotético que contradecía de algún modo cierta información presentada al comienzo del informe y que estaba relacionada con el diagnóstico que ese paciente había recibido. La detección de tales inconsistencias se basó en la idea de que los "diagnosticadores" poseen teorías o creencias sobre trastornos mentales de acuerdo con las cuales los síntomas de tales trastornos conforman una red causal con una estructura específica.

Concretando, el objetivo de este trabajo fue mostrar que el diagnóstico de trastornos mentales no se atiene de forma estricta a los criterios diagnósticos del DSM-IV, sino que está sesgado por teorías y creencias causales. Además, queríamos mostrar que, en este sesgo causal, tienen una implicación importante procesos del sistema 1. Para estudiar la implicación de procesos del sistema 1 en el sesgo causal, se utilizó una técnica de medida online basada en el paradigma de la inconsistencia en informes clínicos. Pudimos detectar los procesos de razonamiento causal dependientes del sistema 1 a partir de los efectos que estas inconsistencias tenían sobre los tiempos de lectura de los participantes. Si el sesgo causal se ponía en marcha, se emplearía un mayor tiempo de lectura en los informes que poseían inconsistencias que en los informes control. Al mismo tiempo, en todos los experimentos, evaluamos la posible

incidencia de estos procesos de razonamiento causal sobre los juicios diagnósticos que los participantes emitían una vez leído el informe clínico. Estos juicios diagnósticos consistían en decidir en qué medida se estaba de acuerdo con un diagnóstico ofrecido, empleando para ello una escala en la que el participante debía posicionarse. De esta forma, si el sesgo causal se ponía en marcha, las puntuaciones en grado de acuerdo con el juicio diagnóstico serían menores en los diagnósticos de los informes con inconsistencia que en los informes control.

A continuación, presentamos los experimentos que se llevaron a cabo en este trabajo. En todos ellos, el estudio de los procesos de razonamiento causal estuvo basado en la detección de procesos de razonamiento rápidos y online durante la lectura de informes clínicos. No obstante, la manipulación de la inconsistencia y de la información de tipo causal se llevó a cabo de maneras diferentes.

Experimento 1

Los objetivos del Experimento 1 fueron: 1) replicar el efecto de estatus causal de Kim y Ahn (2002) y 2) demostrar que tal sesgo causal podía ser consecuencia de procesos de razonamiento causal atribuibles al sistema 1. A partir de los mapas causales sobre los síntomas de diferentes trastornos que dibujaron los participantes del estudio de Kim y Ahn (2002), calculamos una puntuación en centralidad causal de cada síntoma de cada trastorno empleado, lo cual nos permitió seleccionar el síntoma con mayor estatus causal, o lo que también podríamos denominar como el síntoma causalmente más central, y el síntoma con menor estatus, o el síntoma causalmente más periférico. Estos dos síntomas de diferentes trastornos fueron empleados para elaborar las frases target de los informes clínicos que usamos. En la condición inconsistente, esta frase resultaría inconsistente con el diagnóstico ofrecido al principio del informe y, en la condición

control, esta frase no resultaría inconsistente. En la medida en que esperábamos que los sujetos pusieran en marcha una teoría causal del trastorno, esperábamos encontrar un mayor efecto de inconsistencia asociado al uso del síntoma central, comparado con el uso del síntoma periférico. Al mismo tiempo, esperábamos encontrar un mayor impacto del síntoma central, comparado con el periférico, en juicios diagnósticos.

En el Experimento 1 participaron 17 estudiantes de psicología de la Universidad de Málaga y 17 clínicos, estos últimos con 17 años de experiencia media y en su mayoría cognitivo-conductuales. Todos ellos leían 24 informes clínicos: 12 inconsistentes y 12 control, en dos sesiones diferentes. En los informes inconsistentes se les presentaba información preliminar relacionada con el diagnóstico de los trastornos: depresión mayor, fobia específica, trastorno de personalidad antisocial, esquizofrenia, trastorno de personalidad límite y trastorno de personalidad evitativo. En los informes control se les presentaba información preliminar relacionada con el diagnóstico de los trastornos: dependencia de cannabis, sonambulismo, juego patológico, trastorno orgásmico, trastorno de identidad de género e hipocondría. La frase target, en la que se centraba especialmente la medición del tiempo de lectura, consistía en una frase hacia el final del informe en la que se presentaba información aludiendo a la ausencia de un síntoma causal o un síntoma periférico de alguno de los trastornos mencionados para los informes inconsistentes. De esta forma, si esta frase aparecía tras una información preliminar en la que se informaba de un diagnóstico de depresión mayor, fobia específica, trastorno de personalidad antisocial, esquizofrenia, trastorno de personalidad límite o trastorno de personalidad evitativo, resultaría incoherente y los tiempos de lectura serían mayores. Si por el contrario esa misma frase target aparecía tras una información preliminar en la que se informara de un diagnóstico de dependencia de

cannabis, sonambulismo, juego patológico, trastorno orgásmico, trastorno de identidad de género e hipocondría, no tendría por qué resultar incoherente.

En cuanto a los resultados del Experimento 1, se obtuvo efecto de inconsistencia tanto en estudiantes como en clínicos en el sentido de que los tiempos de lectura de las frases target fueron mayores en la condición inconsistente que en la condición control. Sin embargo, sólo en la muestra de clínicos el efecto de inconsistencia estuvo modulado por el estatus causal del síntoma referido en la frase target. Específicamente, el efecto de inconsistencia detectado fue de mayor magnitud cuando la frase target se refería a la ausencia de un síntoma causalmente central que cuando se refería a la ausencia de un síntoma causalmente periférico. En todos los casos, los tiempos de lectura fueron congruentes con los juicios diagnósticos obtenidos: tanto los estudiantes como los clínicos dieron juicios de acuerdo con el diagnóstico más bajos en la condición inconsistente que en la condición control. Sin embargo, sólo en los clínicos estos juicios mostraban un efecto de inconsistencia modulado por el estatus causal de los síntomas, es decir, un efecto de inconsistencia mayor en la condición de síntoma central que en la de síntoma periférico. Por tanto, se replicó el efecto de estatus causal obtenido por Kim y Ahn (2002) en juicios diagnósticos sólo en la muestra de psicólogos clínicos.

Experimento 2

Los objetivos del Experimento 2 fueron 1) evaluar la implicación del razonamiento causal manipulando el orden temporal de síntomas, lo cual es un rasgo definitorio de las relaciones causales, y ofreciendo información explícita sobre conexiones causales entre síntomas, y 2) probar, como en el Experimento 1, si nuestra manipulación podía sesgar el razonamiento diagnóstico de los participantes, evaluando la implicación de procesos de razonamiento causal atribuibles al sistema 1 en tal sesgo.

Esta nueva manipulación nos permitiría solventar algunas limitaciones del Experimento 1 y ofrecer evidencia más fuerte de razonamiento causal. En este caso, la manipulación de información se realizaba alterando el orden temporal en el que se presentaban los síntomas en el paciente, de tal forma que en el informe control se presentaban los síntomas del diagnóstico del trastorno mencionado en la información preliminar en el orden adecuado de acuerdo con la teoría causal del trastorno y, en el informe inconsistente, esos mismos síntomas, tras ese mismo diagnóstico, se presentaban en el orden inverso. Además, añadimos una frase hacia el final del informe, tras la lectura de los síntomas, relacionada con la conexión entre ellos. En los informes inconsistentes, esta frase afirmaba que no existía ningún tipo de relación entre los síntomas, mientras que, en los informes control, esta frase afirmaba que existía una relación de causalidad entre los síntomas en la que el síntoma A causaba el síntoma B, y el síntoma B causaba el síntoma C. Al igual que en el Experimento 1, si los participantes ponían en marcha sus teorías causales de los trastornos, se esperaban tiempos de lectura más largos en los informes inconsistentes que en los informes control al comparar las frases target o frases que contenían los síntomas. De igual forma, esperábamos que la información referente a la secuencia temporal de síntomas y a las conexiones causales entre síntomas influyera en los juicios sobre diagnóstico.

En el Experimento 2 participaron 71 estudiantes de psicología de la Universidad de Málaga y 30 clínicos, éstos últimos con 10 años experiencia media y en su mayoría cognitivo-conductuales. Todos ellos debían leer 12 informes clínicos: 6 inconsistentes y 6 control, en dos sesiones diferentes. Tanto en los informes inconsistentes como control, se ofrecía una información preliminar en la que se presentaba a un paciente que había sido diagnosticado con uno de los siguientes trastornos: anorexia nerviosa, depresión mayor, fobia específica, trastorno obsesivo-compulsivo, trastorno de estrés

postraumático y trastorno de ansiedad generalizada. En este experimento, además, para comprobar que la teoría causal que los sujetos estaban poniendo en marcha para resolver las tareas de lectura y juicio diagnóstico era consistente con la presentada en los informes clínicos, tenían que realizar una tarea adicional en la que debían estimar la eficacia que diferentes tratamientos reales podrían tener sobre los diferentes síntomas presentados, empleando de nuevo escalas. En la medida en que las respuestas en esta prueba de tratamiento fueran consistentes con el uso de la teoría causal empleada en la elaboración del material, podríamos afirmar que la manipulación era apropiada.

Los resultados del Experimento 2 mostraron que los juicios diagnósticos de los estudiantes estuvieron afectados por información sobre la secuencia temporal de síntomas y por la información sobre conexiones causales. Específicamente, los estudiantes estuvieron más de acuerdo con el diagnóstico recibido por los pacientes hipotéticos cuando la secuencia temporal de síntomas y las conexiones causales entre ellos fueron consistentes con las teorías causales de los trastornos diagnosticados que cuando la información ofrecida era inconsistente con tales teorías. Además, los estudiantes emplearon más tiempo en decidir sobre el juicio diagnóstico en la condición inconsistente no-causal, que en la condición consistente causal. Sin embargo, no encontramos evidencia convincente de procesos de razonamiento causal rápidos y online en esta muestra. De igual forma, el tiempo que éstos emplearon en la lectura de frases target relacionadas con la secuencia temporal de los síntomas parecía no estar afectado por si el orden temporal era consistente o inconsistente con la teoría causal del trastorno mencionado en la información preliminar del informe clínico. Los resultados encontrados en clínicos también mostraron un mayor grado de acuerdo con el diagnóstico recibido por el paciente hipotético en la condición consistente causal que en la condición inconsistente no-causal. Los clínicos además emplearon más tiempo en la

tarea de juicio diagnóstico en la última condición que en la primera, sugiriendo que trataron de resolver las inconsistencias causales encontradas en el informe clínico antes de realizar el juicio. Además, encontramos evidencia de razonamiento causal rápido y online en los tiempos de lectura de las frases target de los clínicos. Específicamente, los tiempos de lectura fueron significativamente más largos cuando la información fue inconsistente que cuando fue consistente con las teorías causales del trastorno con el cual el paciente hipotético había sido diagnosticado.

Experimento 3

El objetivo específico del Experimento 3 fue encontrar más evidencia acerca de que la computación de coherencia causal reside en el núcleo de los procesos de comprensión que dependen del sistema 1, tanto durante la lectura como en los juicios diagnósticos. En este caso, tras la información preliminar relacionada con el diagnóstico, se presentaban tres frases ofreciendo información con respecto a la presencia o ausencia de tres diferentes síntomas considerados criterios diagnósticos del trastorno en cuestión de acuerdo al DSM-IV. Tales síntomas formaban parte de una cadena causal (S1→S2→S3) que no era explicitada en el informe clínico, pero que se esperaba que formara parte de la teoría causal del trastorno de los participantes. Las inconsistencias se crearon mencionando explícitamente la ausencia de uno de esos tres síntomas en una de las frases. De esta forma, con esta manipulación, se crearon dos tipos de inconsistencias: *categórica*, entre el diagnóstico y el síntoma ausente, y *causal*, entre la presencia de algunos síntomas y la ausencia de sus antecedentes causales en la cadena causal que conectaba los síntomas. De estos dos tipos de inconsistencia, se podían derivar algunas predicciones de acuerdo a diferentes teorías causales

relacionadas con coherencia y estatus causal. Asumiendo que los participantes activaban automáticamente una teoría causal del trastorno, realizamos las siguientes predicciones:

	Condición (tipo de informe clínico)		
	Paciente 1 ~~S1~~→S2→S3	Paciente 2 S1→~~S2~~→S3	Paciente 3 S1→S2→S3
Frase sobre S1	S1 ausente (cat inconsistente)	S1 presente (consistente)	S1 presente (consistente)
Predicción de TR	Lectura lenta	Lectura rápida	Lectura rápida
Frase sobre S2	S2 presente (caus inconsistente)	S2 absent (cat y caus inconsistente)	S2 presente (consistente)
Predicción de TR	Lectura lenta	Lectura lenta	Lectura rápida
Frase sobre S3	S3 presente (consistente)	S3 presente (caus consistente)	S3 ausente (cat y caus inconsistente)
Predicción de TR	Lectura rápida	Lectura lenta	Lectura lenta
Predicción de juicio diagnóstico	Intermedio	El más bajo	El más alto

Nota. cat y caus corresponden a categóricamente y a causalmente, respectivamente.

En el Experimento 3 participaron 31 estudiantes de Psicología de la Universidad de Göttingen (Alemania) que habían sido entrenados previamente en los trastornos empleados. Los trastornos fueron: depresión, trastorno obsesivo- compulsivo y fobia específica. Como de cada trastorno se crearon 3 tipos de informe (~~S1~~→S2→S3; S1→~~S2~~→S3; S1→S2→S3), cada participante leía un total de 9 informes clínicos en dos sesiones diferentes. Al igual que en el Experimento 2, para comprobar qué tipo de teoría causal estaban poniendo en marcha los sujetos, éstos tenían que realizar una tarea adicional en la que debían estimar en qué medida cada uno de los síntomas empleados causaba cada uno de los restantes, de nuevo empleando escalas para responder.

En cuanto a los resultados del Experimento 3, la evidencia más convincente se obtuvo en los tiempos de lectura. Los tiempos fueron más lentos cuando los participantes leían la frase que informaba de la presencia de un síntoma cuando su antecedente causal estaba ausente. Este resultado indicaba que los participantes detectaban una inconsistencia causal ya que, de acuerdo con la teoría causal activada, dada la ausencia de un antecedente causal, el siguiente síntoma de la cadena causal no debería ocurrir. Los resultados en juicios diagnósticos se ajustaron a las predicciones de la *teoría del modelo causal* (Rehder, 2001; Rehder y Hastie, 2004; Rehder y Kim, 2010), de acuerdo a la cual, los juicios diagnósticos (categóricos) son el resultado de la computación de coherencia. Finalmente, el hecho de que los efectos de inconsistencia encontrados en tiempos de lectura fueran consistentes con los juicios diagnósticos de los participantes sugería que la computación de coherencia depende de procesos de razonamiento causal del sistema 1.

Experimento 4

Los objetivos del Experimento 4 fueron: 1) comprobar si el razonamiento diagnóstico es sensible a la manipulación de los mecanismos causales que conectan unos síntomas con otros, y 2) evaluar si tal sensibilidad a mecanismos causales podía ser atribuida a razonamiento causal del sistema 1 durante la lectura de informes clínicos. Este experimento se centra en procesos rápidos de razonamiento online basados en los supuestos de los participantes con respecto a mecanismos causales. Evaluamos si informando a los participantes explícitamente sobre los mecanismos que conectan los síntomas de un trastorno afectaba al razonamiento causal online y a posteriores juicios diagnósticos. En este experimento, cada informe ofrecía información acerca de un paciente que presentaba 3 síntomas que eran criterios diagnósticos de un trastorno. De

acuerdo al conocimiento general clínico, los 3 síntomas formaban parte de una cadena causal (S1→S2→S3). La única diferencia residía en los mecanismos causales que llevaban de un síntoma al siguiente. En una condición, los mecanismos causales fueron consistentes con la teoría causal del trastorno mencionado en el diagnóstico al comienzo del informe clínico. En la otra condición, se presentaban mecanismos causales plausibles alternativos, los cuales eran inconsistentes con la teoría causal del trastorno. Esperábamos que los participantes fueran sensibles a los mecanismos causales y, por tanto, que detectaran inconsistencias en los mecanismos. Según esto, esperábamos tiempos de lectura más largos en los casos en los cuales los mecanismos no coincidían con la teoría causal del trastorno. Además, esperábamos que la información sobre los mecanismos afectara a los juicios diagnósticos finales a pesar de que el criterio diagnóstico resultara inalterado.

En el Experimento 4 participaron 34 estudiantes de psicología de la Universidad de Málaga, la mayoría de las cuales había recibido información previa sobre los trastornos empleados. Dos factores fueron ortogonalmente manipulados: el mecanismo causal que llevaba de un síntoma al siguiente de la cadena causal y el diagnóstico recibido por el paciente, el cual de nuevo aparecía en el comienzo del informe clínico. Los trastornos que se emplearon fueron: anorexia nerviosa, fobia específica y ansiedad generalizada. Ya que de cada trastorno había cuatro tipos de informe (diagnóstico psicológico y mecanismo psicológico, diagnóstico psicológico y mecanismo bio-médico, diagnóstico médico y mecanismo psicológico, diagnóstico médico y mecanismo bio-médico), cada participante leía un total de 12 informes clínicos en dos sesiones diferentes. En este experimento, tras la tarea de diagnóstico, los participantes tenían que enfrentarse además con una tarea de tratamiento en la cual se les pedía que decidieran, para cada caso, si consideraban que sería más eficaz un tratamiento

psicológico o uno médico. Para cada respuesta, debían utilizar de nuevo una escala en la que en cada uno de los extremos se localizaba cada una de las formas de tratamiento.

En el Experimento 4, los resultados encontrados en tiempos de lectura revelaron la participación de procesos de razonamiento causal basados en creencias acerca de los mecanismos causales que subyacen a los diferentes trastornos. Esta conclusión se apoya en la obtención de tiempos de lectura mayores en las frases informando sobre mecanismos bio-médicos comparados con las frases informando de mecanismos psicológicos, cuando el paciente hipotético había sido diagnosticado con un trastorno mental. Esta conclusión no puede ser explicada por la activación de creencias causales independientes a las creencias de los participantes acerca de trastornos mentales, ya que la diferencia entre tiempos de lectura se revertía cuando el paciente hipotético había sido diagnosticado con una enfermedad médica. El patrón de resultados encontrados en juicio diagnóstico fue el mismo que en tiempos de lectura, sugiriendo de nuevo que los procesos de razonamiento causal rápido y online que tenían lugar durante la lectura podían haber tenido un impacto determinante en los juicios diagnósticos. Por último, los resultados de la tarea de toma de decisión sobre tratamiento corroboraron que las descripciones ofrecidas por los mecanismos psicológicos y bio-médicos fueron interpretadas como tales por los participantes. Específicamente, el tratamiento psicológico recibía mayores puntuaciones en la condición de mecanismo psicológico que en la condición de mecanismo bio-médico, mientras que se obtenía el patrón inverso de resultados en las puntuaciones dadas al tratamiento médico.

Discusión

El patrón de resultados obtenidos en esta serie experimental es muy consistente, al mostrar que tanto estudiantes de Psicología como clínicos con experiencia emplean

procesos de razonamiento rápidos y online, lo que pueden ser identificados como el sistema 1 de Kahneman y Frederick (2002), durante la lectura fluida de informes clínicos para la toma de decisiones en juicios diagnósticos posteriores. Tales procesos de razonamiento implicaban recuerdo rápido de teorías causales sobre trastornos mentales almacenadas en la memoria, así como procesos de inferencia rápidos, los cuales demostraron ser cruciales para computar la coherencia de la información ofrecida. Es probable que tales procesos dependan de procesos guiados por la coherencia y basados en procesos de activación característicos de redes neuronales dinámicas. Esta serie experimental ofrece evidencia convergente demostrando que los procesos de razonamiento del sistema 1 son responsables de la computación de la coherencia que depende de teorías causales de trastornos mentales, las cuales establecen la manera en la cual ciertos síntomas derivan en otros. La dependencia de los razonadores de las teorías causales por medio de procesos de razonamiento rápidos y online se ha demostrado a través de varias manipulaciones que, tomadas en conjunto, han producido un patrón de resultados muy consistente. Además, hemos encontrado una consistencia persistente entre los procesos de razonamiento rápidos detectados por medio de medidas online y los juicios diagnósticos de los participantes tras la tarea de lectura, excepto en los resultados de la muestra de estudiantes en el Experimento 2. Esta consistencia sugiere que los procesos del sistema 1, rápidos y tempranos, están basados en razonamiento causal que puede sesgar los procesos del sistema 2 que podrían funcionar durante la tarea de juicio diagnóstico.

Conclusión

El patrón de resultados encontrados en la presente serie experimental demuestra que tanto los clínicos como los estudiantes de Psicología se involucran en procesos de

razonamiento on-line muy rápidos, que pueden asimilarse a los procesos del sistema 1, durante la lectura de informes clínicos para un juicio diagnóstico posterior. Tales procesos de razonamiento están implicados en la rápida activación de atributos causales y realización de inferencias en el mismo momento en que se recibe información relevante para el diagnóstico de trastornos mentales. Estos procesos serían fundamentales para calcular la coherencia de la información ofrecida, contribuyendo, de este modo, en la construcción de un modelo mental coherente para una adecuada comprensión del caso.

Palabras clave: conocimiento causal, razonamiento clínico, diagnóstico de trastornos mentales, paradigma de inconsistencia, procesos del sistema 1

tapraid5/z1t-assess/z1t-assess/z1t00114/z1t2648d14z | xppws | S=1 | 11/1/13 | 11:02 | Art: 2013-1734

Psychological Assessment

1040-3590/13/$12.00 DOI: 10.1037/a0035151

BRIEF REPORT

Detecting Fast, Online Reasoning Processes in Clinical Decision Making

AQ: au

Amanda Flores, Pedro L. Cobos, and Francisco J. López
University of Málaga, and Instituto de Investigación Biomédica de Málaga (IBIMA), Málaga, Spain

Antonio Godoy
University of Málaga

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 *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., text rev.; American Psychiatric Association, 2000). 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 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 clinicians' causal reasoning. The implications of these results for clinician training are also discussed.

Keywords: diagnostic criteria, clinical reasoning, inconsistency paradigm, causal reasoning

The *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., text rev.; *DSM–IV–TR;* American Psychiatric Association, 2000) generally assumes that all the diagnostic criteria for a mental disorder are equivalent (Kim & Ahn, 2002). Despite this, some AQ: 1 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 that person presented a different set. This result is consistent with other findings related to previous versions of the *DSM* (see Davis, Blashfield, & McElroy, 1993; Garb, 1996; Rubinson, Asnis, & 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–TR*.

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 auto-

Amanda Flores, Pedro L. Cobos, and Francisco J. López, Departamento de Psicología Básica, University of Málaga, Málaga, Spain, and Instituto de Investigación Biomédica de Málaga (IBIMA), Málaga, Spain; Antonio Godoy, Departamento de Personalidad, Evaluación y Tratamiento Psicológico, University of Málaga, Málaga, Spain.

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

tapraid5/z1t-assess/z1t-assess/z1t00114/z1t2648d14z | xppws | S=1 | 11/1/13 | 11:02 | Art: 2013-1734

matic 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 (e.g., 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, judgment-based, slow, reflective, and resource-demanding processes can take place. However, this study focuses on processes that take place in a few 10ths of a second and that demand 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–TR* 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 the person 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 the person presents some symptoms that form part of the *DSM–IV–TR* 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 seems more relevant than the second, 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 that, 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 the *DSM–IV–TR*. 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

tapraid5/z1t-assess/z1t-assess/z1t00114/z1t2648d14z | xppws | S=1 | 11/1/13 | 11:02 | Art: 2013-1734

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) from Málaga University in Spain and experienced clinicians (n = 17) who worked in independent practice in the Málaga area. Their experience as clinicians ranged from 3 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–TR* disorders, as in Kim and Ahn's (2002) 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, 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'Brien, 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–TR* 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 frequent symptoms (i.e., present in numerous *DSM–IV–TR* disorders) were also included. Immediately before the target sentence, there were four filler sentences related to nonclinical information. The filler information would make the previous information on clinical symptoms unavailable from the participants' working memory by the time the target information was read. The last two sentences in the text were the target and the posttarget 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;* American Psychiatric Association, 2000). 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;* American Psychiatric Association, 2000). The posttarget 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, and returning to the 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, participants completed the diagnostic judgment task, in which they had to rate on a continuous scale from 0 to 100 (i.e., from complete disagreement to complete agreement) the extent to which they agreed with the diagnosis provided in the text. Once participants 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 posttarget 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 min to complete.

tapraid5/z1t-assess/z1t-assess/z1t00114/z1t2648d14z | xppws | S=1 | 11/1/13 | 11:02 | Art: 2013-1734

Results

Reading Times

The RTs for the target and posttarget 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 posttarget sentences. In total, 12 target sentence RTs and 10 posttarget sentence RTs were eliminated.

T1 Table 1 shows the mean RTs for the target and posttarget 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 2 (inconsistency: inconsistent vs. control) × 2 (relevance of the symptoms: high vs. low) repeated-measures analysis of variance (ANOVA) on the students' RTs yielded a significant main effect of inconsistency, $F(1, 16) = 24.091$, $MSE = 271{,}235.080$, $p < .001$; $\eta^2 = .56$. None of the other effects were significant (all F values < 2.92). The same trend, although much smaller, was observed for the posttarget sentence. However, an identical 2 × 2 ANOVA on the RTs for the posttarget sentence yielded no significant effect (all F values < 1.27). Regarding the sample of clinicians, the same 2 × 2 ANOVA on RTs for the target sentences yielded a significant effect of inconsistency, $F(1, 16) = 12.801$, $MSE = 1{,}036{,}503.4$, $p = .003$, $\eta^2 = .44$; relevance of symptoms, $F(1, 16) = 9.043$, $MSE = 186{,}673.901$, $p = .008$, $\eta^2 = .36$; and Inconsistency × Relevance of Symptoms, $F(1, 16) = 6.505$, $MSE = 286{,}012.319$, $p = .021$, $\eta^2 = .289$. Simple effects analyses revealed a significant inconsistency effect in both conditions of the relevance of symptoms factor, $F(1, 16) = 14.203$, $MSE = 882{,}461.693$, $p = .002$, $\eta^2 = .47$; $F(1, 16) = 5.899$, $MSE = 440{,}053.978$, $p = .027$, $\eta^2 = .27$, for the high-relevance and low-relevance conditions, respectively.

Table 1 also reveals similar results for the RTs of posttarget 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 2 × 2 ANOVA, which yielded a significant main effect of inconsistency, $F(1, 16) = 5.565$, $MSE = 297{,}226.601$, $p = .031$, $\eta^2 = .26$, and a marginally significant effect of the Inconsistency × Relevance of Symptoms interaction, $F(1, 16) = 3.691$, $MSE = 125{,}731.452$, $p = .073$, $\eta^2 = .19$. 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, $F(1, 16) = 5.873$, $MSE = 329{,}524.764$, $p = .028$, $\eta^2 = .27$, 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–TR* 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 × 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–TR* 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 the *DSM–IV–TR*. 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 2 (inconsistency: inconsistent vs. control) × 2 (relevance of symptoms: high vs. low) repeated-measures ANOVA 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

Table 1
Mean Reading Times (in ms) and Standard Deviations for Target and Posttarget 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	
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Students								
Target sentences	2,992.83	1,025.74	2,205.50	630.87	2,708.16	1,007.04	2,233.68	592.93
Posttarget sentences	2,049.21	642.46	1,942.93	768.85	2,104.15	623.44	1,980.50	952.73
Judgments	54.31	10.96	77.60	10.37	53.63	16.81	72.50	11.89
Experienced clinicians								
Target sentences	3,472.15	1,385.86	2,257.85	562.92	2,826.20	1,084.93	2,273.56	706.01
Posttarget sentences	2,612.94	840.63	1,948.70	704.45	2,420.38	662.66	2,329.13	824.15
Judgments	36.91	19.05	69.56	23.14679	47.01	22.05	67.25	24.67

tapraid5/z1t-assess/z1t-assess/z1t00114/z1t2648d14z | xppws | S=1 | 11/1/13 | 11:02 | Art: 2013-1734

symptoms nor the interaction between the two factors was statistically significant (all F values < 3.16). Regarding the sample of clinicians, the same 2 × 2 ANOVA yielded a significant effect of inconsistency, $F(1, 16) = 46.896$, $MSE = 247.928$, $p < .001$, $\eta^2 = .75$, and Inconsistency × 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.

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 2 (inconsistency: inconsistent vs. control) × 2 (relevance of symptoms: high vs. low) × 2 (type of sample: psychology students vs. experienced clinicians) ANOVA revealed that the Target Inconsistency × Relevance of Symptoms × Type of Sample second-order interaction was not significant regarding the RTs for the target sentence, $F(1, 32) = 0.703$; only marginally significant for the posttarget sentence, $F(1, 32) = 4.02$, $MSE = 125{,}526.813$, $p = .053$, $\eta^2 = .11$; 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–TR* 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 et al., 1990): (a) the status of the *DSM–IV–TR* 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 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–TR*) 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 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* (4th ed., text rev.). Washington, DC: Author.

tapraid5/z1t-assess/z1t-assess/z1t00114/z1t2648d14z | xppws | S=1 | 11/1/13 | 11:02 | Art: 2013-1734

Black, J., & Bower, G. H. (1980). Story understanding as problem-solving. *Poetics, 9,* 223–250. doi:10.1016/0304-422X(80)90021-2

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

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,* 182–190. doi:10.1097/00001888-200002000-00020

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,* 319–322. doi:10.1037/0021-843X.102.2.319

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, NY: Farrar, Strauss, and Giroux.

Kendell, R. E. (1973). Psychiatric diagnoses: A study of how they are made. *British Journal of Psychiatry, 122,* 437–445. doi:10.1192/bjp.122.4.437

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

AQ: 2

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., Asnis, G. M., & Friedman, M. H. (1988). Knowledge of diagnostic criteria for major depression: A survey of mental health professionals. *Journal of Nervous and Mental Disease, 176,* 480–484. doi:10.1097/00005053-198808000-00005

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, the Netherlands: North-Holland.

Schmidt, H. G., Norman, G. R., & Boshuizen, H. P. A. (1990). Cognitive perspective on medical expertise: Theory and implication. *Academic Medicine, 65,* 611–621. 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). Causal relatedness and importance of story events. *Journal of Memory and Language, 24,* 595–611. doi:10.1016/0749-596X(85)90048-8

Trabasso, T., & van den Broek, R. (1985). Causal thinking and the representation of narrative events. *Journal of Memory and Language, 24,* 612–630. doi:10.1016/0749-596X(85)90049-X

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

Westen, D. (2012). Prototype diagnosis of psychiatric syndromes. *World Psychiatry, 11,* 16–21. doi:10.1016/j.wpsyc.2012.01.004

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

Received February 7, 2013
Revision received October 15, 2013
Accepted October 16, 2013 ■