Clinicians as Mechanics?

Causal Reasoning in Clinical Judgment and Decision Making

Colofon

ISBN/EAN: 978-90-9024464-8

Cover image: Photo by William Moore; Box by Jeroen Coumans
Print: Printpartners Ipskamp, Nijmegen

Clinicians as Mechanics?

Causal Reasoning in Clinical Judgment and Decision Making

Een wetenschappelijke proeve op het gebied van de Sociale Wetenschappen

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus, prof. mr. S. C. J. J. Kortmann
volgens het besluit van het College van Decanen
in het openbaar te verdedigen op woensdag 2 september 2009 om 13.00 uur

door

Leontien de Kwaadsteniet
Geboren op 11 mei 1974 te Doetinchem

Promotor:
Prof. dr. Cilia L. M. Witteman

Copromotores:
Dr. Nicole P. C. M. Krol
Dr. John H. L. van den Bercken

Manuscriptcommissie:
Prof. dr. Giel J. M. Hutschemaekers (Voorzitter)
Prof. dr. Laurence Claes (Katholieke Universiteit van Leuven)
Prof. dr. Jules M. Pieters (Universiteit Twente)

Contents

Chapter 1

General Introduction

In this thesis studies are presented that aim to find evidence that mental health clinicians reason causally when thinking about what is the matter with a client, and deciding on effective interventions. This introduction chapter first describes a metaphor that Paul Meehl (1920-2003) came up with in 1954. Meehl compares clinicians to box mechanics deriving their hypotheses from assumed internal mechanisms. Thus, Meehl depicts clinicians as reasoning causally. From this metaphor we move to the "clinical-statistical controversy", and explain why clinicians' reasoning processes should be investigated. We review literature on the importance of causal reasoning in different domains, and on causal reasoning specifically in the clinical domain. We end with an introduction to the studies in the next chapters of this thesis.

Clinicians as mechanics?

How do clinicians arrive at hypotheses about what is wrong with their clients? In his treatment of this question Meehl (1954) compares a clinician with a box mechanic making predictions about an opaque box. This box has ten buttons on one side and ten coloured lights on the other side. Pressing a combination of three buttons leads to lights flashing on and off. Inside the box is a complex mechanism of "interacting gears, brackets, pulleys, springs, sliding surfaces, and the like" (Meehl, 1954, p.57). Meehl distinguishes two ways to make predictions about what lights will flash on and off when a new combination of three buttons is pressed. One way to predict the box's response is by deriving the most probable result from all previously observed combinations of button presses and light flashes, of this and perhaps also of other similar boxes. This he calls the actuarial or statistical way of prediction. Another way is that of an experienced box mechanic. An experienced box mechanic has not only observed combinations of button presses and light flashes, but also knows what the box may look like on the inside, because he has investigated the insides of many similar boxes before. Thus, using this knowledge a box mechanic can make highly specific predictions.

Meehl, who was a practicing clinician himself, proposes that a skilled clinician generates hypotheses in a way similar to how the skilled mechanic would come to a prediction for an opaque box. Based on knowledge of psychological laws, experience with previous clients, and unique client characteristics, a clinician can arrive at highly specific hypotheses. A statistician could not infer such specific hypotheses, because often no prior observations are available to make the necessary deductions from unique characteristics. Meehl thus argues against Sarbin (1944) and Lundberg (1941) who seem to see clinicians as statisticians. The latter suggest that clinicians always derive hypotheses from comparisons with previous cases, with which a new client will always have something in common. These hypotheses may be made in an informal or unconscious way, but are still essentially statistical. With Sarbin and Lundberg however, Meehl argues that, no matter in what way hypotheses are generated, clinicians should always validate their hypotheses using statistical methods if these are available.

In 1954 Meehl discussed 20 studies showing that predictions of human behaviour made using actuarial methods (such as calculating probabilities with regression analysis, or using frequency tables) are equally often or more often correct than predictions which

clinicians made using their own, unspecified clinical methods. Knowing this, Meehl argues, clinicians should always use actuarial methods to test hypotheses about their clients when these are available. Evidence for the superiority of actuarial methods above so-called clinical methods to make predictions about human behaviour has further accumulated. Recent meta-analyses by Grove, Zald, Lebow, Snitz, and Nelson (2000) and Ægisdóttir et al. (2006) show that actuarial methods outperform informal methods used by clinicians in a range of predictions.

Despite this large body of evidence favouring the use of actuarial methods, the "clinical-statistical controversy" (Meehl & Grove, 2006) is still actual (see e.g. Dawes, 2005; Grove, 2005). In practice clinicians do not seem to be convinced of the superiority of statistical methods (Ægisdóttir et al., 2006; Caspar, Berger, & Hautle, 2004; Meehl & Grove, 2006). An explanation for this discrepancy might be that actuarial methods do not easily combine with how clinicians reason about their clients (cf. Caspar et al., 2004). The seeming discrepancy in Meehl's (1954) discussion that clinicians arrive at the most specific hypotheses like mechanics rather than statisticians while at the same time they should use statistical methods to make the most accurate clinical predictions, might reflect this difficulty. Further, for much of what clinicians do in everyday practice statistical methods may not be available (Westen & Weinberger, 2005). Knowledge of how clinicians process client information and arrive at their decisions might help to understand this discrepancy and think of ways to improve clinicians' judgments, by the development of decision aids and improved training methods (Caspar, 1997; Caspar, Berger, & Hautle, 2004; Galanter & Patel, 2005; Garb, 1998, 2005; Nurcombe & Gallagher, 1986; Witteman, Harries, Bekker, & Van Aarle, 2007).

Studies about how clinicians process client information and arrive at their decisions are scarce in comparison to studies which compare clinicians' predictions with actuarial predictions (Caspar, 1997; Groth-Marnat, 2003; Nurcombe 2000). This may be due to the difficulty of studying thought processes, and especially thought processes of practicing professionals (Caspar et al., 2004; Chi, 2006; Ericsson, 2006). This thesis presents studies that investigate mental health clinicians' reasoning, and their causal reasoning in particular.

Causal reasoning

In portraying clinicians as mechanics who derive predictions from assumed mechanisms, Meehl (1954) described clinicians' thought processes as causal reasoning processes. Causal reasoning is central to human thought. Representation of causal relations enables us to predict what will happen, and more importantly: it enables us to predict what the effects of our actions will be, and thus to control our environment (Beach, 1992; Gopnik & Schultz, 2007; Pearl, 2000; Sloman, 2005; Woodward, 2003).

People distinguish events that simply co-occur from events that are causally related (Keil, 2006; Sloman, 2005; Sloman & Hagmayer, 2006; Woodward, 2007). For example, we know that although a seriously running nose and a fever often co-occur, with fever starting somewhat later than the running nose, the running nose does not cause the fever. Rather, there is a common cause, a virus, for both symptoms of influenza. We do not squeeze our noses, or fill them with handkerchiefs, to prevent fever. We combine information about the

co-occurrence of events with information about the temporal sequence of events, their spatial contiguity, and knowledge to distinguish causes and effects from merely co-occurring events (Hagmayer & Waldmann, 2002; Waldmann, 1996). People have reliable intuitions as to what is causally relevant and what is not. For example, we consider colour more likely causally relevant for a living kind (e.g. an insect) than for an artefact (e.g. a tool) (Keil, 2003). Even young children differentiate between causal relations and mere correlations, have intuitions as to what are possible causes in a domain, and they use causal representations to predict and explain events (Keil, 2003; Schulz, Kushnir, & Gopnik, 2007; Wellman & Liu, 2007).

Causal reasoning is used in the explanation of physical events, and it can be applied to all domains in which we intervene, or in which we can imagine interventions (Woodward, 2003). We also understand people's mental states and actions causally, by representing preceding events, feelings, and goals (Wellman & Liu, 2007). Causal representations can make it easier to comprehend and use probabilistic information, such as correlations and base rates (Dawes, 1999; Garcia-Retamero & Hoffrage, 2006; Pearl, 2000; Sloman, 2005; Strevens, 2007). It is unsettled whether explanations are secondary in the sense that we need them because of the primary goal of predicting future events, or whether we have a fundamental tendency to explain, which we, subsequently, can use to predict (Wellman & Liu, 2007). In both cases however, reasoning using causal representations, or causal models, appears an essential, fundamental human capacity.

People use causal representations in a wide range of reasoning and decision tasks, such as diagnostic reasoning (reasoning about possible causes for problems), choosing interventions, making inferences about unknown properties and events, and classifying (assigning items to categories) (Sloman, 2005). It has even been proposed that in general we process information by constructing mental causal representations (Graesser, Olde, & Kletke, 2002). We will discuss studies on causal reasoning in several different domains, and review what is known about causal reasoning in the clinical domain.

Causal representations

Pennington and Hastie's theory of explanation-based decision making (Hastie & Pennington, 2000; Pennington & Hastie, 1993) states that in complex, dynamic situations people construct causal representations which explain available information. To complete these causal models they make inferences, using their background knowledge. People arrive at their decisions by choosing the decision alternative that has the best fit with their causal model. Confidence in a decision depends on how coherent the constructed causal model is, that is: how complete, consistent and plausible, how much of the information is covered by the model, the uniqueness of the model, and the goodness of fit of the model with the chosen decision alternative. Related to this theory is Sloman and Hagmayer's (2006) causal model framework for choice. According to this framework, when people make decisions they also first construct a causal model of the available information. They then proceed by mentally simulating the consequences of alternative choices in this model, to see what option is most favourable.

In the legal domain, deciding on a verdict is a complex task. Trial information comes typically in large amounts, in scrambled sequence, and is usually incomplete as to what

exactly happened. Further, different pieces of information often cannot be interpreted in isolation, but are interdependent (Hastie & Pennington, 2000; Pennington & Hastie, 1993). In a think aloud study Pennington and Hastie (1986) found that American jurors construct causal mental representations of trial information that have a story-structure. Jurors who arrived at the same verdict had constructed similar stories, which were different from the stories constructed by jurors who had arrived at other verdicts (see also Huntley & Costanzo, 2003).

Pennington and Hastie (1988) further showed that jurors construct causal representations while processing trial information, and that these representations precede their verdicts. In a memory study jurors more frequently falsely recognised items as information presented at trial when these fitted with their verdict than when they did not. Further, story construction was facilitated by presenting trial information in a chronological and causally consistent order, for either prosecution evidence or for defence evidence. The number of guilty or innocent verdicts appeared to vary with this manipulation.

People also construct causal representations while processing information when no specific decisions need to be made. Researchers in text and discourse comprehension have proposed that processing information of texts, of discourse, but also for example of movies, is essentially explanation-based (Graesser, Olde, & Kletke, 2002). People, when they read, combine information which is stated in the texts with additional information from their background knowledge, in a mental situation representation. To fill in gaps in the representation readers make inferences. A distinction is made between inferences that are needed for local coherence, allowing comprehension of subsequent sentences, and for global coherence, allowing sentences in the whole text to be integrated meaningfully. It is proposed that readers generally arrive at global coherence by constructing causal mental representations (Magliano, 1999; Langston & Trabasso, 1999; Van den Broek & Gustafson, 1999). Think aloud studies in which the inferences that readers make while processing texts have been analysed, have shown that indeed most inferences are causal (Trabasso & Magliano, 1996; Zwaan & Brown, 1996). Other methods, such as measuring sentence reading times, lexical decision latencies, or word recognition latencies also showed that readers make causal inferences while reading (Magliano, 1999; Graesser, Singer, & Trabasso, 1994).

Clinical causal client representations

Clinical decision making is an instance of decision making in complex, dynamic situations. The problems that clients present with are caused and maintained by different interacting psychological, biological, and environmental factors (Carr, 2006; De Los Reyes & Kazdin, 2006; Haynes, 1992; Kendall, Holmbeck, & Verduin, 2004; Wenar & Kerig, 2006; Kiesler, 1999). Problems usually extend over a longer period of time, and can take different forms in subsequent phases in life (Morton, 2004). Information that clinicians collect about their clients, from interviews, observations, and tests, will never be complete with respect to all possibly relevant causes in the client's life that have resulted in the client's current problems (Meehl, 1954). Given the complexity and the nature of the clinicians' task it seems likely that clinicians will also construct causal mental representations of client information, in order to understand their clients' problems and to decide about the most effective

interventions to diminish these problems (cf. Goldstein & Weber, 1995; Rettinger & Hastie, 2003).

From a prescriptive point of view, this is also how clinicians should proceed. It is recommended that clinicians construct a case formulation for their clients which "aims to describe a person's presenting problems and use theory to make explanatory inferences about causes and maintaining factors that can inform interventions" (Kuyken, Fothergill, Musa, & Chadwick, 2005, p.1188), see for example the handbooks of Carr (2006), De Bruyn, Ruijssenaars, Pameijer, & Van Aarle (2003), Eells (2007), Groth-Marnat (2003), and Wenar & Kerig (2006). It is not clear however whether clinicians do construct client representations that explain the mechanisms underlying clients' problems. Studies on diagnostic judgment and decision making processes are scarce, especially when compared to studies on diagnostic decision making outcomes (Caspar, 1997; Groth-Marnat, 2003; Nurcombe 2000).

Think aloud studies have shown that the clinical judgment and decision making process is a process of generating and testing hypotheses (Nurcombe & Fitzhenry-Coor, 1982) similar to the medical decision making process (Elstein, Shulman & Sprafka, 1978). Psychiatrists generate most hypotheses early, already in the first minutes of a diagnostic interview (Kendell, 1973; Sandifer, Hordern, & Green, 1970). The nature of the hypotheses in these studies has not been specified, and they appear at least partly to be classifying hypotheses.

Investigating the contents of case formulations may inform us about clinicians' causal client representations. Eells, Kendjelic, and Lucas (1998) developed a coding schema to analyse the content of case formulations of clinicians with different theoretical backgrounds: the Case Formulation Content Coding Method (CFCCM). Analysis of intake formulations showed that these mainly summarised descriptive information, and that most formulations lacked a hypothesised mechanism behind clients' problems. Most of the mechanisms that were specified were only rudimentary. Kendjelic and Eells (2007) replicated these findings, but also found that mechanisms became much more specified when clinicians had followed a two-hours training in case formulation, in which the importance of a hypothesised mechanism was emphasized.

Kuyken and colleagues (2005) analysed case formulations which different cognitive therapists gave for the same case description, after training. They found that less than half of the formulations were of sufficient quality, with quality defined in terms of providing a mechanism on which interventions can be based. Kuyken and colleagues further found that reliability of descriptive aspects of the formulations was good, but of explanatory aspects it was poor. Persons and Bertagnolli (1999) similarly found that the mechanisms that therapists provided in cognitive-behavioural formulations had poor reliability. In the study of Kuyken et al. (2005) quality and reliability of the case formulations was higher for therapists who were more experienced and who had more professional qualifications.

In short, clinicians do not generally provide elaborate mechanisms for their clients' problems in their case formulations, and furthermore they often seem to disagree about the mechanisms they provide (see also Garb, 1998). It is possible however, that clinicians do endorse causal beliefs about their clients which they do not explicate in their written formulations. It seems in fact rather likely that clinicians will endorse causal beliefs about

their clients, given that they have to decide about effective interventions, and that people have been found to base their intervention choices on causal representations.

Intervention choices

An example from research by Hagmayer and Sloman (2005) shows how knowledge of underlying mechanisms affects intervention choices. Hagmayer and Sloman told students to suppose that research had indicated that men who do the dishes have a better health. They asked the students whether they would advise a friend who wanted to improve his health to do the dishes. Whether they did appeared to depend on how the improved health was explained. If the students were told to suppose that better health was a consequence of the fact that men who care for their health also care for equality issues, they were less likely to advise their friend to do the dishes, than when they were told to suppose that better health resulted from the extra physical activity that doing the dishes gave. In the last case, doing the dishes directly affects health. In the first case, doing the dishes does not affect one's health but improved health is a consequence of a common cause that also causes doing the dishes. Thus, differences in the structure of causal models affected which advice students gave.

Causal representations have also been found to influence which interventions people think are effective in more realistic situations. Students' ratings of the effectiveness of actions to prevent coronary heart disease appeared to be strongly related to their causal models of the risk factors. When for example smoking was considered to strongly increase the risk of coronary heart disease, to stop smoking was rated as a highly effective preventive action (Green & McManus, 1995). Similarly, people's action tendencies in scenarios of environmental risks, such as chemical pollution or a new epidemic, depend on their causal mental representations of these risks (Böhm & Pfister, 2000).

Pliske and Klein (2003) modelled experts' decision processes in several domains, such as warfare technique, fire fighting and nursing. They concluded that in complex, uncertain situations experts decide what to do by constructing causal representations of the situation, and envisioning the effects of an action by mental simulation (see also Klein, 1997; cf. Endsley, 1997). Their model for experts' decision making in complex and unfamiliar situations is similar to Sloman and Hagmayer's causal model framework for choice, except that Pliske and Klein believe that professional decision makers will not usually compare outcomes of different decision alternatives, but will choose the first alternative that leads to satisfactory consequences.

Clinical intervention choices

Many studies have addressed the question how clinician variables (e.g. experience, personality) influence the outcomes of interventions (Beutler, Malik, Alimohamed, Harwood, Talebi, Noble, & Wong, 2004), but it has hardly been investigated how clinicians arrive at their intervention decisions. Nelson and Steele (2008) asked for clinicians' considerations when choosing interventions. They found that evidence for the effectiveness of interventions and flexibility of interventions were the most important considerations. Witteman and Kunst (1997) and Witteman and Koele (1999) investigated to what extent clinicians' interpretations of client data predicted their intervention choices. They found no relation: clinicians who gave the same interpretation of client data chose different interventions, and clinicians who

chose the same interventions gave different interpretations of client data. Theoretical preferences turned out to be the best, though weak predictor of clinicians' intervention choices. Witteman and Kunst (1997) and Witteman and Koele (1999) looked at group similarities in interpretation and intervention combinations. It therefore remains possible that on an individual level interpretations and intervention choices were matched.

Furnham (1995) found that lay people's theories about phobia and their beliefs about the effectiveness of treatments for phobia were coherently related. For example, people who believed in Freudian accounts of the aetiology of phobia were likely to also believe that psychoanalytic treatment will be effective. Yopchick and Kim (2008) presented psychology undergraduates with different causal structures for how symptoms were related in artificial mental disorders. They found that students' ratings of the effectiveness of interventions depended on the place in the causal structures of the symptoms that were affected by the interventions. Interventions that affected symptoms that were initial causes were rated as more effective than interventions that affected symptoms that were intermediate causes. Thus, it seems that at least non-clinicians' ratings of the effectiveness of clinical interventions are related to their causal beliefs.

Classifications

Causal representations also affect classifications. People have theories about how features of concepts of natural kinds (e.g. dogs, tomatoes) and artefacts (e.g. hammers, chairs) are causally related (Medin & Rips, 2005; Murphy & Medin, 1985). Features that are causally central in a concept's theory are perceived as more important, and receive greater weight in the classification of something as belonging to a category, than features which are causally peripheral. This means that people rate features on which more other features depend, or: features that cause more other features, as more important in classification than features which have fewer other features depending on them. This is called the "causal status effect" (Ahn, 1998; Ahn, Gelman, Amsterlaw, Hohenstein, & Kalish, 2000; Ahn, Kim, Lassaline, & Dennis, 2000). For example, having dog-DNA is rated as very important to categorize an animal as a dog, since dog-DNA leads to dogs having many of their characteristic features. Having four legs is rated less important, since having four legs does not lead to many other dog features.

Rehder and Kim (2006) further found that objects with features that have multiple causes are sooner classified as belonging to a category than objects with features that have one cause. Also, when a coherent causal pattern of features is present, consistent with the causal knowledge one has, an object will be classified as belonging to a category sooner. Finally Rehder and Kim found a "primary cause effect", which means that causes that initiate other features in a mechanism are perceived as stronger evidence for category membership than intermediate causes, or (final) effects (for a critical view relating the primary cause effect to the causal status effect, see Rehder & Kim, 2006).

Clinical classifications

Clinicians often classify clients' symptoms and problems as representative of a disorder specified in the Diagnostic and Statistical Manual of Mental Disorders (APA, 2000). To do so, it is required that specified criteria are met. In a theory-drawing task Kim and Ahn (2002) asked clinicians to indicate how they believed symptoms were causally related for

DSM-IV disorders. They found that clinicians weight symptoms that in the DSM-IV receive equal weights differentially, depending on the position of the symptoms in their theories. Specifically, hypothetical clients presenting with symptoms that are more causally central in clinicians' theories, meaning that more symptoms depend on them, were more often diagnosed as having a specific disorder than clients presenting with symptoms that are causally peripheral or isolated in the theories. For example, in clinicians' theories for Anorexia Nervosa the symptom "refuses to maintain weight" causes several other symptoms of Anorexia Nervosa (e.g. excessive exercise, dieting, preoccupied with food), while "absence of the period for more than 3 months" does not cause any other symptoms. The first causally central symptom was found to be more important for classification than the latter causally peripheral symptom. Further, causally central symptoms were recalled more often, and falsely recognised more often, than causally peripheral or isolated symptoms. Thus, a causal status effect from clinicians' theories was found in clinicians' classifications (see also Cobos, Florez, & López, 2008), similar to the causal status effect found in categorisation of natural kinds and artefacts (Ahn, 1998; Ahn, Gelman, et al., 2000; Ahn, Kim, et al., 2000).

It should be noted that since the theories clinicians drew in the studies of Kim and Ahn (2002) contained only symptoms stated in the DSM-IV, and the DSM-IV refrains explicitly from specifying aetiological theories, these theories do not contain many deep root causes. On the other hand, Kim and Ahn's findings are especially interesting given that clinicians are trained to use this atheoretical classification system.

Theories about how symptoms are causally related may affect clinicians' classifications in another way too. Students decided that patients with symptoms that stemmed from a common cause via separate causal paths were more likely to have a hypothesized medical condition than patients with symptoms that stemmed from this common cause via the same causal path. This is called a "causal diversity effect" (Kim & Keil, 2003). Similarly, to investigate whether a patient had a medical condition students preferred to find out whether causally diverse symptoms were present rather than symptoms with similar origins (Kim, Yopchick, & De Kwaadsteniet, 2008). So far, this causal diversity effect has only been shown with psychology undergraduates reasoning about medical conditions. Experienced clinicians may however also show a preference for information which is causally diverse in theories about psychiatric disorders, above information with similar causal origins.

Finally, Proctor and Ahn (2007) found an effect of the causal structure of knowledge on clinicians' inferences, similar to the causal status effect in classifications. They found that clinicians make inferences from symptoms that are known to be causes rather than from symptoms that are known to be effects. For example, if clinicians learned that large mood swings cause excessive social anxiety in a client, they thought it was more likely that this client would engage in reckless behaviour, than when they learned that the excessive social anxiety caused the large mood swings. On the other hand, clinicians thought a client was more likely to have a tendency to blush when they learned that this client's excessive social anxiety caused large mood swings, than when the relation was reversed (Proctor & Ahn, 2007). Thus, the structure of causal beliefs affects clinicians' classifications, and may affect their inferences.

Summary

Causal reasoning seems basic to how people process information and make decisions, especially in complex situations. Research has shown that laypeople and professionals construct causal mental representations of information, that they make causal inferences while doing so, and that they base their judgments and intervention choices on them. Further, classification decisions are affected by the causal structure of people's theories. It remains unsettled however whether practicing mental health clinicians engage in causal reasoning when they process information to understand their clients and to decide upon the most effective interventions. Knowledge of how symptoms are causally related in a disorder or with a specific client has been found to influence clinicians' classifications and inferences. But at the same time, mechanisms explaining clients' problems are hardly specified in clinicians' case formulations. Also it is not clear how clinicians arrive at their intervention choices. The question how clinicians decide upon interventions has hardly been addressed. In all, evidence seems lacking for Meehl's clinicians-as-mechanics idea.

This thesis: Clinicians as mechanics?

This thesis contains studies that aimed to learn more about how clinicians reason. So, in Meehl's words their goal is "that of giving a behavioral description of what the clinician does" (Meehl, 1954, p.37). In particular, these studies address the question whether, as mechanics, clinicians construct causal client representations reflecting the mechanisms they assume to underlie clients' problems, when they process client information, decide on what may the matter with a client and what may be effective interventions, and make classifications. Although there is a lot of evidence that underlines the importance of causal representations in reasoning and decision making, the central role of causal representations is not yet evident in clinicians' reasoning. Knowledge of clinicians' reasoning processes will be useful for the development of methods to improve clinicians' performance that may address their natural reasoning strategies (Witteman et al., 2007). The following chapters present studies aimed to elicit the role of causal representations in clinical judgment and decision making, using different methods.

Chapter 2 describes a think aloud study in which we elicited clinicians' thoughts while they reasoned about what could be the matter with a client and what they would propose as follow-up steps. The think aloud method is used to tap into people's reasoning. It is assumed that people can reliably verbalise the information they are attending to in their short term memory, if they can do so directly (and no retrieval processes are needed due to a too large delay in time) and if what they are attending to can be verbalised (of automated processes for example only input and output can be verbalized, not what happens in between) (Ericsson, 2006; Ericsson & Simon, 1993).

We presented participants with the intake video and with information from a battery of tests. We analysed the thoughts clinicians verbalised after they had received all this information, in two ways. First, we analysed the causal structure of participants' mental representations, similar to how Pennington and Hastie (1986) extracted the causal structure of jurors' mental representations. We also analysed to what extent the follow-up steps that participants proposed were related to their mental causal models. Complementary to this

structure analysis, we analysed the contents of participants' verbalisations, to see what specific causal factors participants mentioned, and what follow-up steps, similar to how Eells and colleagues (1998) analysed the content of clinicians' case formulations (see also Eells, Lombart, Kendjelic, Turner, & Lucas, 2005). Thus, we aimed to extract clinicians' causal mental client representations, and see whether they base follow-up steps on these.

Chapter 3 describes the analysis of clinicians' verbalisations made while watching the intake video. We categorised the inferences participants made while processing information, in analogy to the categorisation of inferences of think aloud protocols from readers in text comprehension research (see Trabasso & Magliano, 1996). We aimed to see whether clinicians make more causal inferences, compared to descriptive and predictive inferences. Additionally, we analysed the specific contents of participants' causal inferences, to see whether clinicians make the same.

In Chapter 4 a study is presented in which we further investigated whether clinicians base their intervention choices on causal models. We analysed to what extent participants' ratings of the effectiveness of different interventions could be predicted from their causal client representations. To extract participants' causal client representations we used a cognitive mapping method. Cognitive mapping is a method which elicits how participants think about a problem by asking them to draw causal relations they perceive between relevant variables (see e.g. Hodgkinson, Maule, & Bown, 2004). In contrast to the extraction of mental causal representations from think aloud protocols as described in chapter 2, which can be seen as an indirect method, cognitive mapping directly asks for clinicians' causal beliefs.

Chapter 5 describes a series of studies in which we investigated how the causal structure of theories about disorders may affect clinicians' classifications. In particular, we aimed to find out if experienced clinicians who reasoned about psychiatric disorders would show a causal diversity effect. We looked if experienced clinicians rate evidence that is causally diverse stronger than evidence with causally similar origins, according to theories for disorders from peer reviewed literature. We also looked if experienced and novice clinicians rate evidence that is causally diverse stronger than evidence with causally similar origins, according to their own theories.

The last chapter, Chapter 6, summarises and discusses the findings of the studies presented in chapters 2 to 5. It ends with implications for practice and suggestions for further research.

Chapter 2

Causality in Clinicians' Mental Representations of Client Information
A Think Aloud Study

Abstract

We investigated whether mental health clinicians mentally represent client information in causal models, and whether they base their interventions on these causal models. We showed 20 clinicians an intake video and test results, and asked them to think aloud about these questions: "What do you think is the matter with this client; what would you propose as follow-up steps?" Both causal structure analysis and content analysis of participants' verbalizations indicated that their mental representations were predominantly descriptive and lacked causal complexity. Also, most verbalizations about further assessment and interventions did not seem to be related to participants' causal models. Finally, we found large variation in the causal factors that participants mentioned and in the interventions they proposed. We discuss explanations for our findings and implications, and suggest further research.

Studies have shown that the reliability and validity of clinical decisions is poor (Dawes, 1994; Garb, 1998; Grove, Zald, Lebow, Snitz, & Nelson, 2000; Meehl, 1954), and that experienced clinicians often do not outperform inexperienced clinicians (Dawes, 1994; Eells, Lombart, Kendjelic, Turner, & Lucas, 2005; Garb, 1998; Krol, De Bruyn, & Van den Bercken, 1992; Lichtenberg, 1997). Many studies on clinical decision making have compared diagnostic judgments, such as DSM-classifications or behaviour predictions, with norms, such as established e.g. by expert consensus or policy capturing. Results show that clinicians' judgments deviate from these norms, and thus it is concluded that they perform poorly (Caspar, 1997; Dawes, 1994; Nurcombe & Gallagher, 1986). An important limitation of such studies is that they do not reveal underlying processes. Studying the clinical decision making process can provide useful knowledge for the development of educational methods and decision support, which may effectively improve decisions (Caspar, 1997; Caspar, Berger, & Hautle, 2004; Galanter & Patel, 2005; Garb, 1998; Nurcombe & Gallagher, 1986; Witteman, Harries, Bekker, & Van Aarle, 2007). Unfortunately, process studies which reflect the dynamics and complexity of the clinical situation are rare (Caspar, 1997; Groth-Marnat, 2003; Nurcombe 2000).

In complex, dynamic situations the construction of a coherent mental representation seems crucial for decision making: the right decision depends on a correct understanding of the situation (Endsley, 1997; Pliske & Klein, 2003). It is assumed that if problem information concerns human events, mental representations will be structured as causal models or stories, containing causal and temporal relations (Beach, 1992; Goldstein & Weber, 1995; Hastie & Pennington, 2000; Rettinger & Hastie, 2003). Causal mental representations enable us to understand, predict what will happen and, perhaps more importantly, what the effects will be of our interventions (Beach, 1992; Sloman, 2005; Sloman & Hagmayer, 2006).

The theory of Explanation-Based Decision Making (Hastie & Pennington, 2000) describes how in complex domains decisions result from causal reasoning about information. According to this theory, decision makers use previously acquired knowledge and experience to construct causal explanations which integrate information, direct inferences, and on which, when completed, decisions are based. Pennington and Hastie (1986; 1988) found that American jurors who were presented with large amounts of ill-structured evidence information constructed specific causal models: stories. These stories are summary representations in which evidence information is linked causally and temporally. Jurors use their general world knowledge to make inferences to complete the representations. Jurors' verdicts follow from their stories: different story representations appeared to lead to different verdicts. Furthermore, confidence in a verdict was found to be determined by the coherence and the uniqueness of the constructed story.

Text processing theorists also assume that most comprehension, not only of narrative texts, is essentially explanation-based (Graesser, Olde, & Klettke, 2002; Magliano, 1999). They assume that people process information by constructing causal mental representations which go beyond the information actually presented, and that they use background knowledge to make inferences which provide local and global coherence to the constructed situation models. Think aloud studies showed that explanation-based inferences are fundamental and are the most prevalent inferences readers make, together with

referential inferences made to determine the referents of pronouns used (Magliano, 1999; Trabasso & Magliano, 1996).

Another model that subscribes to the importance of a causal mental representation of information for decision making in complex contexts is the Recognition-Primed Decision Model (Pliske & Klein, 2003). This model is based on studies of experienced professionals in several domains. According to the RPD-model, decision makers try to causally explain a situation which is not immediately understood, by means of feature matching and story construction. Decision makers link observed events to causal factors and use mental simulation to predict consequences of possible actions.

In the clinical domain, deciding what is the matter with a client and how (s)he should be treated is a complex and dynamic task of assembling and integrating unstructured, incomplete information, from several sources. Based on the theories described above, we expect clinicians to construct causal models: integrative representations of client information in which symptoms, behaviours, and events are causally and temporally understood. We further expect that these causal models precede and determine clinicians' decisions. This seems in accordance with characterizations of clinical thinking given by psychiatrists: "What's wrong, how did it get it that way, and what can we do about it?" (Nurcombe, Drell, Leonard, & McDermott, 2002b, p. 350), and: "I want to make a story in which these symptoms will make sense and fit together in a temporal sense." (Nurcombe, Drell, Leonard, & McDermott, 2002a, p. 94). The identification of causal variables and the integration of information in a causal, explanatory model is considered crucial in clinical assessment, because these (modifiable) variables affect the effectiveness of treatment (Eells, Kendjelic, & Lucas, 1998; Haynes, 1992; Haynes & Williams, 2003; Kuyken, Fothergill, Musa, & Chadwick, 2005; Nurcombe, 2000). In this study we aimed to identify causal structures in clinicians' mental representations of client information, and to assess the relation of the proposed follow-up steps to these causal structures. In addition, we aimed to identify what causal factors and follow-up steps clinicians consider, and whether they agree about these.

Method

To elicit clinicians' mental representations we used the think aloud method. This method can provide insight in thought processes during problem solving (Ericsson, 2006; Ericsson & Simon, 1993; Van Someren, Barnard, & Sandberg, 1994). When thinking aloud participants directly verbalize information which is attended to and activated in short term memory. It is claimed that thinking aloud does not change the thought process, except that it moderately slows it down (Ericsson & Simon, 1993).

Participants

Twenty clinicians with different levels of experience participated. We approached clinical child psychologists who had participated in an earlier study and from our network. Four participants were novices: master students of the faculty who had finished nine months internships in which they also had been engaged in diagnostic tasks. All novices were women with a mean age of 25.3 years (SD = 2.6). The other participants, twelve women and four men, had a mean experience of 13.3 years (SD = 9.8), and a mean age of 41.3 years (SD = 12.3).

Materials

We presented participants with a 33 minutes video of an intake interview of two diagnosticians with a nine year old boy, Eric, and his parents. Eric and his parents had been referred to a Dutch youth care institution because of the boy's problematic behaviour at home and at school. We additionally presented the results of nine tests administered to the boy (a developmental history questionnaire, the Wechsler Intelligence Scale for Children (III), the Child Behaviour Check List, the Teacher Report Form, a Stroop test (measuring inhibition capacity), the Bourdon-Vos (measuring concentration span), the Family Relations Test, the Theory Of Mind test, and a test measuring children's perceived competencies). We also included an observation report of a free play situation and a summary of the observations of the boy's behaviours during the tests. The test results presented were copies of the original forms. The parents had given us written consent to use the video and test materials for our study. The case was chosen after consultation with two experienced child psychologists, who both judged this case to be representative and with sufficient complexity.

Procedure and task

All task instructions were typed out, to insure that all participants received the same instructions. Participants were free to read the instructions aloud or silently, and could always ask for clarification. In the first instruction we told participants that they would be asked to think aloud, that is: "to verbalize all your thoughts while they are coming up. You do not have to think about what you want to say and you do not have to explain your thoughts. You just have to verbalize immediately what you are thinking. As in 'think talking', let your thoughts speak. You could imagine yourself sitting alone in a room and talking to yourself." (cf. Ericsson & Simon, 1993; Van Someren et al., 1994). After participants had read this instruction, the experimenter gave them two exercises in thinking aloud.

In the second instruction we explained to participants that they would see a video on a laptop showing an intake conversation featuring a nine year old boy and his parents, and we asked participants to verbalize all thoughts they had while they were watching this video. We explained how they could interrupt the video by pushing the space bar whenever they wanted in order to verbalize thoughts. We asked participants to think aloud about the following questions: "What do you think is the matter with this boy; what would you propose as follow-up steps?" We had chosen these specific questions, after discussions of our study design with experienced child psychologists, to elicit clinical reasoning as it would occur in practice. In practice these questions seem to be the ones most frequently asked by parents. Further, we expected that these questions would elicit causal reasoning in participants, in two ways. First, although the first question may seem to ask for descriptive information, parents with this question ask for an explanation of their child's problems. Second, as we argued in the introduction, treatment planning should be based upon a causal understanding of a client's problems.

We told participants that all their verbalizations would be tape recorded and that all data would be treated anonymously. We emphasized that what they said would not be judged as being right or wrong. When the participants had indicated that they understood the instruction, the experimenter started the video. If participants did not seem to be

verbalizing their thoughts, the experimenter reminded them to think aloud (Ericsson & Simon, 1993; Van Someren et al., 1994).

The boy's test results were given when participants had finished talking at the end of the video. Participants were instructed to keep thinking aloud about the questions while reading the test results. They could choose to read the results aloud or silently. If necessary, the experimenter reminded participants to think aloud. When they had finished reading and talking, the experimenter asked participants to summarize their thoughts about the initial questions. All participants' verbalizations were recorded on audiotape.

After the experiment, the participants received a short form with questions asking for their age, gender, education, years of experience, work setting, number of cases seen which were similar to the presented case, confidence in their judgments, and whether they adhered to a specific theory.

The experienced participants received a 50 euro gift certificate for their participation; students received 10 euro in cash. The experiment took one and a half to two hours per participant. The first author administered the task mostly at the participants' work settings. The students and two of the experienced participants visited the first author in her office.

Analysis of the think aloud protocols

All verbalizations were typed out verbatim in protocols for each participant. We extracted the structure and content of participants' mental representations from the verbalizations they made after they had watched the video and studied all test results, including the summaries they gave. First, we analyzed the causal structure of diagnosticians' mental representations, analogous to Pennington and Hastie (1986) who extracted jurors' mental representations of trial information from think aloud protocols obtained after the jurors had seen a video of an enacted murder trial. Next, we investigated whether the follow-up steps the clinicians proposed were related to their mental causal models. Third, we analyzed the contents of diagnosticians' verbalizations using the Case Formulation Content Coding Method (CFCCM) from Eells, Kendjelic, Lucas, and Lombart (received in personal communication, October 2007, see Eells et al., 1998; Eells et al., 2005). We looked at what causal factors clinicians considered, and what follow-up steps.

Causal structure analysis

Case statements. Two coders independently identified *case statements* in the protocols, in analogy to Pennington and Hastie's *explicit story references*. The coders did not have knowledge of the presented case. We defined case statements as all verbalizations concerning Eric and other persons in Eric's life. Coders did not distinguish between information actually given and inferences made by participants. They did not code references to events in the intake interview as case statements, in analogy to the exclusion of references to events at the trial from Pennington and Hastie's explicit story references. For example, "Mother says" is not a case statement, however what she is saying can be a case statement: "(that) he is always busy and often gets angry". Further, we defined case statements *not* to be: verbalizations concerning the think aloud task or concerning participants' own thinking processes, evaluations of information (e.g., "This is important"), open questions (for more information), and incomplete or incomprehensible verbalizations. We also excluded verbalizations concerning further assessment and treatment proposed,

and their expected effects. These were identified separately (see below: *Analysis of proposed follow-up steps*).

Case states. The first author segmented all identified case statements into separate *case states*: simple sentences. For example, the case statement "he is always busy and often gets angry" was divided into two states: "he is always busy" and "(he) often gets angry". A list of states resulted for each participant. Both coders then checked whether they saw repetitions of the same states in the lists, for each participant separately, and marked all repetitions.

Links. Subsequently, both coders identified *temporal* links and *causal* links in the lists of states of all participants. Coders identified links by connective words such as *because*, *therefore*, *then*, or by verbs such as *causes* and *reinforces*. We defined a temporal link as a link which expresses a sequence of states in time. We also considered words that relate a state to a specific moment in time a temporal link (e.g., "*at the age of five* he moved"), although this does not link two states, but only indicates at which point in time a state occurred. We defined a causal link as a link between two states, in which one state is a cause of the other. Our definitions are similar to the semantic relationships in stories which Rumelhart (1975) defined. In our definition of causal links we also included intentional links, relating a character's intention to an action (cf. "Motivate", Rumelhart, 1975). A frequently used link is the *and* link, expressed by words such as *and*, *also*, and *as well*, which indicates a combination of two or more states. This combination may be solely an enumeration of more than one state. Although not typical for stories or causal models, coders did identify *and* links, because these may also announce an enumeration of more than one state causing, being caused, or being linked temporally. The coders also decided which states the links related. To do so, coders were allowed to look in the original protocol text to be able to correctly interpret the extracted states and links.

Graphs. The authors constructed *conceptual graphs* from the states (repetitions excluded) and links identified by the coders, to depict the causal structures extracted from the protocol text. Conceptual graphs consist of case states presented in boxes and links between these states presented as arrows between the boxes (cf. Pennington and Hastie, 1986). *And* links were only taken into account when they were not merely an enumeration of information, but meant for example *also caused* or *also after*. Consider these three case states and their links: "because (causal link) he always acts so impulsively / he frightens children / and (*and* link) he gets involved in fights". Here the last state is considered to be caused by the first state, as is the second state.

Analysis of follow-up steps

Follow-up statements. The two coders went through the protocols again, this time to identify all verbalizations that suggested additional assessment, treatment plans, information expected to be gained from further assessment, and effects expected from treatment: the *follow-up statements*.

Follow-up states. The first author again segmented all follow-up statements in semantic units: simple sentences. For example, the follow-up statement "Refer him to another school so he becomes less special" was divided into: "Refer him to another school"

and “so he becomes less special”. For each participant this resulted in a list of *follow-up states*. The coders again checked for repetitions and marked these in the lists.

Links. Subsequently, the coders identified *causal*, *temporal*, and *and* links in all lists of follow-up states. Additionally, they identified *reason* links and *referential* links. We defined reason links as the words referring to the reasons participants gave for their follow-up states. Typical reason links are words such as *so*, *therefore*, and *because*. A reason provided for a specific follow-up state can be a case state, for example when the reason link relates a treatment plan to a content state (e.g., “the boy is socially incapable (case state), / therefore (reason link) he should engage in a training in social skills (follow-up state)”). A reason can also be another follow-up state, if the reason link refers to what effect is expected from treatment, or which information is expected to be gained from assessment (e.g., “observe the boy in class / in order to (reason link) see how he reacts to his classmates”). We defined a referential link as words referring to something stated in another follow-up state or case state. Typical referential links are recognized by words such as *here*, *this*, and *where*. For example, in the follow-up state “further investigate those physical aspects” the words “those physical aspects” refer to the case states: “there were troubles with his birth / and *(and* link) he fell on his head”. Coders also decided, again using protocol texts if needed for correct interpretations, which states the links related.

Graphs. The authors connected the follow-up states identified (repetitions excluded) by means of the identified links to each other and to case states. When a reason or a referential link connected a follow-up state to a case state in a participant’s conceptual graph resulting from the first analysis, we added this follow-up state, and the follow-up states connected to this follow-up state, to that participant’s conceptual graph. Here too, we only took *and* links into account when they meant *also temporally linked*, *also causally linked*, *also linked by reason*, or *also linked referentially*.

Agreement. We calculated interrater agreement for the identification of case statements and follow-up statements on the basis of agreement on words in five randomly selected protocols. It resulted in Cohen’s kappa of 0.80 for case statements and 0.78 for follow-up statements. We calculated agreement on repetitions on the basis of five protocols, which resulted in Cohen’s kappa of 0.79 for case states and 0.77 for follow-up states. These kappas can be considered substantial (Landis & Koch, 1977) or excellent (Fleiss, 1981). We calculated agreement on the links by dividing the number of links agreed on by the total number of links identified by both coders[1]. On the basis of five protocols, agreement was 91.2 % for the links in case states and 82.8 % for the links in follow-up states. For 87.1 % of the links agreed in case states and 89.6 % of the links agreed in follow-up states, coders decided similarly about which states were related. After each coding step, the first author marked the differences between coders and presented them to the coders. Usually coders then came to a shared conclusion about these differences. If not, differences were resolved by the first author.

[1] We did not calculate kappa-coefficients here, as there is no plausible way to determine expected agreement.

Content Analysis

This analysis we performed following the procedure described in the CFCCM: "Manual for case formulation and treatment plan coding", by Eells, Kendjelic, Lucas, and Lombart (received in personal communication, for a description of this coding method, see Eells et al., 2005). Eells and colleagues developed this coding method to reliably analyze the specific contents and quality of written and 'think aloud' case formulations of psychotherapists, who may adopt different theoretical frameworks. We translated the CFCCM into Dutch.

Idea units. Three coders segmented all verbalizations into idea units. In the CFCCM, Eells et al. define idea units as "the expression of a complete thought", following the segmentation procedure that Stinson, Milbrath, Reidbord, and Bucci (1994) proposed for segmentation of psychotherapy transcripts. This segmentation procedure, based on intuitions of what "complete ideas" are, has been reliably used by Stinson et al. (1994) and Eells et al. (2005). Idea units are of equal size or larger than the semantic units we identified in the structural analysis: case states and follow-up states are simple sentences, idea units can consist of one or more simple sentences. We based the final segmentation of the protocols on coders' agreement about the beginnings of new units. Each time two or three coders agreed on the beginning of a new idea unit, this was taken as such.

Categories and elements. In the CFCCM Eells et al. distinguish four major coding categories. Category A, descriptive information, represents idea units that convey information presented in the case description. Considerations of a diagnosis in terms of DSM-classifications are coded under category B. Category C, inferred information, is used for idea units that contain new information inferred from the presented information. Considerations of causal factors are coded in this category, that is: idea units expressing ideas about psychological, biological, or social mechanisms and predisposing and precipitating factors. We considered strengths and potential therapy-interfering factors as causal too. Finally, category D represents idea units considering treatment planning and further assessment, whether these are formulations of specific treatment types (e.g. individual cognitive therapy) or more general foci of attention. An idea unit can be assigned to one to four of the main categories at the same time. For example, the idea unit: "His poor fine motor skills fit with Non-verbal Learning Disorder" is assigned to two categories: category A (descriptive information) and category B (diagnostic information). Idea units which could not be assigned to any category (e.g. participants' verbalizations about their thought processes, or the task, or incomprehensible verbalizations) received the code *NC* (no code), and were not used in any further analyses.

Each of the four major categories of the CFCCM consists of numbered elements which further specify the kind of information expressed. Within a category, an idea unit can be assigned to only one specific element. Thus, the three coders assigned each idea unit to one element within one or more categories (or coded it as NC). All differences between the coders were resolved in discussion.

Sub-elements. The more specific sub-elements of elements in the CFCCM were not always appropriate for a child's case formulation; it seems that the CFCCM is mainly written for case formulations for adult clients. For example, in the CFCCM separate sub-elements

are defined for predisposing experiences in childhood and predisposing experiences in adulthood. Also, a frequently used intervention in children's cases is parent support, and this was frequently mentioned by our participants. This is however not represented in the CFCCM among the different treatment types. Therefore, instead of using the sub-elements of the CFCCM, we had two coders identify the specific contents of the idea units which were coded as elements in categories B, C, and D.

For category B, the coders identified different classifications mentioned for DSM-IV axis I (for axis II, participants mentioned no specific classifications). For the causal elements coded in category C (inferred information) and the elements coded in category D (treatment planning) coders identified which different predisposing experiences, current stressors, psychological, biological, and social mechanisms, strengths, potential interfering factors, and which different treatment types or foci of attention participants mentioned. For the mechanisms (psychological, biological, social), coders identified the different causes within the mechanisms, not the effects. First one coder went through all protocols, creating new sub-elements for each idea unit expressing a content not identified before. Next, the second coder coded the idea units using the first coder's distinction of sub-elements, adding or merging sub-elements as she thought appropriate. Because participants considered many different causes and foci of further research and treatment, coders distinguished between clinical features of cognition, affect, behaviour, physical condition, and interpersonal adjustment, following Carr (2006, Table 11.2). Also, for social factors, coders distinguished between parental factors, family factors and school factors, similar to Carr's distinction of contextual factors (2006, Figure 2.1). Coders resolved disagreements in discussion.

Agreement. We calculated agreement about segmentation into idea units using the same formulas as Eells et al. (2005) and Stinson et al. (1994) did, from Scott and Hatfield (1985). For five randomly chosen protocols, the three coders arrived at a mean of 71 % pairwise agreement (ranging from 57 to 81 %), using the conservative formula (A / (A + D), A meaning beginnings of new units agreed on, D beginnings disagreed on), or 83 % (ranging from 72 to 91 %) using the less conservative formula (2 A / (x + y), where x and y are the total number of beginnings for two coders). To establish content coding agreement, we calculated multirater kappas for five randomly chosen protocols, using the formula from Siegel and Castellan (1988) (cf. Eells et al., 2005). This resulted, for coding in elements, in a mean kappa of .64 (varying from .54 to .76, for different protocols). To compare, agreement in Eells et al. (2005) amounted to kappa values of .61, .81, .62 and .69 for coding into elements, for the four major categories. As in Eells et al. (2005), our coders seemed to agree more on their assignments of idea units to main categories, than on assigning idea units to specific elements within categories: the mean multirater kappa for agreement about categories was .79 (varying from .74 to .84). To compare, kappa agreement values in Eells et al. (2005) were .61, .81, .62 and .69, for coding into categories, for the different categories[2]. For assigning idea units to sub-elements coders arrived at 90 % agreement, calculated for all

[2] We did not calculate kappas for the different categories separately, because we based our agreement calculations on fewer idea units than Eells et al. (2005) did. Especially in category B there were relatively few idea units, which in the five random protocols were exclusively assigned to one element (Axis I classification).

idea units assigned to elements in category B, explanatory elements in category C, and elements in category D (99 % for category B, 71 % for category C, and 90 % for category D).

Results

Causal structure

Case statements

Participants verbalized on average 602 words (*SD* = 218) after they had watched the video and studied the test results. Of these words on average 28.2 % (*SD* = 9.4) were contained in *case statements*. Segmentation of the case statements resulted in a mean of 17.2 *case states* (*SD* = 5.9) (repetitions excluded). Participants verbalized on average 2.7 *causal* links (*SD* = 2.6), 0.5 *temporal* links (*SD* = 1.0), and 3.6 *and* links (*SD* = 2.1). On average 6.2 case states (*SD* = 4.7), or 32.8 % (*SD* = 22.3), were linked by means of causal or temporal links, or the *and* links which meant *also causally linked* or *also temporally linked* (see method).

Linked case states were visualized in *conceptual graphs*. Three participants had not verbalized any causal or temporal links, so for them no conceptual graphs could be constructed. For nine participants conceptual graphs consisted of more than one disjoint part. The length of the longest chain, that is: the longest connected path in a conceptual graph, consisted on average of 2.4 linked case states (*SD* = 1.5), for six participants the longest chain consisted of more than two case states. To illustrate, Figure 1 and Figure 2 show the conceptual graphs we constructed for two participants. Case states that were not causally or temporally linked are listed beneath the graphs. Figure 1 shows a conceptual graph that consists of two disjoint parts. In Figure 2 the longest chain consists of six case states, which is the longest of all participants' chains.

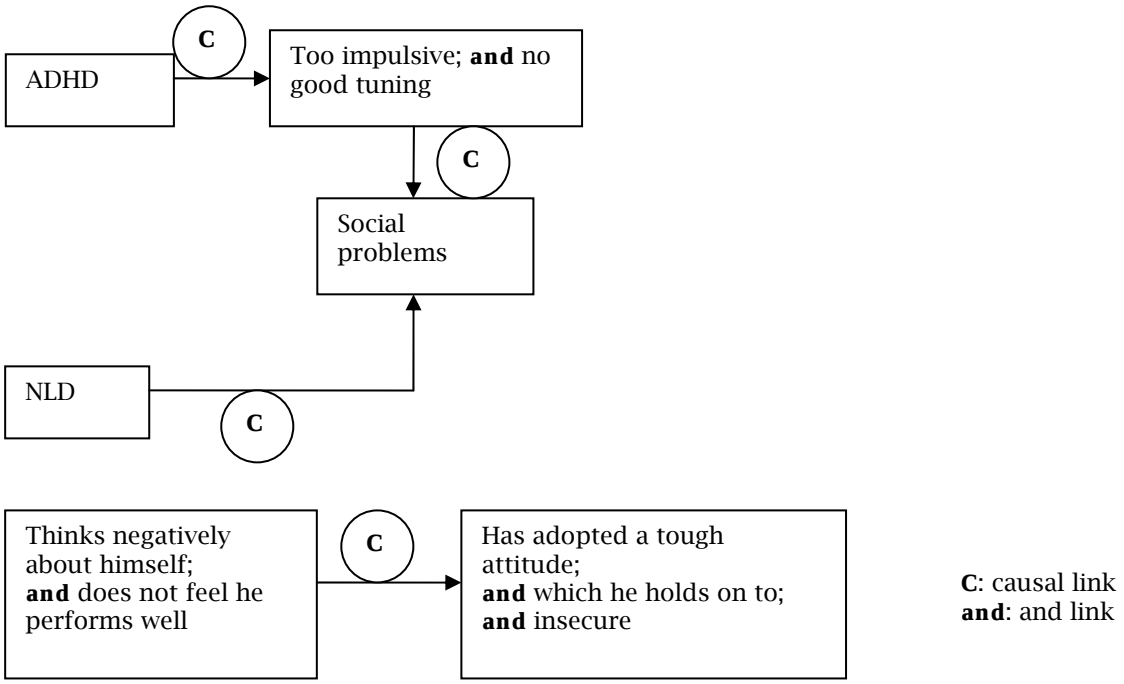
Follow-up steps

On average 26.0 % of participants' verbalizations (*SD* = 12.7) were contained in *follow-up statements*. Segmentation of the follow-up statements resulted in a mean of 11.2 *follow-up states* (*SD* = 6.1) (repetitions excluded). Participants verbalized on average 0.3 *causal* links (*SD* = 0.7), 0.7 *temporal* links (*SD* = 1.2), 4.3 *and* links (*SD* = 3.3), 3.0 *reason* links (*SD* = 1.5), and 1.4 *referential* links (*SD* = 1.8). Here, temporal links between follow-up states represent the order of follow-up steps proposed by a participant, or that some information or effect was expected after some assessment or some treatment.

On average 7.8 follow-up states (*SD* = 4.6), or 70.8 % (*SD* = 23.6), were linked by means of reason links, referential links, causal links, temporal links, or the *and* links which meant *also causally linked*, *also temporally linked*, *also linked by a reason link*, or *also referentially linked*. However, not all linked follow-up states were linked to case states, and not all follow-up states linked to case states were linked to case states in a conceptual graph. On average 5.6 follow-up states (*SD* = 4.0), or 55.9 % (*SD* = 36.0), were linked to case states. This percentage includes follow-up states that are indirectly linked to case states: follow-up states that are causally, temporally, by reason, or referentially linked to other follow-up states which are connected to a case state. On average 3.4 follow-up states (*SD* = 3.8), or 32.8 % (*SD* = 35.3), were linked to case states in a conceptual graph, including follow-up states which were linked indirectly.

Three participants had no follow-up states linked to case states. Eight participants had no follow-up states linked to case states in a conceptual graph. Three of these were the participants for whom we had not been able to construct a conceptual graph in the first analysis, as they had not verbalized any causal or temporal links between case states.

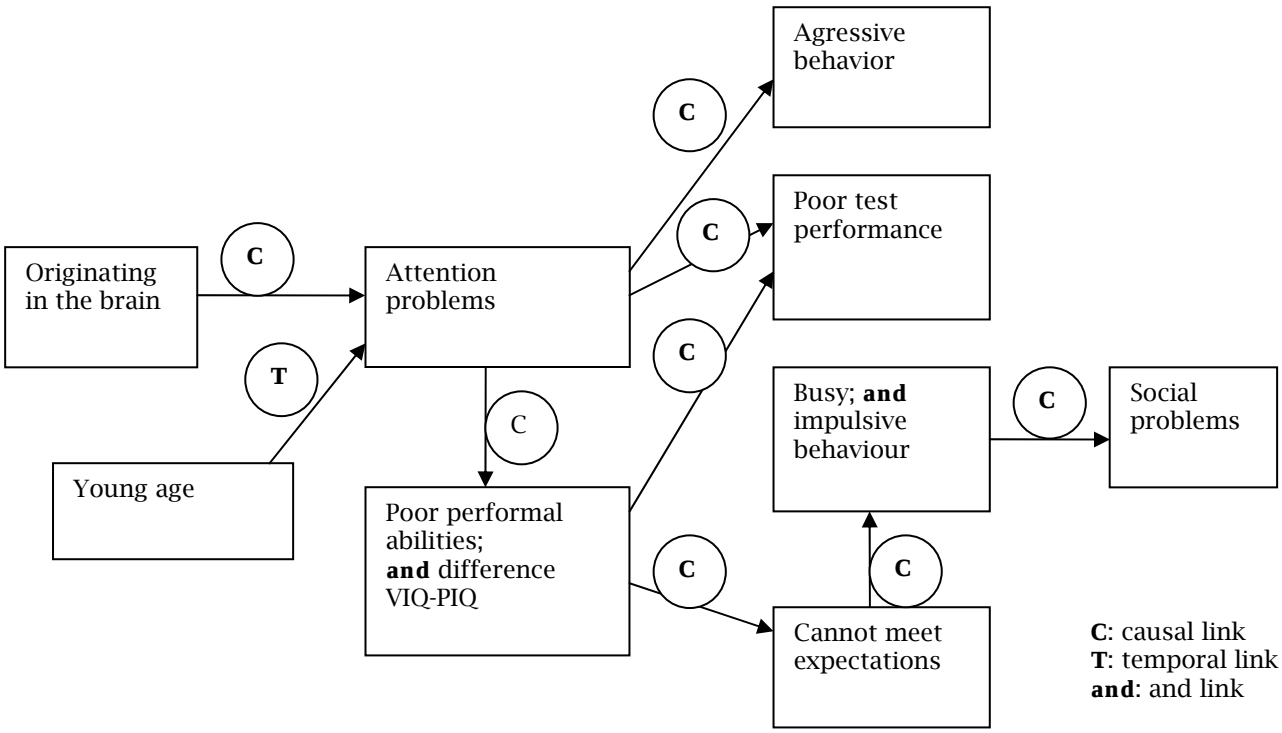
Figures 3 and 4 show how follow-up states are linked to case states in the conceptual graphs shown in Figures 1 and 2, as well as follow-up states linked to case states which were not part of the conceptual graphs for these participants.



Unlinked case states:

- Interference problems
- Concentration problems too
- Experiences social competence as good
- Other areas less (competent)
- WISC scores performal worse than verbal
- In TRF he comes just under the limit
- Negative atmosphere at home
- He does not experience this so (negatively)
- He can express his emotions

Figure 1. Conceptual graph and unlinked case states for one participant.



Unlinked case states:

- Sufficient social insight
- Axis-disorder is excluded
- Not clearly ADHD
- Also problems with hyperactivity
- With structure he performs better
- Has most difficulties with math
- His Bourdon-Vos is not so bad
- Very weak at the Stroop
- Counteracts
- He misses certain insights
- And he misses integrative capacity
- He cannot handle daily practical things

Figure 2. Conceptual graph and unlinked case states for another participant.

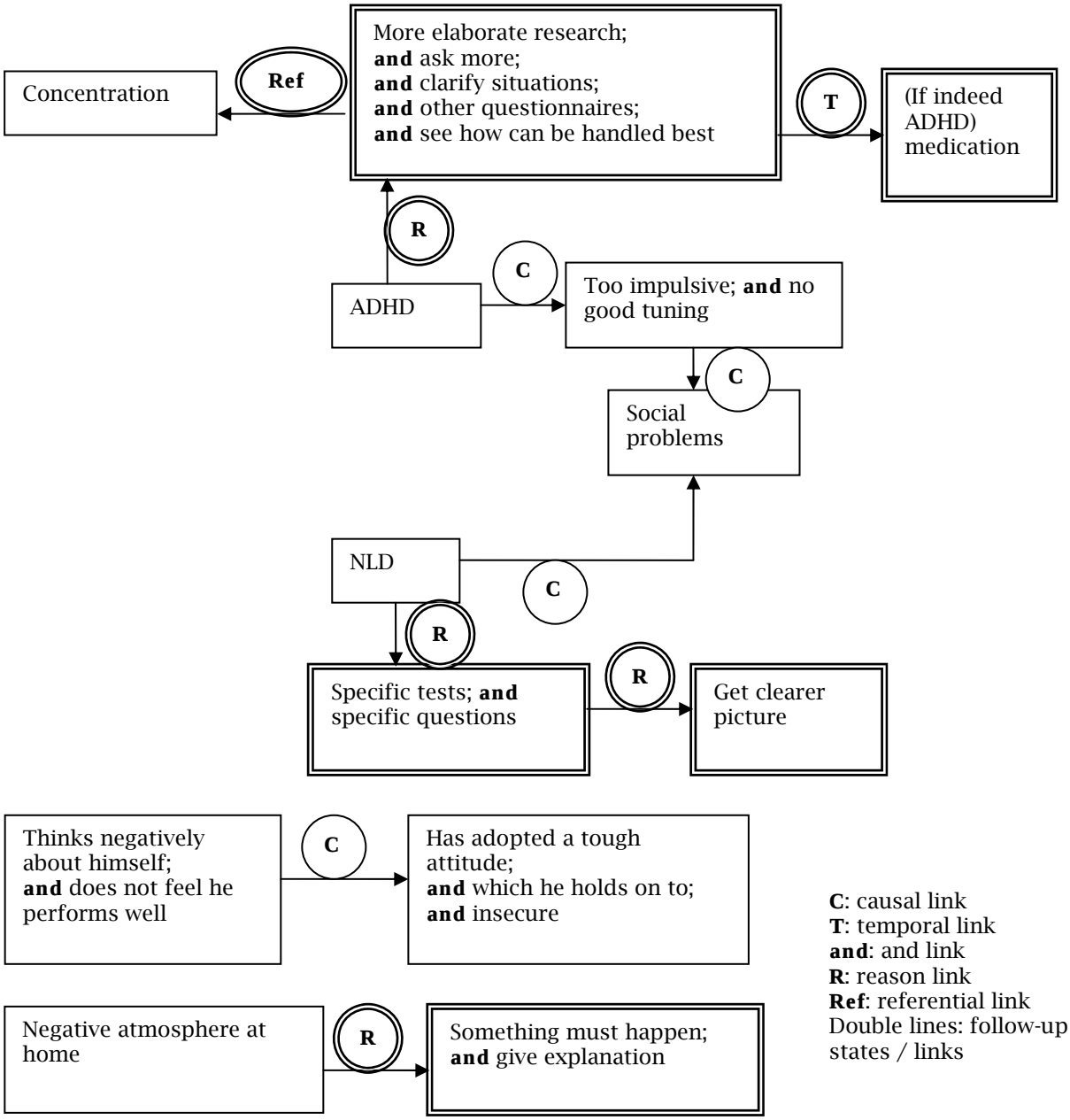


Figure 3. Follow-up steps added to conceptual graph and unlinked case states for participant of Figure 1.

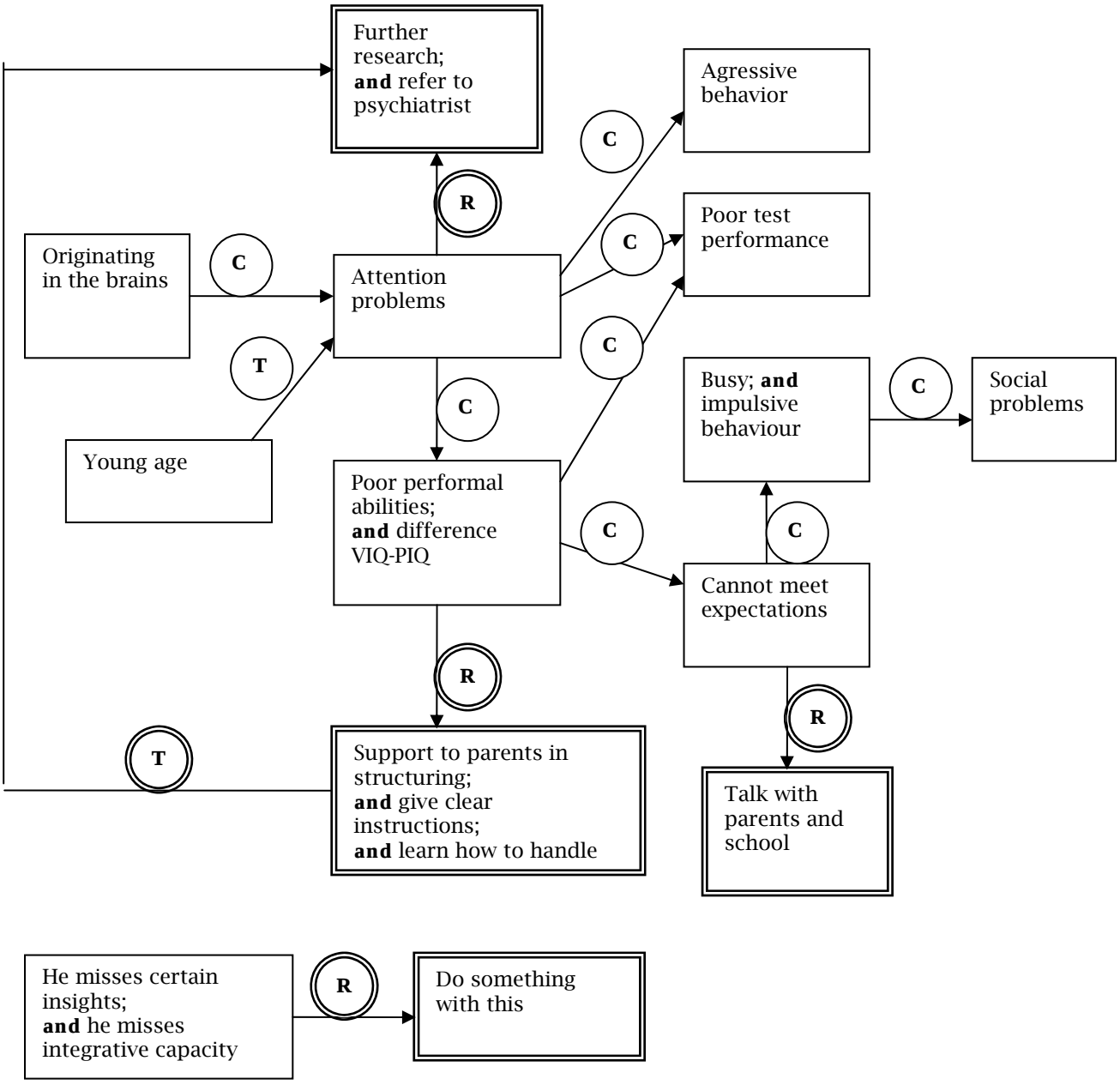


Figure 4. Follow-up steps added to conceptual graph and unlinked case states for participant of Figure 2.

Background variables

Because results show a large variation in the proportions of case states and follow-up states that are linked, we correlated these outcome measures with participants' experience (measured by the number of similar cases participants had seen before, although years of experience gave similar results). Experience was not significantly correlated with these outcomes, nor with the number of causal links, or the length of the longest causal path (Spearman's rank order correlations were between -.25 and .09; all *p*'s > .2). When asked for their confidence in their judgments of what was the matter with the boy and in their proposed follow-up steps, participants scored on average 5.2 and 5.3 respectively, on a scale of 1 (very little) to 7 (very much). Although confidence was not significantly correlated with the proportions of linked states, it was significantly correlated with experience: the more experienced with similar cases, the more confident participants were (Spearman's rho = 0.67 and 0.65, respectively; both *p*'s < .01). Most participants did not adhere to one specific theory, so we could not relate our outcome measures to theory preferences.

Content analysis

Aver aged over participants, coders assigned 11,3 idea units (*SD* = 5,6), which is 30,0 % of all idea units, to category D (treatment planning). Further, on average 8,5 idea units (*SD* = 5,8) were assigned to category A (descriptive information), 3,5 (*SD* = 3,1) to category B (diagnostic information), and 8,5 (*SD* = 4,7) to category C (inferred information). Within category C, on average 4,9 idea units (*SD* = 2,8) were assigned to explanatory, causal elements: psychological, biological, and social mechanisms, predisposing and precipitating factors, strengths, and potential therapy-interfering factors. Thus, on average, of all units assigned to either categories A, B, or C, 25,0 % was assigned to causal elements (*SD* = 13,0), 57,6 % to descriptive elements (*SD* = 15,0) (category A elements, plus these elements from category C: inferred symptoms and evaluations of level of functioning), and 17,3 % to category B[3] (*SD* = 14,8).

Table 1 shows which (sub) elements in the four main categories of the CFCCM coders identified in participants' verbalizations, and with how many participants. Elements of the CFCCM to which coders assigned no idea units are not presented.

This table shows that all participants considered descriptive information (category A); most of the time they mentioned problem and symptom information from the case (*n* = 18). For 18 participants at least one idea unit was assigned to category B (classifications); these participants all referred to sub-element "ADHD" (Attention Deficit/Hyperactivity Disorder). We should note here, that consideration of a classification is sufficient to get a code. Thus, if participants mention for example ADHD, while saying that they think this classification should not be given, the code for having considered this classification is still assigned. For category C, idea units of 11 respectively 13 participants were assigned to

[3] We calculated the proportion of idea units assigned to explanatory, causal elements of the total number of idea units assigned to categories A, B, and C, to make this comparable to the outcome measures for causal linking of case states in the causal structure analyses. We calculated the proportion of idea units assigned to treatment planning elements (category D) of the total number of idea units, assigned to any category, to make this comparable with the proportion of follow-up statements identified in the protocols, in the causal structure analysis. Note that in the content analyses units of analysis (idea units) are larger than in the structure analysis (simple sentences); that in the content analyses repetitions are not excluded from analyses; and that in the content analyses open questions are included.

elements we consider to be descriptive: evaluations of level of functioning, and inferred symptoms or problems. For the causal elements, we see that one participant considered predisposing factors, and two participants considered precipitating factors. Most participants (n = 16) mentioned one or more psychological mechanisms, but as can be inferred from the sub-elements, participants mentioned different causes within this element. Even within the distinction between cognitive and affective sub-elements participants mentioned different specific causes, as is indicated between brackets. Six participants mentioned biological mechanisms, 11 participants mentioned social mechanisms. Within category D most participants (n = 17) indicated that they wanted to perform further assessment. However, they differed in specifically what they wanted to assess. Half of the participants proposed parent support (n = 10) as specific treatment type. Participants agreed less about other specific treatment types, specific techniques, and foci of attention.

Causal structure and causal content

We correlated the number of causal elements identified for a participant in the content analysis (the sum of the number of idea units coded as psychological, biological, and social mechanisms, predisposing and precipitating factors, strengths, and potential therapy-interfering events) with causal measures of the causal structure analysis, that is: the number of causal links, percentage linked case states, and length of the longest causal path. This resulted in Spearman's rho of 0.71, 0.61, and 0.76 respectively (all p's < .01).

Conclusions and discussion

We analyzed the structure and the contents of clinicians' mental representations of client information, to see whether clinicians represented information in causal models. We also investigated whether the follow-up steps they proposed were related to these causal models. Results of the structure analyses show that on average one third of the case states verbalized were causally and temporally connected, mostly causally. For most participants the causal structures we reconstructed in conceptual graphs were simple; the longest path of connected states often did not exceed two states. Also, conceptual graphs often consisted of two or more disjoint parts. Moreover, we found that most follow-up steps that clinicians proposed were not related to causally and temporally structured representations. Almost half of all follow-up steps were not even linked to any case state. Analysis of the contents of participants' verbalizations, using the CFCCM, gave comparable results: one fourth of the idea units (excluding units assigned to the category of treatment planning) were identified as having causal content, while most verbalizations contained descriptive information. While there was large variation between participants, it seems that most participants did not construct complex, coherent causal representations on which they based their follow-up steps. Furthermore, we found that participants mentioned different causal elements and gave different suggestions for further assessment and treatment.

If we compare our results to those of Pennington and Hastie (1986), we see that in the jurors' protocols 91 % of all events verbalized was causally and temporally linked, mostly causally. Also, jurors' story representations were related to their verdicts. We had expected to find more elaborate causal structures in clinicians' mental representations of client information as well as more connections between follow-up states and case states, since

Table 1

Elements and Sub-elements identified in Content Analysis

Category, Element	***n***	**Sub-element**	***n***
A Descriptive information	**20**		
a. Symptom and problem identification	18		
b. Development (including mental & physical health)	6		
c. Social relations (family / school / peers)	6		
d. Appearance	3		
e. Need more descriptive information	12		
B Diagnostic information	18		
1. Axis 1 DSM-IV-TR diagnosis	18	a. Attention Deficit/Hyperactivity Disorder	18
		b. Pervasive Development Disorders	8
		c. Oppositional Defiant Disorder	2
		d. Reading Disorder (Dyslexia)	1
		e. Not specified	3
2. Axis 2 DSM-IV-TR diagnosis	1	a. Not specified	1
3. Axis unspecified	2		2
C Inferred information	19		
1. Evaluation level of functioning	11		
2. Inferred symptoms and problems	13		
3. Predisposing factors	2	a. Pre- and perinatal complications	1
		b. Injury	1
		c. Not specified	1
4. Precipitating factors	1	a. Moving house	1
5. Psychological mechanisms	16	a. Cognitive (attention, learning, self-image)	7
		b. Affective (impulse regulation, anger)	9
		c. Behavioural (hyperactivity)	2
		d. Interpersonal	3
		e. Comorbid disorder (NLD)	1
		f. Not specified	3
6. Biological mechanisms	6	a. Genetic predisposition	2
		b. Neurological factors	3
		c. Physical factors (sensory, metabolism)	2
7. Social mechanisms	11	a. Parental factors (parenting style, expectations)	5
		b. Family interactions	4
		c. School factors (teachers, expectations, peers)	3
		d. Social factors not specified	4
8. Strengths	9	a. Strengths/adaptive skills	6
		b. Motivation	1
		c. Psychosocial support	1
		d. Not specified	1
9. Potential interfering factors	2	a. Attitude child	1
		b. Attitude father	1

Table 1

Continued

Category, Element	n	Sub-element	n
D Treatment planning	20		
1. Assessment/evaluation	17	a. Development	1
		b. Child: cognitive factors	3
		c. Child: affective factors	1
		d. Child: behaviour/interpersonal factors	3
		e. Child: neurological factors	1
		f. Child: physical factors	3
		g. Diagnosis	6
		h. Parenting style	4
		i. Family interactions	3
		j. School interactions	2
		k. Evaluation	1
		l. Referral	5
		m. Not specified	6
2. Type of treatment	13	a. Parent support	10
		b. Family therapy	3
		c. Remedial teaching	2
		d. Training (social skills, self-efficacy)	1
		e. Cognitive therapy	3
		f. Referral	1
3. Specific techniques	15	a. Child-focused (cognitive, affective)	3
		b. Parenting intervention	6
		c. School intervention	5
		d. Psycho-education	5
4. Treatment expectations	5	a. Improvement child (cognitive, affective)	3
		b. Improvement parenting style	1
		c. Improvement family interactions	1
5. Focus on signs and symptoms	4	a. Cognitive factors	3
		b. Affective factors	1
6. Focus on psychological mechanisms	5	a. Cognitive	2
		b. Affective	1
		c. Interpersonal	1
		d. Not specified	1
7. Focus on social factors	9	a. Parenting style	4
		b. Family interactions	6
		c. School interactions	5
		d. Special education	1
		e. Not specified	2
8. Medication			7
9. Focus on strengths			3

Note. Elements from the CFCCM to which coders assigned at least one idea unit are included in this table, and the sub-elements identified within these elements. For both elements and sub-elements the number of participants for which a (sub) element was coded is indicated. For mechanisms, causes are indicated, not effects.

identification of causal variables seems important for choosing effective interventions (Eells et al., 1998; Haynes, 1992; Haynes & Williams, 2003; Kuyken et al., 2005; Nurcombe, 2000; Sloman, 2005; Sloman & Hagmayer, 2006).

Our results resemble those of Eells et al. (1998), who found that written clinical case formulations from intake interviews were largely descriptive, and simple rather than complex. In these formulations, symptoms, problems, precipitating or predisposing life events were hardly integrated in explanatory mechanisms. Kendjelic and Eells (2007) obtained similar results for a group of clinicians who received no training in case formulation. In psychotherapists' think aloud case formulations Eells et al. (2005) found that the largest proportion of idea units was coded as inferred information (category C): 37 %. However, this percentage includes idea units about inferred symptoms and problems and evaluations of level of functioning, elements that we did not consider to be causal. Kuyken et al. (2005) assessed the reliability and quality of case formulations that clinicians had constructed using a systematic method for cognitive case formulation. Although agreement on descriptive aspects was good, agreement was lower for inferential aspects. Further, only 44 % of the formulations was judged as at least "good enough", in terms of the parsimony, coherence, and meaningfulness needed to base effective interventions on them.

Why are the causal representations of client information we extracted simple, diverse, and poorly related to treatment planning? We discuss four explanations.

First, the differences in our findings compared to those of Pennington and Hastie (1986) may be explained by the different task demands for jurors and clinicians. American jurors have to decide about one specific event: the crime, while clinicians have to decide how to treat problems which usually span a whole period in life, or a whole life. Jurors appeared to construct stories to integrate trial information. Stories, although they may be hierarchic (have embedded episodes) (Pennington & Hastie, 1986; Rumelhart, 1975), are told in a linear fashion: one event after another unfolding in time. Clients' problematic behaviours are not a single event, and are not the consequence of a single cause. Behaviour is caused by many interacting influences: biological, cognitive, and environmental, which are mostly probabilistic (Haynes, 1992; Kiesler, 1999; Morton, 2004). As a result of the interactions, behaviours can take different forms in subsequent phases in life (Morton, 2004). Explanations of behaviours of clients should be multidimensional, taking into account all influences, their interactions, and changes in time. This multidimensionality of (explanations for) behaviour complicates decisions for interventions (De Los Reyes & Kazdin, 2006). Gaps in a linear story will be more easily discerned and filled in than gaps in more complex, multidimensional models (Keil, 2003). Thus, it might be more difficult to construct a comprehensive causal representation for a clinician deciding upon a client's behaviours, than it is for jurors deciding about a crime.

Second, the lack of causal structure in clinicians' mental representations may be explained by the fact that comprehensive scientific knowledge about interacting causes is still lacking for mental disorders and psychopathological behaviour (Kiesler, 1999; Cicchetti & Sroufe, 2000), as is evidence for why interventions are effective (Bieling & Kuyken, 2003; De Los Reyes & Kazdin, 2006). Insufficient or inconsistent knowledge of causes and effects of interventions will impede the construction of complete, coherent causal models to which

follow-up steps can be connected. Perhaps clinicians' knowledge base resembles physics students' knowledge base of physics: loose fragments which are hardly connected in unifying causal theories (Di Sessa, Gillespie, & Esterly, 2004; Keil, 2006). This might explain why in our study conceptual graphs often consisted of two or more disjoint parts, why causal paths were usually short, and why many follow-up steps remained unlinked to case states. Also, a lack of consensual knowledge on aetiology and effective treatments can explain the large variation in causal contents and treatment planning.

A third explanation for our results is that clinicians do not only engage in causal reasoning to process client information and to decide upon treatment, but also in schematic processing, such as classification of problem behaviour in DSM-IV categories. If information is ordered in schemas which clinicians activate while processing information, clinicians might mention elements of schemas in their protocols, but we will not find much causal and temporal linking. Leon and Perez (2001) found that clinicians activate "clinical diagnosis inferences" while reading clinical texts. This activation is a process of categorization of signs, symptoms, and behaviours. A similar process may partly account for our findings: if we consider the lists of unlinked states in Figures 1 and 2, it seems that some of these states can be interpreted as symptoms belonging to categories, or elements of schemas that participants may have of ADHD or NLD (Nonverbal Learning Disorder).

Interestingly, participants sometimes suggest categories as cause, as when 'ADHD' is thought to cause impulsive behaviour (see Figure 1). This seems to be in line with the claim of Leon and Perez (2001) that the clinical diagnosis inferences clinicians activate are explanations: according to Leon and Perez clinicians infer underlying psychopathological disorders from symptoms and behaviours, which are thought to cause the specific behaviours. In the DSM-IV-TR (American Psychiatric Association, 2000) ADHD is defined as a pattern of persistent inattentive, hyperactive, and impulsive behaviours. Thus interpreted, saying that ADHD causes impulsiveness is circular. Kim and Ahn (2002) found that despite the atheoretical design of the DSM-IV, clinicians' representations of DSM-IV disorders do have causal structure, and that this causal structure influences clinicians' classifications.

The content analysis shows that participants mentioned classifications, and it seems that participants agree more about classification than about causal elements. Perhaps higher agreement about classification follows from the consensual nature of the classification systems (e.g. DSM-IV; ICD-10).

Clinicians might be engaged in both causal reasoning and schematic processing. However, if schemas are activated, and clinicians mention elements of these schemas when verbalizing their thoughts, one would expect follow-up steps to be related to these elements. But we found that almost half of all follow-up steps was not related, either directly or via other follow-up steps, to any case state. Also, despite agreement on classification, participants did not agree about treatment planning. These findings are in line with those of Witteman and Koele (1999), who found that clinicians' client interpretations could not predict their treatment decisions. Perhaps the follow-up steps that clinicians proposed are based on routine or institutional standards, rather than (only) on causal or schematic mental representations of client information.

Our fourth explanation concerns our methodology. It might be argued that our participants did not succeed in arriving at elaborate, coherent causal representations due to the case we used, which might have been too difficult, or because we gave insufficient information. However, most clinicians were familiar with similar cases, and thought the case was representative. We deliberately provided a lot of information, more than clinicians will generally have before deciding on treatment.

Another methodological explanation is, that possibly clinicians were constructing causal models or stories, and did base their follow-up steps upon these, but that we have not uncovered them because links are not always verbalized explicitly, i.e. with connective words or verbs. Similarly, causal elements might be considered though not always verbalized explicitly. Our structure coding was very strict, copying from Pennington and Hastie (1986) who demanded explicitness before considering a relation to be causal. They judged that in "Johnson was a violent man. That makes me think he intended to kill Caldwell" the two events are not causally or temporally connected; while in "Johnson was angry so he decided to kill him" they are (Pennington and Hastie, 1986, p. 249). However, simply reading the protocols yielded no clear indication of implicit causal or temporal structuring of information, such as chronological ordering. We further note that our results seem comparable to those of Eells et al. (1998) and Kuyken et al. (2005), who analyzed written case formulations. Also, the correlations between results of the analyses of causal structure and causal content add to the reliability of both methods for analyzing causality in diagnosticians' reasoning. Still, when clinicians become more experienced they might rely more on implicit, automatic routines, which are not accessible for verbalization (cf. Einhorn, Kleinmuntz, & Kleinmuntz, 1979; Ericsson & Simon, 1993). Although concurrent verbalization of thoughts may accurately reflect underlying cognitive processes (as opposed to retrospective verbal reports on mental processes, see Nisbett & Wilson, 1977), it is impossible to report all thoughts. Validation of results of the think aloud method with other methods is recommended (Caspar, 1997; Einhorn et al., 1979; Ericsson, 2006; Magliano, 1999).

Limitations of this study are inherent to the labour-intensive think aloud method. We used only one case, and our sample is small. Because of the small sample size statistical calculations incorporating participants' characteristics are not powerful. Therefore, we can neither support nor exclude explanations for the large variation we found in our structure and content analyses. Clinicians' experience may however be expected to influence causal reasoning, in the sense that more expert diagnosticians might have more complex, elaborate, and comprehensive causal representations of clients, to which intervention choices are more evidently linked (Boshuizen, 2004; Eells & Lombart, 2003; Eells et al., 2005; Kuyken et al., 2005). Theoretical adherence might also be expected to influence causal reasoning; however Kuyken et al. (2005) found that cognitive therapists who received an additional training in a standardized cognitive case formulation method still disagreed about explanatory inferences.

Future research

In future studies, the role of causal reasoning in diagnosticians' information processing and decision making needs to be further investigated. Task demands might be varied to learn how these influence clinicians' reasoning processes. Also sample characteristics such as amount of experience might be varied. Other qualitative methods might be used, such as explicit probing for causal explanations as in cognitive mapping techniques (cf. Hodgkinson, Maule, & Bown, 2004; Kim & Ahn, 2002; see Chapter 4 of this thesis). In addition, more quantitative and experimental methods might be used to refine models of clinicians' cognitive processes and validate models' predictions (Einhorn et al., 1979; cf. Gaines, Brown, & Doyle, 1996; Magliano, 1999). Further, research should establish what clinicians base their interventions on, because so far this is unclear (cf. Eells et al., 1998; Kuyken et al., 2005; Witteman & Koele, 1999). Also, it remains to be established whether high quality causal client representations or case formulations lead to better treatment decisions, in terms of effectiveness (Bieling & Kuyken, 2003; Eells, 2007). Finally, the disagreement about causal, explanatory elements that diagnosticians mentioned here, or described in case formulations (Eells et al., 1998; Kenjelic & Eells, 2007; Kuyken et al., 2005, see also Garb, 1998) needs further attention. Specifically, it seems important to learn what the consequences are of this disagreement and how it might be reduced.

Implications

Because of the limitations of this study mentioned above, the implications we suggest should be treated with caution. Given that clinicians' tasks are indeed inherently complex, and that there is a lack of fundamental causal knowledge, decision support and educational methods might be directed at aiding in the construction of causal models, testing causal hypotheses, and at making gaps in knowledge visible. It is not self-evident that clinicians are aware of gaps. Rozenblit and Keil (2002) found that people overestimate their ability to explain complex causal systems. Clinicians in our study were confident about their assessments, and more so when they were more experienced. A relation between clinicians' (over)confidence and experience has been reported more often (e.g. Oskamp, 1965; Lichtenberg, 1997). Since we hardly found comprehensive causal representations, a poor relation of follow-up steps to case states, and large variation in causal contents and treatment planning, this confidence is perhaps out of place.

Morton (2004) has developed a graphic tool which can help represent complex ideas in a comprehensible form, called "causal modelling". Using this tool facilitates both understanding and comparison of different theories for disorders (Krol, Morton, & De Bruyn, 2004) as well as one's causal understanding of individual clients. Also, use of "Functional Analytic Clinical Case Models" (Haynes & Williams, 2003), which depict and quantify complex causal relations, can aid in identifying important causal relations and subsequently selecting the most effective treatments for individual clients. We advocate the use of such tools. Kendjelic and Eells (2007) found that training in case formulation improves their quality, and increases the number of explanatory inferences in them. Training in case formulation skill might also be enhanced by the use of computers (Caspar et al., 2004).

Also, it seems important to stimulate students and clinicians to read and use recent scientific causal insights. In this light, findings of Woods, Brooks, and Norman (2005) are

interesting. They found that, after a time lag, students diagnose more accurately and recall more medical information if this information was learned by explaining the causal mechanisms than when information was learned by providing probabilities. Students seem to remember causal information better, or information is more easily reconstructed when it is understood causally.

Acknowledgments

We thank Eric and his parents, and Rowella Kuijpers, for sharing their personal information with us. We thank Karine Groen, Chantal Berens, and Emmy Stamsnieder for conscientiously coding the protocols. We thank Tracy D. Eells for providing us with the CFCCM, the case formulation content coding manual, which we found very useful. We thank all our participants. And we thank all those researchers and clinicians who commented on earlier drafts and presentations, among who are Laurence Claes, Patricia Bijttebier, Nancy Kim, Rowella Kuijpers, and anonymous reviewers. Their comments contributed greatly to this manuscript.

Chapter 3

Looking for Causes?
Causal Reasoning in Mental Health Clinicians' Information Processing

Abstract

Twenty clinicians watched a video of an intake-interview, while thinking aloud about these questions: "What do you think is the matter with this client; what would you propose as follow-up steps?" Based on the assumption that clinicians' reasoning process is explanation-based, we expected that clinicians would make mostly causal inferences while processing the intake information. We found, however, that most inferences were descriptive. Further, we found that, as early as at the intake, different clinicians considered different explanations. We suggest that the role of causal reasoning in clinicians' information processing may be limited, because a causal understanding of clients may not always be necessary to make treatment decisions, and because diagnosticians may lack complete causal knowledge.

Clinicians collect information about their clients through interviews, tests, and observations, in order to identify their clients' problems and to decide on effective interventions for these problems. This information is integrated in case formulations, which are supposed to give a comprehensive description of the client, and causally explain their behaviours and symptoms, referring to past events and to interacting physical, social, and psychological factors. Further, these case formulations allow predictions or prognoses as to which interventions will be most effective (e.g. see handbooks: Carr, 2006; De Bruyn, Ruijssenaars, Pamijer, & Van Aarle, 2003; Eells, 2007; Wenar & Kerig, 2006). Causal reasoning thus seems essential to clinicians' information processing and decision making (see also Garb, 2005).

In complex, dynamic situations decision makers reason causally to arrive at their decisions. They integrate ill-structured, incomplete, and implication-rich information in causally structured mental representations. From these representations they then derive their decisions (Goldstein & Weber, 1995; Hastie & Pennington, 2000; Pliske & Klein, 2003; Sloman & Hagmayer, 2006). To arrive at coherent representations decision makers make inferences, using their background knowledge. Many inferences are made during the decision process, before actually reaching a final conclusion, as is shown in research on jurors' decision making (Pennington & Hastie, 1988; 1994).

Similarly, in text comprehension people make a lot of inferences, using their knowledge, to construct coherent causal situation representations (Graesser, Olde, & Klettke, 2002; Graesser, Singer, & Trabasso, 1994; Langston & Trabasso, 1999; Van den Broek & Gustafson, 1999). Think aloud studies showed that people make explanatory inferences during reading rather than associative inferences which elaborate on descriptions, or predictive inferences which predict consequences (Trabasso & Magliano, 1996; Zwaan & Brown, 1996). Similar findings have been obtained with time-based behavioural measures, such as sentence reading times and lexical decision times (Graesser et al., 1994; Magliano, 1999).

Comprehension is part of decision making, and integration of information in coherent causal representations is central to both decision making and text comprehension (Rettinger & Hastie, 2003; cf. Whitten & Graesser, 2003). This also applies to clinical reasoning (Arocha & Patel, 1995; Leon & Perez, 2001). We may therefore expect that clinicians make causal inferences while they process client information to arrive at coherent causal representations, using their world knowledge and specific clinical knowledge. We also expect them to make these inferences early in the process, in line with Sandifer, Hordern, and Green (1970) who found that psychiatrists generated almost half of their diagnostic hypotheses after watching the first three minutes of diagnostic interview videos (see also Kendell, 1973, Nurcombe & Fitzhenry-Coor, 1982).

However, this expectation is not supported by research on diagnosticians' case formulations. Eells, Kendjelic, and Lucas (1998) found that the case formulations clinicians made from intake interviews were predominantly descriptive. Less than half of the case formulations contained inferred causal mechanisms which explained clients' problems, and only a few of these contained mechanisms which were rated adequate or strong. Similarly, Kuyken, Fothergill, Musa, and Chadwick (2005) judged that cognitive therapists' case formulations were often not sufficiently coherently structured and did not provide a

meaningful enough account of the client's problems, to allow the derivation of effective interventions. Mental client representations as extracted from clinicians' think aloud protocols also appeared to lack causal complexity, with follow-up steps proposed poorly related to them (De Kwaadsteniet, Krol, & Witteman, 2008, see Chapter 2 of this thesis)

However, in the studies of Eells et al. (1998), Kuyken et al. (2005), and De Kwaadsteniet et al. (2008) the analyses concerned clinicians' remarks, either written or verbalised, made *after* all of the information had been presented. Such post hoc reports may not be representative of reasoning processes as these occur at the time of actually processing the information (Ericsson, 2006; Ericsson & Simon, 1993). This was shown, for example, by Eva, Brooks, and Norman (2002). They compared medical students' and residents' think aloud protocols obtained on-line with think aloud protocols obtained after all the case information had been presented. In the post hoc protocols fewer propositions were verbalised, fewer hypotheses were generated, hypotheses were generated later, and fewer causal links were verbalised than in the on-line protocols. Causal explanations that clinicians generate while processing information may not be elaborated on afterwards, or only briefly mentioned (cf. Nisbett & Wilson, 1977), perhaps because these appear obvious, or because clinicians feel too uncertain about them.

In the present study, we analysed clinicians' think aloud protocols that were obtained on-line. Clinicians verbalised their thoughts while they were watching a video of an intake interview. Assuming that diagnostic reasoning indeed is an explanation-based process similar to decision making in other complex domains and in text comprehension, we hypothesised that we would find more causal than associative or predictive inferences, and that questions and inquiry plans verbalised by the clinicians would mostly ask for causal rather than descriptive information.

In addition we investigated the specific contents of the inferences, questions, and inquiry plans, in order to see whether different clinicians offer the same explanations and inquiry plans. Reliability of clinicians' causal inferences is notoriously poor (Dawes, 1994; Garb, 1998, 2005). Kuyken et al. (2005) and Persons and Bertagnolli (1999) found that cognitive therapists could reliably formulate descriptive aspects, but that the inferential aspects of the formulations were less reliable. Also, variability in causal statements that clinicians verbalise in post hoc think aloud protocols was large (De Kwaadsteniet et al., 2008). By analysing the contents of clinicians' inferences early in the diagnostic process, we want to find out whether disagreement already arises early.

Analysis of clinicians' on-line information processing should thus inform us about clinicians' causal reasoning processes, while they are constructing their mental representations. This information is useful to improve diagnostic reasoning, for example through the development of decision aids and training methods (Caspar, Berger, & Hautle, 2004; Garb, 2005; Witteman, Harries, Bekker, & Van Aarle, 2007).

Method

Participants

Twenty clinicians with different levels of experience participated. Four participants were master students at the end of their final year. These students, all women, had had nine

months of clinical experience with children of primary school age during their internships. The sixteen experienced participants, twelve women and four men, worked in fourteen different institutions, as a clinical (child) psychologist or as a remedial educationalist. Their mean experience was 13.3 years (*SD* = 9.8).

Materials

We presented participants with materials of a diagnostic investigation of a nine year old boy, for which we had received his parents' written consent. These materials included a video of 33 minutes of the intake interview, the results of nine tests, and a description of the behaviour observed during these tests. The boy and his parents had been referred to a Dutch youth care institution because of the boy's increasingly problematic behaviour at home and in school.

In the intake interview, first the parents and the boy talked with two clinicians. After ten minutes the boy left with one of the clinicians to go to a play room. The parents continued their conversation with the other clinician. The video was shown on a laptop, and participants wore headphones. Participants could interrupt the video by pressing the space bar. When they interrupted the video the pauses were registered.

Procedure

Except for the student participants and two experienced participants who came to the experimenter's office, participants were visited by the experimenter at the institution where they worked. After two unrelated exercises to get used to verbalising thoughts aloud, the experimenter presented clinicians with the intake video. A written instruction asked the participants to watch the intake as if they were present, or as if they were going to be asked to take over this client from a colleague. The participants were asked to think aloud, while watching the video, focusing on the following questions: "What do you think is the matter with this boy; and what would you propose as follow-up steps?" The questions were copied on a sheet of paper, which was on the table in front of the participants, as a memory aid.

Participants could at any time interrupt the video in order to express their thoughts. They could also verbalise their thoughts without interrupting the video. If participants didn't speak for several minutes, the experimenter reminded them to think aloud. Otherwise, the experimenter remained silent.

At the end of the video, the experimenter prompted participants to continue to think aloud about the questions. When participants stopped speaking, she asked them to study the additional test information, all the while thinking aloud. When they again stopped speaking, she asked for a summary. Finally, participants filled in a short questionnaire asking for personal information (such as age and experience). They were thanked and received their rewards: a gift certificate of 50 euro for the clinicians, or 10 euro for the students. The experiment took one-and-a-half to two hours per participant.

Analysis

The analysis of the data focused on the verbalisations made by the participants while they watched the intake video and later, after the video stopped, before they were presented with the additional information. In the 33 minutes of the intake a lot of information was presented, which should be sufficient to analyse clinicians' reasoning while they form

mental representations. Two coders independently segmented and coded the typed-out verbalisations.

Segmentation. First, coders segmented all verbalisations into semantic units, using the transcription procedure of Berman and Slobin (1994) (cf. Trabasso & Magliano, 1996). Each unit should describe one situation (event, action, or state). If, for example, a participant said "he is very impulsive and often oppositional", coders divided this into the units "he is very impulsive" and "and often oppositional". Coders neglected participants' "narrator comments" which concerned their own thinking (e.g. "I think that", "It appears that") (following Berman & Slobin, 1994), or which concerned persons in the video (e.g. "Mother says", "The clinician asks").

Categories. Next, coders assigned each unit to one of seven categories: *information from the video*, *inferences*, *questions*, *inquiry plans*, *treatment plans*, *knowledge*, and a rest category. Information from the video was defined as repetitions of what the parents, the boy, or the clinicians said in the video, or consisting of observations of behaviours in the video. Repetitions of what had been said did not need to be given in the same words, as long as the gist remained the same. Repetitions of information or observations which were stated in general terms were also considered to be information from the video, as long as these general formulations did not add anything new.

Coders identified units as inferences when the verbalisations of the participants added new information concerning persons in the video, which at least at that point had not yet been presented. To establish whether information had been presented before or not, coders used a literally typed-out report of the intake video.

We defined questions as verbalisations asking for more information. Usually questions can be recognised by words such as *who*, *what*, *how*, etcetera (e.g. "What hobbies does he have?"). Verbalisations conveying inquiry plans were coded separately (e.g. "I would check his hearing"), as were verbalisations conveying treatment plans and their expected effects (e.g. "Medication might help", and "to help him concentrate"). Verbalisations in which participants explicated their theoretical or experiential knowledge were categorised separately (e.g. "I often see young children being spoiled"). Finally, in the rest category were coded verbalisations not about the persons in the video, but for example about the task (e.g. "Which key should I press"), or about participants' own thinking processes (e.g. "Let me think back") and incomprehensible verbalisations.

Coders coded all units identified in participants' verbalisations, excluding only experimenter's verbalisations and typist's comments (e.g. "laughs"). Here too, coders neglected narrator comments.

Subcategories. For inferences, questions and inquiry plans, coders distinguished four subcategories: *descriptive*, *explanatory*, *classifying*, or *predicting*. Inferences were coded as descriptive when they provided additional information or details about symptoms, problems, or persons, so the who, what, how, etcetera of events or situations (e.g. "(I think) there will be many conflicts between mother and son"). They were coded as explanatory when they provided causes or reasons for events or situations (e.g. "It might just be provocation"). These inferences can be imagined to answer why questions. Inferences were coded as classifying when they mentioned specific clinical categories for symptoms or

behaviours (e.g. from DSM-IV, as in "(It seems to me) ADHD"). Finally, inferences were coded as predicting when they conveyed what a participant thought would happen later in the video, or later in the boy's life (e.g. "He might drop out of school"). The level of certainty expressed (e.g. by words such as *perhaps*, *probably*, *probably not*) was irrelevant for assigning an inference to one of these subcategories.

Analogously, questions and inquiry plans were coded as descriptive when they asked for more descriptive information (e.g. "How is his motor-functioning?"), as explanatory when they asked for possible causes or reasons (e.g. "Why does he become so angry?"), as classifying when they asked for possible classifications (e.g. "Test the ADHD-hypothesis"), or as predicting when they asked how situations would evolve in time. For inquiry plans, we added the subcategory *open* for units in which inquiry plans expressed only the plan to gain additional information, without further specifying what kind of information (e.g. "I would do additional research").

Content coding. Coders analysed the specific contents of those inferences, questions, and inquiry plans that had been subcategorised as either explanatory or classifying. For this, we used the Case Formulation Content Coding Method (CFCCM) of Eells, Kendjelic, Lucas, and Lombart (received in personal communication, October 2007). This coding method has been developed to code the contents of written case formulations and of case formulations as verbalised in thinking aloud (for a description, see Eells, Lombart, Kendjelic, Turner, & Lucas, 2005).

We mapped the inferences, questions and inquiry plan that had been subcategorised as classifying onto the category 'diagnostic information' of the CFCCM, which codes verbalisations that consider a diagnosis in DSM-terms. We mapped the inferences, questions and inquiry plans that had been subcategorised as explanatory onto the category 'inferred information' of the CFCCM, specifically to those elements which concern explanatory inferences: predisposing and precipitating factors or events, psychological mechanisms, biological mechanisms, social mechanisms, and protective factors.

Based on participants' verbalisations we created a subdivision in the elements of the CFCCM, because sub-elements in the CFCCM do not always seem appropriate for children's cases (cf. Chapter 2 of this thesis). For this subdivision we used Carr's (2006) identification of factors influencing problem development (see Figure 2.1 in Carr, 2006) and his distinction of cognitive, affective, behavioural, physical, and interpersonal clinical features (see Table 11.2 in Carr, 2006). For explanatory units, coders distinguished what specific predisposing and precipitating factors participants mentioned, what specific causes participants mentioned in the psychological, biological, and social mechanisms, and what protective factors they mentioned. Inferences, questions, and inquiry plans might not mention any specific explanation (e.g. just asking "why", or mentioning "child factors"). Coders gave these units the code *not specified*. In the classifications, coders distinguished the different axis I DSM-IV-TR-classifications (APA, 2000) that participants mentioned.

Reliability. We calculated reliability on the basis of five randomly chosen protocols. Coders agreed about segmentation into units for 90.6 % of the words in the protocols. For assigning units to one of the seven categories, Cohen's kappa for agreement was .74. For the subcategories Cohen's kappa was .69. We calculated content coding reliability for

explanatory and classifying units in all protocols, since there were not so many of them. Cohen's kappa for agreement about content codes was .77. These kappa show substantial coding agreement (Landis & Koch, 1977). For disagreements, coders provided alternative codes. Usually, with these alternatives agreement was reached. Remaining disagreements were resolved in discussion.

Results

Categories and subcategories

Coders segmented the protocols into on average of 198.7 units (*SD* = 140.1). One-third was identified as information from the video, and almost a quarter as inferences. Of the inferences, almost two-third was categorised as descriptive. Of the questions an even higher percentage was descriptive, and the same was true of the inquiry plans (see Table 1 for the percentages).

Table 1

Mean Percentages (and Standard Deviations) of Protocol Units Coded into the Different Categories and into Sub-Categories

Category	**Mean %**	**(*SD*)**	**Sub-category**	**Mean %**	**(*SD*)**
Information from the video	34.6	(19.7)			
Inferences	22.3	(12.2)	descriptive	65.9	(18.6)
			explanatory	17.6	(13.5)
			classifying	14.5	(14.8)
			predicting	2.0	(3.0)
Questions	10.5	(8.3)	descriptive	75.8	(21.9)
			explanatory	21.8	(21.8)
			classifying	1.7	(4.8)
			predicting	0.7	(2.5)
Inquiry plans	6.3	(4.2)	descriptive	70.9	(25.3)
			explanatory	6.4	(9.3)
			classifying	3.1	(6.6)
			open	19.6	(18.5)
Treatment plans	2.3	(6.0)			
Knowledge	1.4	(1.9)			
Rest	23.6	(10.6)			

Paired sample t-tests showed that the proportion of descriptive inferences was significantly higher than the proportion of either explanatory, classifying, or predicting inferences ($t(19) = 7.4$; $t(19) = 7.6$; $t(19) = 14.8$, all p's <.01). Similarly, paired sample t-tests showed that the proportion of descriptive questions and inquiry plans was larger than the proportions of explanatory, classifying, predicting, or open questions and inquiry plans (t-values ranged between 5.6 and 15.0, all p's <.001).

Content coding

Table 2 shows to which elements of the CFCCM and to which sub-elements coders assigned the inferences, questions, and inquiry plans that had been coded as explanatory and classifying.

As can be seen, participants mentioned only a few predisposing and precipitating factors. Moving house was mentioned most often, by six participants. Seventeen participants mentioned psychological mechanisms, with cognitive, affective, and interpersonal causes. Within these causes, participants mentioned different specific factors, as is indicated in the table. Four participants mentioned a biological mechanism, with different causes. Seventeen participants mentioned social mechanisms. Here too, participants mentioned different causes, and within these causes again different specific factors. One participant mentioned a protective factor. Nine participants mentioned explanations which could not be assigned to any of the content elements (e.g. "It is something in the boy"; "It is a combination of child and family factors").

Participants often repeated causal factors in their explanatory inferences, questions, and inquiry plans. On average, participants mentioned 5.3 different causal factors (ranging from none to 19).

Seventeen participants mentioned a classification and all these participants mentioned ADHD. Eight participants mentioned at least one alternative axis I classification. Two participants mentioned that the problems might be classified according to DSM-criteria, without further specification. Participants also often repeated the same classification in their inferences, questions, and inquiry plans. On average, participants mentioned 1.7 different classifications (ranging from none to six).

Conclusions and Discussion

We expected the diagnostic decision making process to be explanation-based, similar to decision making in other complex, dynamic domains (Hastie & Pennington, 2000), and similar to text comprehension (Graesser et al., 2002). Thus, we had expected that while processing the video information clinicians would make many causal inferences, and seek mostly causal explanations. These expectations were not borne out. We found that clinicians' thoughts while they watched the intake video were mostly repetitions of information from the video. Participants did generate quite some inferences, but these were mostly additions of descriptive information. It may not be surprising that participants generated fewer questions, inquiry plans, or treatment plans, since they were expecting to receive more information after the intake video. However, most of the information they asked for was again descriptive, not explanatory.

These findings are contradictory to findings in text comprehension research, where most of the inferences that readers make seem to be causal rather than descriptive (Trabasso & Magliano, 1996; Zwaan & Brown, 1996). These findings do, however, corroborate earlier results that showed no evidence for elaborate causal reasoning by clinicians who were thinking aloud *after* the presentation of information (De Kwaadsteniet et al., 2008). Also, written case formulations often appear to contain insufficient causal mechanisms to form a basis for treatment plans (Eells et al., 1998; Kuyken et al., 2005).

Table 2

Number of Participants who Considered Specific Causal Factors and Classifications as Identified in the Content Analysis

Element	***n***	**Sub-element**	***n***	**Further specification**	***n***
Explanations					
1. Predisposing factors	2	a. Pre- and perinatal complications	1		
		b. Age	1		
2. Precipitating factors	10	a. Moving house	6		
		b. Events in school	2		
		c. Events at home	3		
		d. Not specified	4		
3. Psychological mechanisms	17	a. Cognitive factors	13	General intelligence	5
				Integration / Learning problems	6
				Self-image	4
				Conscience	2
				Other	3
		b. Affective factors	13	Impulse control / Regulation	10
				Anxiety	4
				Mood	3
				Other	5
		c. Interpersonal factors	9	Provocative behaviour	5
				Interactions	4
				Attitude	2
				Other	2
		d. Not specified	1		
4. Biological mechanisms	4	a. Neurological factors	3		
		b. Physical factors	2		
		c. Genetic predisposition	1		
		d. Not specified	2		
5. Social mechanisms	17	a. Parental factors	12	Parenting style	12
				Other	3
		b. Family interactions	1		
		c. School factors	4	Teaching	3
				Peers	2
		d. Unstructured situations	4		
		e. Social factors not specified	5		
6. Strengths	1	a. Psychosocial support parents	1		
7. No specific explanation	9				

Table 2
Continued

Classifications			
1. Axis 1 DSM-IV-TR diagnosis	17	a. Attention Deficit/Hyperactivity Disorder	17
		b. Pervasive Development Disorder	5
		c. Oppositional Defiant Disorder / Conduct Disorder	6
		d. Learning Disorder / Reading Disorder (Dyslexia)	3
		e. Depression	1
		f. Attachment Disorder	1
		g. Not specified	5
2. Axis unspecified	2		

Note. The table presents the elements from the CFCCM, the sub-elements we distinguished as well as further specifications within these sub-elements that participants mentioned in their classifying and explanatory inferences, questions and inquiry plans. Coders identified causes in the mechanisms, not effects.

First, we will discuss methodological aspects that might have influenced our findings. Then we will discuss why our participants, and clinicians more in general, might not usually be engaged in elaborate causal reasoning while processing client information.

When collecting our think aloud protocols we adhered closely to the procedure described by Ericsson and Simon (1993). They claim that, although participants will not verbalise all their thoughts while processing information, thinking aloud should yield a reliable subset of their thoughts (Einhorn, Kleinmuntz, & Kleinmuntz, 1979; Ericsson, 2006; Ericsson & Simon, 1993). We gave our participants the opportunity to interrupt the video at any moment to speak and to speak without interruption, so they could verbalise their thoughts as they came up. We did so specifically to prevent participants from needing to retrieve their thinking process from memory, since this might lead to biased responses (cf. Nisbett & Wilson, 1977).

We do not expect that differences in coding are responsible for the low number of causal inferences we identified. We analysed the inferences clinicians made in analogy to think aloud studies in text comprehension (Trabasso & Magliano, 1996; Zwaan & Brown, 1996) and juror decision making (Pennington & Hastie, 1994). Specifically, our distinction between descriptive, explanatory, and predicting inferences matches the distinction between associations, explanations, and predictions of Trabasso and Magliano (1996).

The think aloud instruction we used may not have explicitly prompted clinicians to reason causally. However, we believed that a more specific instruction might induce other thinking processes than clinicians normally have. We assumed that causal reasoning would occur spontaneously, as this seems required to think of effective interventions (Beach, 1992; Sloman, 2005; Sloman & Hagmayer, 2006). In contrast, in other on-line think aloud studies in text comprehension the instructions explicitly asked participants to understand each individual sentence, and to inform the experimenter about their understanding (Trabasso & Magliano, 1996; Zwaan & Brown, 1996). The stories that those participants read had been

constructed as coherent causal networks, in which the story structure explains the presence of each sentence. In on-line think aloud studies on diagnostic reasoning participants have been instructed to verbalise their (preliminary) diagnostic conclusions at set points (see Eva et al., 2002; Sandifer et al., 1970). We think that as a consequence of such directing instructions, participants might consider more explanations and classifications than they would do without such directions, as in our study (cf. Ericsson & Simon, 1993).

The issue thus is whether clinicians normally engage in elaborate causal reasoning in order to decide what is the matter with a client and what follow-up steps need to be undertaken. Here, we focus on two important considerations: the need to engage in elaborate causal reasoning, and the availability and accessibility of causal knowledge needed to do so. In text comprehension, readers only generate many causal inferences if they have an explicit goal to understand the text, and if relevant information is available and easily accessible to working memory (Frank, Koppen, Noordman, & Vonk, 2003; McKoon & Ratcliff, 1992; Trabasso & Magliano, 1996). For jurors, an exact reconstruction of past events is necessary to be able to decide upon a verdict. Jurors need to know the precipitating factors, and the motives of a suspect, before deciding on their guilt. Clinicians might not consider the predisposing, the precipitating, and maintaining factors in detail before deciding upon a treatment. Groenier, Pieters, Hulshof, Wilhelm, and Witteman (2008) found that clinicians judged causal analysis of clients' problems least necessary, compared to describing clients' problems and choosing treatments. Clinicians might indicate treatments without having a clear understanding of the mechanisms causing clients' problems, but they might for example base treatments on a classification.

Although classifications are no explanations (APA, 2000; Morton, 2004), it looks as if they sometimes are being used as such. For example, Leon and Perez (2001) call categorisations that clinicians activate, such as kleptomania, explanatory inferences, "because it is seen as the manifestation of a deep psychological disorder that causes the behavior" (Leon & Perez, 2001, p. 189) (see also Chapter 2). Classifications may provide coherence to a mental representation (Arocha & Patel, 1995; Leon & Perez, 2001; Van den Broek & Gustafson, 1999).

Apart from not having the explicit goal to causally understand their clients, clinicians may lack relevant knowledge. In the clinical domain, causal explanations of clients' problems are complex: psychological and or behavioural problems develop in a continuous interaction between many biological, psychological, behavioural, and environmental factors (Kiesler, 1999; Morton, 2004; Nurcombe, 2000). Detailed knowledge about what all these factors are and how they interact is in development (Bieling & Kuyken, 2003; Clarkin & Levy, 2004; Kiesler, 1999). Uncertainty about specific causes for clients' problems might prevent clinicians from drawing causal inferences. Leon and Perez (2001) showed that with more experience, and thus more clinical knowledge, clinicians make categorising inferences sooner. From this, we reason that if relevant causal knowledge is made available, and easily accessible (e.g. by training or experience), clinicians might make more causal inferences.

Finally, we found that different clinicians considered different causal factors. This might also be a consequence of incomplete common causal knowledge (Bieling & Kuyken, 2003). Further, clinicians often do not get feedback about the explanations they propose,

about the actual effectiveness of the interventions they propose, and about what specifically makes interventions effective (Garb, 2005; Lichtenberg, 1997). This might increase the idiosyncrasy of clinicians' causal theories. Differences in education, in institutions and in clients may further increase the idiosyncrasy of clinicians' causal beliefs. This can be an explanation of our current finding that different clinicians' causal judgements begin to deviate as early in the process as during the intake interview. And if clinicians disagree about causes already early in the process, poor reliability of clinicians' final causal conclusions is to be expected (cf Chapter 2; Garb, 1998, 2005; Kuyken et al., 2005; Persons & Bertagnolli, 1999).

Implications and future research

Our sample was small, we used one case, and our experimental situation is different from a real diagnostic situation. Therefore our findings should be interpreted with caution, and need replication with more participants and more cases. It would be interesting to test whether clinicians make implicit causal inferences, using implicit measures such as priming tasks (cf. Magliano, 1999; Leon & Perez, 2001). Further, from our suggestion that causal knowledge is insufficiently available and shared to arrive at reliable causal client representations, the testable prediction follows that providing clinicians with relevant causal knowledge will lead to more elaborate causal reasoning *and* to more reliable causal representations.

If arriving at coherent causal explanations for clients' problems is seen as a prerequisite for the planning of effective interventions, as handbooks prescribe, it would be worthwhile to use causal modelling methods to aid and train clinicians to construct these (e.g. Haynes & Williams, 2003; Morton, 2004). Subsequently, attention should be paid to the testing of these models and the evaluation of the effectiveness of interventions based on them (Haynes, Spain, & Oliveira, 1993; Staines, 2007).

Acknowledgements

We thank Eric and his parents, and Rowella Kuijpers, for their permission to use their personal information for our case materials. We thank Jos Wittebrood for the interruption program. We thank our coders: Miranda Leuverink and Jolin Swinkels. We thank Tracy D. Eells for providing us with the CFCCM, the case formulation content coding manual. We thank all our participants. And we thank our colleagues for their comments on earlier drafts. Their comments contributed greatly to this manuscript.

Chapter 4

Causal Client Models in Selecting Effective Interventions: A Cognitive Mapping Study

Abstract

An important reason to choose an intervention is the expectation that it will be effective. We investigated whether clinicians base their predictions of the effectiveness of interventions on models representing the mechanisms causing and maintaining a client's problems. Forty clinical child psychologists drew causal models and rank ordered interventions according to their expected effectiveness, for two cases. We predicted participants' judgments of effectiveness (i) from their individual causal models, and (ii) from average rankings, mirroring common beliefs. We found that for a majority of participants judgments of effectiveness were better predicted by their causal models, than by the average rankings. Additionally, we found little agreement about the models, and low to moderate agreement about the effectiveness expected from different interventions. Taken together, these results indicate that clinicians indeed base their predictions of the effectiveness of interventions on causal models, but that these models are idiosyncratic.

How do clinicians decide upon interventions for their clients? Hundreds of different interventions exist that clinicians might choose to use, and this number keeps growing (Wenar & Kerig, 2006; Lambert, Bergin, & Garfield, 2004). Probably the most important reason to choose a specific intervention is the expectation that it will be effective. Other factors may also influence clinicians' intervention choices, such as availability, flexibility, fit within budget and time, and client's consent (Nelson & Steele, 2008; Nezu & Nezu, 1993).

The effectiveness of an intervention depends on how it changes the system of interacting psychological, biological, and environmental factors that cause and maintain a client's problems (Carr, 2006; De Los Reyes & Kazdin, 2006; Haynes & Williams, 2003; Kendall, Holmbeck, & Verduin, 2004; Wenar & Kerig, 2006). For example, the effectiveness of different interventions such as social skills training, parent support, or family therapy to reduce a child's aggressive behaviour will depend on whether this aggressive behaviour is caused by a lack of social skills, whether poor parenting reinforces the aggressive behaviour, or whether a child acts aggressively in reaction to parents' marital problems (Wenar & Kerig, 2006).

Given that an intervention is effective to the extent that it changes the mechanisms underlying a client's problems such that problems diminish, we expect that clinicians base their predictions of the effectiveness of interventions on client-specific causal models which specify these mechanisms. Such models allow predictions of the consequences of different interventions (Beach, 1992; Sloman, 2005; Sloman & Hagmayer, 2006). This is also what is recommended in the literature: clinicians should construct case formulations which "aim[s] to describe a person's presenting problems and use theory to make explanatory inferences about causes and maintaining factors that can inform interventions" (Kuyken, Fothergill, Musa, & Chadwick, 2005, p.1188) (see also e.g. Carr, 2006; Eells, 2007; Groth-Marnat, 2003; Wenar & Kerig, 2006).

Causal Assumptions Affect Clinicians' Reasoning

A number of studies has investigated the impact of beliefs about causal structure on clinicians' reasoning. For example, Kim and Ahn (2002) showed that clinicians' theories of how symptoms are causally related in a disorder influence their diagnoses and memory for case information. Clinicians judge symptoms that are causally more central in their theories (i.e., symptoms that affect many other symptoms) as stronger evidence for a diagnosis, and they recall these symptoms better than symptoms that they rate as causally peripheral (i.e., symptoms that do not influence other symptoms). Recently, Cobos, Florez and López (2008) replicated these findings and additionally found that knowledge that a causally central symptom is absent slows down clinicians' information processing more than knowledge that a causally peripheral symptom is absent. Proctor and Ahn (2007) showed that assumptions about the causal structure underlying a patient's symptoms affected the inferences that clinicians draw. Clinicians are more willing to make inferences about unobserved symptoms from symptoms that affect other symptoms than from symptoms that do not affect other symptoms.

Causal Models in Intervention Selection

Research in other domains provides converging evidence that people use models representing causal mechanisms to decide upon effective interventions (Hagmayer & Sloman,

2005, in press; Sloman & Hagmayer, 2006). Individual causal models of coronary heart disease appear strongly correlated with how non-medical and medical students rate the effectiveness of different preventive interventions (Green & McManus, 1995). Similarly, a strong relation is found between the perceived causal structure of scenarios presenting environmental risks, such as chemical pollution or a new epidemic, and people's action tendencies (Böhm & Pfister, 2000).

In the clinical domain, Furnham (1995) showed that lay people's beliefs about which interventions are effective to treat phobia are related to their beliefs about what causes phobia. This indicates that lay people might infer the effectiveness of different interventions from their causal theories. In a similar vein, Yopchick and Kim (2007) investigated how the structure of causal models influenced students' judgments of the effectiveness of interventions for artificial mental disorders. It turned out that interventions which affected symptoms that are initiating causes, with no further causes, were judged to be more effective than interventions which affected symptoms that do have causes.

We know of no evidence to show that clinicians' causal models influence their intervention choices. Witteman and Kunst (1997) and Witteman and Koele (1999) found no evidence for a connection between interpretation of client data and proposed interventions. Clinicians who chose the same intervention considered different client aspects, and clinicians who considered the same client aspects chose different interventions. However, given the poor reliability of clinicians' causal inferences (Garb, 2005; Kuyken et al., 2005; Persons & Bertagnolli, 1999; see also Chapter 3 in this thesis), it seems likely that their causal client models are idiosyncratic. It may thus be the case that although on the group level no relation was found, on an individual level clinicians use their individual causal models when they select interventions.

Case formulations should specify mechanisms underlying clients' problems, from which effective interventions can then be derived. Nevertheless, research findings indicate that mechanisms are often not well elaborated (Eells, Kendjelic, & Lucas, 1998; Kuyken et al., 2005). In addition, from the analysis of clinicians' think aloud protocols it appeared that clinicians' causal client models are rather simple, and that proposed interventions are often unrelated to these models (see Chapter 2 in this thesis). In contrast to these findings, Kendjelic and Eells (2007) showed that after only two hours of training clinicians' case formulations became more elaborate, comprehensive, complex, precise, and contained more references to underlying mechanisms. Thus, clinicians may not usually explicate causal explanations of clients' problems even when they do endorse implicit causal models.

Aim of this Study

The studies reported above indicate that causal assumptions do affect clinicians' reasoning, while there is currently no evidence that clinicians choose interventions based on causal models of their clients' problems. This may be because (i) clinicians don't explicate their causal models if they are not prompted to do so, and (ii) research has focused on average causal models rather than on the causal assumptions of individual clinicians. In the present study we aimed to see whether clinicians base their expectations of the effectiveness of different interventions on individual models explaining the mechanisms underlying a client's problems.

We elicited clinicians' individual causal client models with a cognitive mapping technique (Axelrod, 1976; Hodgkinson, Maule, & Bown, 2004; Huff & Jenkins, 2002; see also Kim & Ahn, 2002, for a similar method used with clinicians). We also asked clinicians to rate the effectiveness of various interventions. We analyzed whether clinicians' individual causal client maps could predict which interventions they rated as most effective. As an alternative, we analyzed whether common beliefs about the interventions' effectiveness could predict individual effectiveness ratings. Additionally, we analyzed clinicians' agreement about the contents of the causal models. We hypothesized that clinicians' effectiveness ratings would be better predicted by their idiosyncratic causal models than by the average ranking expressing shared beliefs.

Method

Participants

Forty Dutch therapists participated who had worked for at least two years with children of primary school age. Their mean experience was 14.4 years (*SE* = 1.7), their mean age 41.2 years (*SE* = 1.9), and they had on average seen 28 children in the last three months (*SE* = 6.4). Participants were randomly assigned to one of two groups. Later analyses showed that randomization was successful: there were no significant differences between groups in experience ($t(38) = -1.1$), age ($t(38) = -1.2$), or number of children seen in the last three months ($t(37) = 0.9$; all p's $> .2$). In the first group there were three men, in the second group four. This is representative for psychotherapists working with children in the Netherlands. Ten participants in the first group and twelve in the second worked in youth care institutions; nine participants in the first group and five in the second in an educational setting. The remaining participants worked in private practice or in a medical setting. In both groups, most participants said they did not adhere to a specific theory, the rest mentioned different theories (e.g., system theory, learning theory).

Materials and procedure

Materials. Two case descriptions were constructed based on descriptions of depressed children from De Wit (2000). Each description was one and a half pages long (810 words). One case description was about a boy named Thijs, ten years old, presenting with depressive symptoms from which he suffered since his parents divorced a year earlier and he lost contact with his father. The other case description was about a boy named Brian, eleven years old, who also presents with depressive complaints which he had developed about a year earlier, when he repeated sixth grade and lost contact with his best friend. We increased complexity in Brian's case by adding information indicating comorbidity of Attention-Deficit/Hyperactivity Disorder, Oppositional Defiant Disorder, and a reading problem (Carroll, Maughan, Goodman, & Meltzer, 2005; Drabick, Gadow, & Sprafkin, 2006; Ostrander & Herman, 2006). We provided similar lengths of text in both cases to describe the situation at home and in school, as well as relationships with peers. In addition, results of six tests were presented with both cases, WISC-III (intelligence test), TRF (behaviour questionnaire filled in by teacher), CBCL (behaviour questionnaire filled in by parents), a depression questionnaire for children, and a questionnaire for children's own perceived competencies. Brian's case also contained scores on a questionnaire for ADHD, and Thijs'

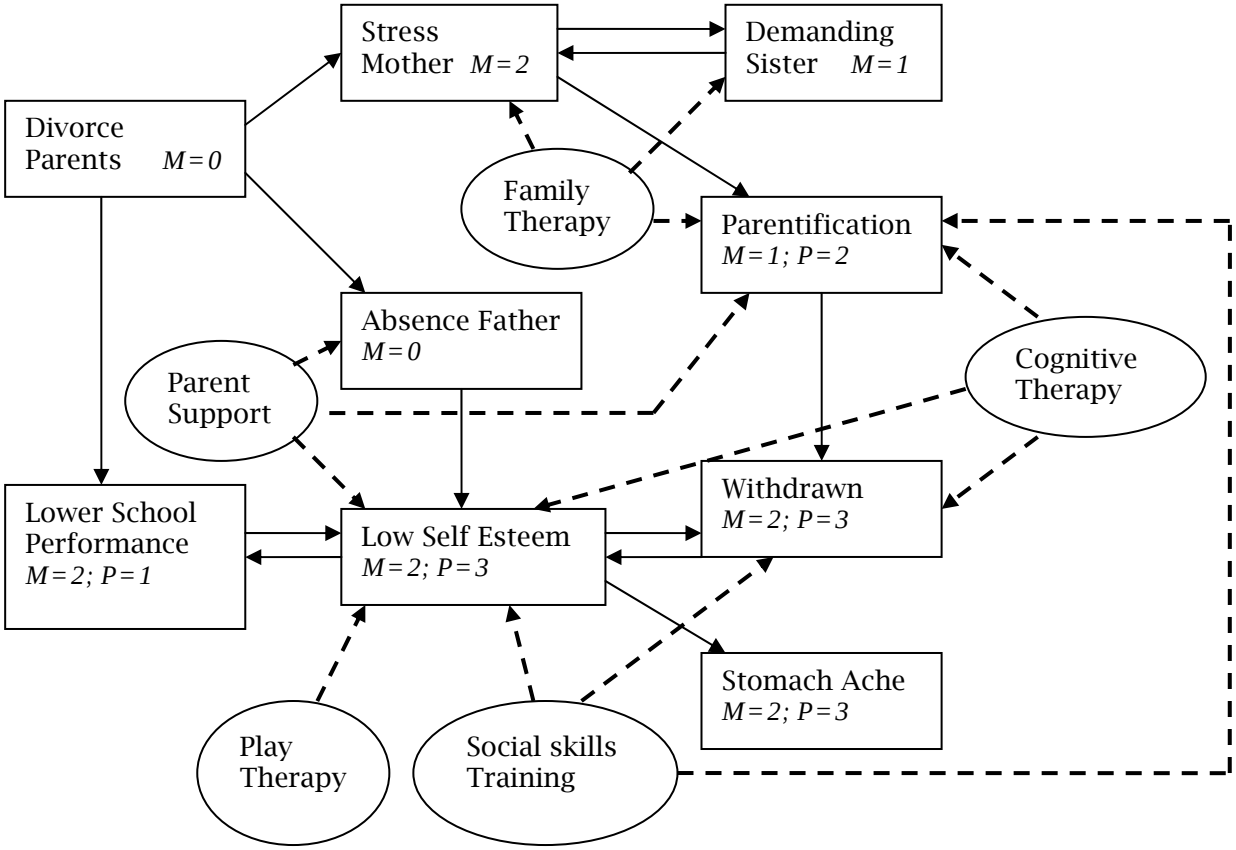
case contained scores on an additional depression questionnaire. Case descriptions were revised by two experienced clinicians to ensure that they were realistic and provided sufficient information to decide on an intervention (see Appendix A for complete case descriptions).

Causal Mapping Task. For the causal mapping task participants received written instructions and an example, which showed six variables connected by several unidirectional and one bidirectional causal links. The task was split into four steps. In the first step participants were asked to write down on index cards the problems they identified, and the causal factors that initiated, increased, maintained, decreased, or prevented these problems. Participants could use as many cards as they thought appropriate, though not more than ten (a pretest confirmed ten to be an adequate number). In the second step participants were asked to rate the importance of the problems. They were instructed that ratings should reflect the degree of harm or distress experienced by the child or others (e.g. parents, teacher), or of the risks inherent to a problem (cf. Haynes & Williams, 2003). The single most important problem should be rated as '1', the one or two second most important problems as '2', and the remaining problems as '3'. In the third step participants were asked to arrange the problems and causal factors on an A3-sized paper and to indicate causal relations by arrows. Participants were allowed to draw bidirectional relations and loops. Participants were also asked to indicate for each relation whether it was initiating / increasing, or preventing / decreasing, with a plus or minus. Also, they had to rate the strength of the relations as weak ('1'), moderate ('2'), or strong ('3'). Finally, in step four participants were asked to indicate for each problem or causal factor how modifiable they thought it was. We explained that the modifiability of a factor could depend on many things, for example resistance to change, or availability of effective treatments (cf. Haynes & Williams, 2003). Easily modifiable factors were to be rated as '2', moderately modifiable as '1' and non-modifiable as '0'.

Interventions Rating Task. In the interventions rating task participants were presented with a list of 10 possible interventions taken from the literature (Carr, 2006; De Wit, 2000): (a) remedial teaching, (b) family therapy, (c) individual cognitive therapy, (d) medication, (e) individual therapy for mother, (f) training impulse control, (g) parent support, (h) individual play therapy, (i) teacher support, and (j) social skills training. Participants were asked to choose and rank the five most effective interventions, that is, the five interventions leading to the highest reduction of the problems identified. They were instructed to assume that each intervention would be the only intervention undertaken, and that other factors such as cost, time, and effort were irrelevant. When finished, they indicated, on a scale of 1 to 7, how confident they were that the intervention they had rated as most effective would indeed be the most effective.

Linking Interventions to Variables in the Causal Models. After having constructed the causal map, and having rated the interventions for effectiveness, participants were asked to indicate which factors or problems in their maps were immediately affected by the five interventions they had chosen. For this, they added the interventions to their maps, and drew causal relations between the interventions and the variables affected. Again

participants rated these causal relations as strong, moderate, or weak, by assigning a '3', '2', or '1' to each arrow. Figure 1 shows an example of a completed causal map.



Note. Variables are depicted as rectangles, interventions as ovals. Straight arrows indicate causal relations among variables, dashed arrows causal effects of interventions. Pluses or minuses and numbers indicating strength are not indicated, for clarity. All causal relations in this model were positive. *P* indicates rated problem importance (1 most important – 3 least important), *M* indicates rated modifiability (2 easily modifiable – 0 not modifiable).

Figure 1. Example of a causal map generated by a participant for Thijs.

Questionnaires. For both cases participants received a short questionnaire, in which they were asked to indicate how many similar cases they had seen before, how difficult and how complex they thought the present case was in relation to other cases they met, whether the description contained sufficient information to rate the effectiveness of interventions, and whether their maps were representative for how they thought about the case.

Procedure. At the start, both groups were told that we were interested in how they thought about client information and decided upon interventions. All instructions were typed out, and participants were encouraged to ask for explanations at any time. The order of the tasks was different for the two groups. In the mapping-first group, participants constructed a causal map for a case first. Then they ranked interventions for their

effectiveness. While making their rankings they could refer back to their causal map, which was still in sight, although they were not asked or prompted to do so. Next they indicated how the five most effective interventions affected variables in their map. Finally, they filled in the questionnaire. Then, the same procedure was repeated for the second case description. In the ranking-first group, the order of tasks was reversed. Participants did the intervention ranking task for the first and the second case before they proceeded to the causal mapping task. Hence, they had to rank interventions for effectiveness without being prompted about causal models.

In both groups the order of cases was counterbalanced. At the end, participants in both groups answered questions about their age, gender, experience, the institution and function they worked in, their theoretical adherence, and they could give comments. Participants were thanked and received a gift certificate of 25 euro. The task took on average one and a half hours. Two research assistants administered the task at the participants' work places, except with two participants who came to the office of the author.

Analyses and Results

Analyses and results will be reported together for better understanding. First participants' assessment of their own models, their rankings, and of the given information will be reported. Then we will report whether clinicians' individual causal client models can predict their rankings of interventions for their effectiveness, and whether an average ranking can. Finally we will report clinicians' agreement about causal models and interventions.

Participants generated 40 causal models for Thijs and 39 models for Brian. One participant failed to generate a causal model for Brian due to personal time constraints. Participants indicated that their maps were representative for how they thought about the case. Means on a scale from '1 does not reflect my thinking' to '7 perfectly reflects how I think about the case' were 4.6 (SE = .2) for Thijs and 5.1 (SE = .2) for Brian, both of which deviated significantly from 4, the midpoint of the scale (Thijs $t(39)=3.2$, p <.01, Brian $t(38)=5.6$, p <.01). Further, participants judged the amount of information as sufficient to rate the effectiveness of the interventions. On a scale of 1 (insufficient) to 7 (too much) they gave mean ratings of 3.6 (SE = .2) for Thijs, and 3.7 (SE = .2) for Brian. Both scores did not deviate significantly from 4, which we interpreted as 'sufficient' ($t(39) = -1.6$, p >.1 ; $t(38) = -1.2$, p >.2). Participants also indicated they were confident that the intervention they rated highest would indeed be the most effective intervention for the boy in the case description: on a scale of 1 (not at all confident) to 7 (highly confident) they gave mean ratings of 5.4 (SE = .2) for Thijs, and 5.2 (SE = .2) for Brian.

Prediction of Individual Rankings from Causal Models and from Average Rankings

To analyze whether participants' rankings of the chosen interventions for their effectiveness can be predicted from their individual causal models we first had to define strategies to derive predictions from the individual causal maps. A strategy describes which information from a causal map is used and how this information is integrated to derive the effects of interventions. The first strategy we call the *maximum impact strategy*. It is based on the functional analysis approach proposed by Haynes and Williams (2003). Its basic idea

is that first the impact of each variable upon all other variables (including the client's problems) is calculated taking into account the strengths of the causal relations among the variables, the importance of the problems and the modifiability of the variables. Second, the effect of each intervention is calculated by summing up the impacts of the variables directly affected by the intervention (see Appendix B for details).

The second strategy we call the *maximum symptom reduction strategy*. It was inspired by causal model theories of decision making (Hagmayer & Sloman, 2009; Sloman & Hagmayer, 2006). This strategy only takes into account the strength of the different causal relations, but not the importance of the problems or the modifiability of the variables. First the reduction of the variables that are directly targeted by the intervention is calculated. Then the resulting reduction of the variables that are affected by these variables is computed, and so on. As the causal influence of an intervention fades quickly, only four iterations had to be used to capture an intervention's effect. Finally, the reductions resulting from each intervention are summed (see again Appendix B for details).

The two causal strategies were compared to a third strategy, which does not consider causal models. This non-causal strategy may be called the *common knowledge strategy*. Its basic idea is that participants simply refer to their expert knowledge about the effectiveness of different interventions. All our participants had seen similar clients before (a median of 20 for Thijs and 40 for Brian) and had ample experience in treating children (14 years on average). Thus they probably share a knowledge base, which they may have employed to guide their judgments. As an indicator for their shared knowledge about the effectiveness of the different interventions we looked at how often participants chose interventions as one of the five most effective, see Table 1. If participants use this common knowledge rather than their own causal models, rankings derived from these frequencies should be the best predictor for their individual rankings. Note that the relative frequency of participants choosing an intervention as effective and the mean ranks assigned to these interventions are very highly correlated, $r = .96$.

As a first analysis we compared participants' individual rankings to the predictions derived from the three strategies described above. For the two causal strategies (maximum impact strategy and maximum symptom reduction strategy) predictions for the five interventions that were considered most effective – and which were thus included in the map – were derived from participants' idiosyncratic causal models. The resulting predictions (impact values in case of the maximum impact strategy and reduction values in case of the maximum reduction strategy) were transformed into ranks. For the non-causal strategy the ranks implied by the frequencies depicted in Table 1 were used. For each participant an individual rank order was computed for the five interventions that this participant actually chose as being one of the five most effective. Finally the ranks derived from the predictions of each strategy were correlated with the rankings actually made by each participant, using Spearman's Rho. While computing the predictions, it turned out that one participant created a causal model for both Thijs and Brian which implied that all interventions would be equally effective. Therefore these two causal models were excluded from the analyses reported hereafter.

After computing the rank correlations between predictions and actual rankings for each case and each participant, we averaged these correlations for each strategy for each case (Thijs vs. Brian) and each group (causal mapping task before ranking vs. causal mapping task after ranking). As some predicted rank orders perfectly matched the actual rank order assigned by participants, we were not able to transform correlations into Fisher z-values before averaging. In consequence, the reported means underestimate the actual mean correlations. Table 2 shows the resulting twelve mean correlations. All are above .3, some as high as .62.

Table 1

Relative Frequencies and Mean Ranks (SE) of Interventions

Interventions	Thijs		Brian	
	Relative frequency	Mean ranks (*SE*)	Relative frequency	Mean ranks (*SE*)
Remedial teaching	.13	5.68 (.15)	.54	4.51 (.26)
Family therapy	.70	3.15 (.32)	.51	4.67 (.30)
Individual cognitive therapy	.83	3.55 (.26)	.46	4.87 (.26)
Medication	.05	5.90 (.06)	.69	3.36 (.33)
Individual therapy mother	.63	4.48 (.23)	.51	4.85 (.26)
Training impulse control	.00	5.98 (.03)	.49	4.64 (.27)
Parent support	.95	2.55 (.22)	.90	2.23 (.25)
Individual play therapy	.85	3.28 (.29)	.28	5.41 (.18)
Teacher support	.43	5.20 (.19)	.44	4.87 (.22)
Social skills training	.38	5.15 (.20)	.15	5.74 (.12)

Note. Ranks ranged from 1 (best) to 5 (least effective). Interventions not included within the five most effective interventions received rank 6.

In order to test whether these mean correlations were significantly different from chance, we created a theoretically derived test distribution. First we created all possible combinations of two rank orders with ranks 1-5, allowing for double ranks. Next we correlated these rank orders using Spearman's Rhos. Then we randomly drew 3000 sets of 20 correlations, and averaged each set. Five percent of the resulting mean correlations were larger than .19 and 1 % was larger than .27. Thus, all mean correlations depicted in Table 2 are significantly above chance, which means that with all three strategies we are able to predict participants' effectiveness rankings to some extent.

Table 2

Mean Correlations between Participants' Effectiveness Rankings and the Effectiveness Rankings Derived from Different Strategies Conditional on Case and Task Order Group

	Causal mapping before ranking task				Causal mapping after ranking task			
	Thijs		Brian		Thijs		Brian	
Strategy	Mean	(*SE*)	Mean	(*SE*)	Mean	(*SE*)	Mean	(*SE*)
Maximum impact	.59	(.11)	.45	(.08)	.44	(.09)	.49	(.06)
Maximum symptom reduction	.62	(.10)	.33	(.10)	.49	(.08)	.62	(.06)
Common knowledge	.46	(.08)	.51	(.08)	.48	(.09)	.48	(.09)

One may expect to find a higher correlation for predictions derived from the causal model for participants who ranked the interventions directly after constructing their causal model. Using independent-sample t-tests to compare the mean correlations for each strategy between conditions, only one difference reached significance. The maximum symptom reduction strategy predicted participants' effectiveness ranking better when participants did the rating tasks first than when they did the causal mapping task first ($t(36) = -2.5$; $p<.05$, equality of variances not assumed). No other test reached significance (all ts <1, $p> .4$). Thus the only difference we found contradicted the above speculation, and despite descriptive differences there seems to be no overall effect of task order.

So which strategy predicted participants' rankings of effectiveness best? No clear pattern emerged on the group level, as can be seen from the findings reported in Table 2. Correlations were not systematically higher for causal strategies than for the non-causal strategy or vice versa. The same is true when the two causal strategies are compared with each other, although the maximum symptom reduction strategy seems to be slightly more predictive. In addition, mean correlations were all substantial, but not very high (overall average of .49). However, it is quite plausible that individual participants used different strategies to assess effectiveness. Some may have used causal strategies, others common knowledge, and others may have used completely different strategies. Therefore we went

back to each individual participant and each case and compared the correlations for the three strategies. We classified participants into four categories using the following scheme: If the correlation of one of the causal strategies was at least rho ≥ .5 and was at least .2 higher than for the common knowledge strategy, the participant was classified as using a causal strategy. If the correlation of the common knowledge strategy was at least rho ≥ .5 and .2 higher than either of the two causal strategies, the participant was classified as using a non-causal, common knowledge strategy. If the correlations for the causal strategies and the common knowledge strategy were both rho ≥ .5 and did not differ by at least .2 participants were classified as tied between causal and non-causal. Finally if neither strategy reached rho ≥ .5 participants were classified as using another strategy. We decided to use rho ≥ .5 as the threshold for classification because the overall mean correlation between any of the strategies' predictions and the actual rankings was .49. So for all participants who were classified by the scheme described above as using a specific strategy the predictions had a higher than average correlation with the actual rankings.

Results of this classification are presented in Table 3. For Thijs 44% of the participating clinicians seem to have used a causal strategy, while only 10% seem to have followed common knowledge. In the case of Brian 47% of clinicians seem to have used a causal strategy over 24% who seem to apply common knowledge. For both cases there was a substantial number of participants for whom causal and non-causal strategies made equally good predictions. Surprisingly, only a rather small number of participants had to be classified as 'other' (a total of 6 participants for each case).

Table 3

Percentage of Participants Using either a Causal or a Non-Causal Strategy

	Thijs	Brian
Causal strategy (maximum impact or maximum symptom reduction strategy)	44	47
Non-causal strategy (common knowledge strategy)	10	24
Tie between causal and non-causal	31	24
No prediction	15	15

As a final analysis we investigated how well the strategies that participants seemed to have used predicted their rankings of effectiveness. To do so we calculated the mean correlations between predicted rankings and actual rankings. These findings are given in Table 4. Mean correlations are reported separately for the maximum impact and the maximum reduction of symptoms strategies for all participants using a causal strategy. As Table 4 shows mean correlations were much higher than the ones reported in Table 2.

Especially the maximum symptom reduction strategy, which simulates the reduction of problems resulting from an intervention in the causal model, does a good job for both cases. It is able to account for more than 50% of the variance in rankings of participants pursuing a causal strategy.

Table 4

Mean Rank Correlations between Predictions Derived from Individual Causal Maps or Common Knowledge and Effectiveness Rankings of Interventions

	Thijs	Brian
Causal strategy 1: Maximum impact strategy	.82	.64
Causal strategy 2: Maximum symptom reduction strategy	.76	.74
Non-causal strategy: Common knowledge strategy	.63	.76

Content of Causal Models and Agreement

Table 5 depicts the number of variables (i.e., problems and causal factors) and causal relations in participants' causal models, regardless of their specific content. As Table 5 shows, participants tended to construct more complex causal models for Brian than for Thijs. Paired samples t-tests showed that participants used more variables ($t(38) = 2.4$; $p < .05$), drew more positive relations ($t(38) = 2.4$; $p < .05$), and drew more bidirectional relations ($t(38) = 3.7$; $p < .01$) for Brian than for Thijs.

Table 5

Descriptive Statistics of Causal Models Generated by Participants: Means and (SE)

	Thijs				Brian			
	N	Positive relations	Negative relations	Bi-directional relations	*N*	Positive relations	Negative relations	Bi-directional relations
Variables	7.98 (.27)	13.7 (.83)	.23 (.10)	2.38 (.38)	8.64 (.21)	15.5 (1.06)	.46 (.19)	3.80 (.53)
Interventions	5	12.2 (.75)			5	12.5 (.91)		

These findings conform to the fact that Brian but not Thijs shows several signs of comorbidities. Interestingly, the difference between the models is not reflected by participants' subjective assessments. Mean difficulty ratings for Thijs were 3.6 (*SE* = 0.2) on a scale from '1 = not at all' to '7 very', for Brian they were 3.3 (*SE*= 0.2). Mean complexity ratings for Thijs were 3.6 (*SE* = 0.2), and for Brian 3.6 (*SE* = 0.2). Paired t-tests showed no differences for the cases ($t(37) = 1.1$, $p > .2$; $t(37) = 0.2$, $p > .8$).

In order to compare the content of participants' individual causal models, variables had to be coded into the same categories. Categories were defined by three coders (the author and two research assistants) on the basis of ten randomly selected maps. This initial analysis resulted in 13 content categories for Thijs and 20 for Brian. Next, the coders independently coded the variables of the remaining maps into these categories, adding a category each time they thought that the content of a card was not adequately reflected by any of the categories created before. In the end, for Thijs 15 content categories and for Brian 23 content categories were defined. For Thijs, all three coders agreed for 90.8 % of the cards; for 8.9 % two of the three coders agreed. For Brian all three coders agreed about 83.2 % of the cards; for 15.4 % two of the three coders agreed.

Table 6 shows the categories and their relative frequencies for both cases. In addition, the mean importance of the variables as indicated by participants is shown, with higher values indicating more important problems. Values were computed by assigning 8 points to the most important problem, 4 points to the two second most important problems, 2 points to the least important problems and 1 to variables not representing a problem. This was done separately for each model drawn by each participant. In order to assess agreement among participants, kappa values were calculated over categories that were mentioned by at least five participants. Kappa was .20 for Thijs and .29 for Brian, which indicates a poor agreement given conventional standards (Fleiss, 1981). Nevertheless, both kappa values were statistically significant ($p < .01$)

Table 6

Relative Frequencies of Categories and Importance Ratings

Thijs			Brian		
Categories	Relative frequency	Mean importance	Categories	Relative frequency	Mean importance
Divorce	.93	3.16	Emotional stress mother	.95	2.27
Depressive complaints	.93	5.14	ADHD characteristics	.92	4.50
Stress mother	.90	2.67	Parenting problems	.87	2.71

Table 6
Continued

Demanding sister	.75	1.67	Behavioural problems at home	.82	4.34
Absence father	.68	2.52	Low school performance	.69	2.70
Parentification	.65	3.00	Relation to father	.59	1.61
Few social contacts	.60	2.79	Depressive complaints	.51	3.10
Low school performance	.58	2.73	Low self-esteem	.49	3.68
Low self-esteem	.43	3.47	Behavioural problems school	.41	3.75
Physical complaints	.40	2.31	Few social contacts	.33	1.77
Protective factors	.18	1.00	Repeating grade	.26	2.10
			Protective factors	.21	1.17
			Friend moved	.18	1.29
			Negative social interactions	.18	3.14
			Genetic predisposition	.15	1.17

Note. Relative frequencies are only given for categories identified in maps of at least 5 participants. The importance of problems was recoded so that higher values indicate more important problems. A value of 1 indicates that the category is no problem at all.

Causal relations were retained in coding. Whenever coding resulted in more than one causal relation between two categorical variables, we copied the strongest relation. For example, ‘sad mood’ and ‘being withdrawn’ could be mentioned on different cards, and both be caused by ‘poor school performance’. In coding, ‘sad mood’ and ‘being withdrawn’ become one category: ‘depressive complaints’. Here, we copied the strongest causal relation from ‘lower school performance’ to one of the two original variables as the relation between

'poor school performance' and 'depressive complaints'. Table 7a shows the frequencies of causal links, conditional on the number of participants with both categories in their models, for Thijs; Table 7b shows these results for Brian. For example, 34 participants included both 'divorce' and 'depressive complaints' in their model for Thijs and 24 of these participants assumed that divorce was a factor contributing to the depressive complaints. Agreement among participants was calculated by computing kappa using pairs of categories mentioned by at least 5 participants. Agreement was again poor both for Thijs ($\kappa = .29$) and for Brian ($\kappa = .26$), although kappa values were again statistically significant (p's < .01).

Agreement about the effectiveness of the ten possible interventions for both Thijs and Brian was depicted in Table 1, which shows the relative frequencies of participants choosing each of the interventions as one of the five most effective. Based on these relative frequencies again kappa values were calculated, yielding a moderate agreement for Thijs ($\kappa = .43$), but a really poor agreement for Brian ($\kappa = .12$). Both kappas were again significant (p's<.01). A series of t-tests for independent samples yielded no significant differences between the groups. Participants chose the same interventions as most effective regardless of whether they first constructed a causal model, or did so after they had made their rankings.

We also calculated participants' agreement about the effects of the ten interventions. To do so, we again first determined the numbers of participants with each possible combination of interventions and categories. If at least five participants had a certain combination we determined the frequency of a causal link. The relative frequencies are shown in Tables 8a and 8b. Based on these relative frequencies again kappas were computed, yielding a moderate agreement for Thijs ($\kappa = .44$) and Brian ($\kappa = .41$). Thus participants at least seem to agree which categories are affected by the interventions.

General Discussion

An important reason to choose a specific intervention is that it is expected to be effective. Because the effectiveness of an intervention depends on how it changes the mechanism underlying a client's problems, we expect clinicians to base their predictions of how effective interventions will be on causal client models representing these mechanisms. We elicited clinicians' causal models for two cases, and asked clinicians to rate the effectiveness of interventions. We hypothesized that clinicians' effectiveness ratings would be better predicted by their idiosyncratic causal models than by the average effectiveness ranking.

We defined two strategies to derive effectiveness predictions for interventions from individual causal maps, the maximum impact strategy and the maximum symptom reduction strategy. In addition, we derived average rankings from the relative frequencies of participants choosing an intervention as one of the five most effective. We took this average ranking to represent participants' common beliefs about the effectiveness of different interventions in similar cases. Averaged over participants, both causal strategies and the average ranking predicted effectiveness ratings better than chance. However, looking at individual participants it appeared that for a majority of participants the causal strategies predicted their effectiveness ratings better than the average ranking did. For most other

Table 7a

Frequencies of Causal Relations Conditional on the Number of Participants Having Both Categories in their Causal Maps for Thijs

	Divorce	Stress mother	Demanding sister	Absence father	Parenti-fication	Few social contacts	Depressive complaints	Physical complaints	Low self-esteem	Low school perform.	Protective factors
Divorce		20/33	11/29	17/24	12/23	6/22	24/34	4/15	6/17	3/21	1/8
Stress mother	3/33		15/29	1/24	21/23	8/21	22/33	2/15	6/18	3/21	1/7
Demanding sister	1/29	16/29		1/22	12/20	1/17	11/29	2/13	4/16	1/17	0/7
Absence father	1/24	5/24	5/22		5/18	3/16	17/26	1/14	7/11	3/15	0/7
Parentification	1/23	7/23	5/20	0/18		4/16	12/24	6/12	8/14	1/15	0/5
Few social contacts	0/22	2/21	1/17	1/16	0/16		11/22	3/12	6/9	3/16	1/5
Depressive complaints	2/34	7/33	3/29	3/26	0/24	17/22		7/17	8/15	19/23	0/7
Physical complaints	0/15	0/15	0/13	0/14	0/12	1/12	4/17		1/7	2/13	
Low self-esteem	0/17	1/18	1/16	0/11	4/14	4/9	11/15	3/7		4/8	0/5
Low school performance	0/21	1/21	0/17	0/15	0/15	0/16	8/23	5/13	4/8		
Protective factors	1/8	1/7	0/7	0/7	1/5	1/5	3/7		2/5		

Note. Empty cells indicate combinations which fewer than five participants had in their model. Categories in the left column are indicated as causes, categories in the top row as effects.

Table 7b

Frequencies of Causal Relations Conditional on the Number of Participants Having Both Categories in their Causal Maps for Brian

	Stress mother	Relation to father	Parenting problems	Few social contacts	Depressive complaints	Low self-esteem	Low school performance	Protective factors	ADHD characteristics	Behav. prob. home	Behav. prob. school	Friend moved	Repeating grade	Negative social inter.	Genetic predis.
Stress mother		6/23	21/32	0/13	9/19	6/18	1/25	0/7	5/34	14/32	0/15	0/7	1/9	0/7	0/5
Relation to father	12/23		14/19	2/11	5/10	4/12	0/15	0/6	3/23	7/19	0/8	0/7	1/6		
Parenting problems	21/32	5/19		0/11	6/17	4/17	1/24	0/6	12/31	20/27	3/15	0/7	0/9	0/6	
Few social contacts	1/13	0/11	0/11		1/5	4/7	0/9		1/13	3/11	1/2		0/5	0/3	
Depressive complaints	1/19	1/10	2/17	2/5		2/6	2/14		1/18	6/16	3/6		1/5	2/4	
Low self-esteem	2/18	1/12	0/17	2/7	2/6		4/13		3/17	8/15	5/8		1/5	2/4	
Low school performance	2/25	1/15	0/24	1/9	9/14	9/13			3/25	3/23	4/10		5/6	0/6	0/5
Protective factors	0/7	1/6	2/6						1/6	1/5					
ADHD characteristics	8/34	2/23	11/31	4/13	7/18	6/17	13/25	0/6		17/29	13/14	0/7	3/9	5/7	0/6
Behav. prob. home	10/32	1/19	8/27	4/11	6/16	8/15	0/23	0/5	3/29		3/14	0/5	0/7	4/6	0/5
Behav. prob. school	0/15	0/8	0/15		3/6	3/8	1/10		3/14	2/14					
Friend moved	0/7	0/7	0/7						1/7	3/5					
Repeating grade	1/9	1/6	0/9	2/5	2/5	4/5	2/6		2/9	2/7					
Negative social interactions	0/7		0/6				0/6		2/7	1/6					
Genetic predisposition	1/5						2/5		3/6	1/5					

Note. Empty cells indicate combinations which fewer than five participants had in their model. Categories in the left column are indicated as causes, categories in the top row as effects.

Table 8a

Frequencies of Direct Effects Conditional on the Number of Participants Having Both a Category in their Causal Map and Chosen an Intervention as one of the Five Most Effective for Thijs

	Divorce	Stress mother	Demanding sister	Relation father	Parenti-fication	Few social contacts	Depressive complaints	Physical complaints	Low self-esteem	Low school performance
Remedial teaching	0/5			0/5			1/5			
Family therapy	7/26	18/24	21/21	10/17	12/18	4/18	6/25	1/12	5/13	1/16
Individual cognitive therapy	6/30	0/29	2/24	3/22	10/23	10/22	25/30	6/15	15/15	7/20
Medication										
Individual therapy mother	8/23	22/23	3/19	3/18	4/17	1/14	1/23	0/12	2/8	0/14
Training impulse control										
Parent support	5/35	29/34	18/28	10/25	16/25	7/23	1/35	1/15	5/16	0/21
Individual play therapy	8/32	0/32	2/25	3/22	4/22	4/19	28/31	6/12	15/15	5/18
Teacher support	0/15	0/16	0/13	0/12	0/11	4/11	10/17	1/5	7/8	11/12
Social skills training	0/14	0/13	0/13	0/11	1/9	9/12	9/14	0/6	6/9	0/6

Note. Empty cells indicate combinations which fewer than five participants had in their model. In the left column interventions are listed, which affect categories as mentioned in the top row.

Table 8b

Frequencies of Direct Effects Conditional on the Number of Participants Having Both a Category in their Causal Map and Chosen an Intervention as one of the Five Most Effective for Brian

	Stress mother	Relation to father	Parenting problems	Few social contacts	Depressive com-plaints	Low self-esteem	Low school perf.	Protec-tive factors	ADHD char.	Behav. prob. home	Behav. prob. school	Friend moved	Repea-ting grade	Genetic predis. father
Remedial teaching	0/19	0/13	1/17	1/6	5/14	2/10	17/17		1/20	3/18	2/8		3/6	
Family therapy	14/19	9/12	17/19	1/8	7/12	1/7	0/12	0/5	3/18	8/14	1/6		1/5	
Individual cognitive therapy	0/18	0/8	0/17	2/7	8/11	8/9	0/12		4/16	10/16	4/8		2/6	
Medication	2/27	0/18	1/23	4/11	3/12	2/13	4/19		27/27	12/25	4/11		0/6	2/5
Individual therapy mother	20/20	0/10	3/19	0/5	1/12	0/11	0/17		0/17	2/17	0/9		0/5	
Training impulse control	0/20	0/14	0/16	5/8	2/9	4/9	1/16	0/5	16/20	11/19	5/9			
Parent support	20/34	15/21	29/31	1/11	3/17	7/18	1/25	0/6	5/34	12/30	1/16	0/6	0/8	1/7
Individual play therapy	0/9		0/10		4/6	4/6	0/5		2/10	1/6				
Teacher support	1/17	0/11	0/17	6/8	4/6	9/12	8/12		7/16	8/14	8/8	0/5	3/7	
Social skills training	1/5		1/5			1/5	0/5		2/6					

Note. Empty cells indicate combinations which fewer than five participants had in their model. In the left column interventions are listed, which affect categories as mentioned in the top row.

participants the causal strategies and average ranking predicted effectiveness ratings equally well. Only for a minority of participants did the average ranking predict ratings best. Thus, clinicians' effectiveness ratings were indeed consistent with their individual causal models, and more so than with common beliefs about the effectiveness of the interventions.

Analyses of the agreement about the contents of the causal models showed that clinicians' models were idiosyncratic. Participants focused on different variables and assumed different causal relations. This idiosyncrasy is not unexpected. Clinicians have been found to disagree about case formulations, and especially about causal inferences (Garb, 2005; Kuyken et al., 2005; Persons & Bertagnolli, 1999; see Chapter 2). Agreement about the effectiveness of interventions was moderate to poor. Interestingly, participants agreed moderately about what variables are affected by the interventions for both cases. Thus, it seems that although participants think differently about mechanisms underlying the boys' problems, they share beliefs about the effects of the interventions.

Taking these findings together we conclude that, despite the idiosyncrasy of the causal models, participants' effectiveness judgments are more in line with their individual models than with shared beliefs about the effectiveness of these interventions. Thus, our hypothesis is supported. Our findings are also consistent with previous studies showing that causal assumptions affect clinicians' diagnosis (Kim & Ahn, 2002) and inferences about unobserved symptoms (Proctor & Ahn, 2007). Further, similar findings had been obtained for students' ratings of the effectiveness of actions to prevent coronary heart disease (Green & McManus, 1995) or reduce environmental risks (Böhm & Pfister, 2000). Also related are findings in the area of legal decision making. Jurors were shown to have different mental causal models of what happened in a crime, and these different models led to different verdicts for the same suspect (Huntley & Costanzo, 2003; Pennington & Hastie, 1988).

Strategies for Predicting the Effectiveness of Interventions

We could predict most participants' effectiveness ratings quite well with the causal strategies we defined, and for some participants even perfectly so. Hence, our findings indicate that most clinicians considered their own causal model when judging the effectiveness of different interventions. What we do not know is which precise strategy each clinician used. Looking at the complexity of the causal models drawn, it seems that clinicians may not derive the effectiveness of interventions from their models by taking into account all variables directly and indirectly affected, the strength of all causal relations, causal feedback loops, the modifiability of variables, and problem importance. Instead, they may use simpler heuristics (cf. Gigerenzer & Goldstein, 1996). For example, as the effect of an intervention rapidly fades the further away a problem is from the point of intervention, participants may refrain from inferring distant, but weak effects of an intervention. Note that such simpler heuristic strategies are still causal, because they still infer effectiveness of interventions from participants' own causal client models.

Further, it is possible that clinicians base their expectations for how effective different interventions will be not on a specific causal model constructed for an individual client, but on a more general causal model of behaviour. Robins, Mendelsohn, Connell, and Kwan (2004) found that people disagree about causes of specific behaviours of others and themselves. However, people tend to arrive at similar explanations for other people's and

their own behaviours. Thus, Robins et al. suggest that people may have a personal, implicit theory about what generally causes behaviour in themselves and others, and that different individuals may have different theories. Similarly, therapists in our study might have implicit theories about what generally causes (problematic) behaviour, emotions, and cognitions, and use these theories to infer the effectiveness of interventions (cf. Hong, Chiu, Dweck, Lin, & Wan, 1999).). Alternatively, they may have disorder-specific causal models (cf. Kim & Ahn, 2002). However, inspection of the contents of the causal models shows that participants used different variables for Thijs and Brian (see Table 6). Also, for variable categories which are present in models for Thijs and Brian participants show different beliefs about causal relations (see Tables 7a and 7b). For example, for both boys depressive complaints were identified as a category. However, relatively more participants believed that depressive complaints lead to fewer social contacts, low self esteem, and low school performance for Thijs than for Brian. Thus, the causal models that participants constructed for both boys seem to be client-specific rather than general.

As a non-causal strategy to infer the effectiveness of interventions we defined the common knowledge strategy, to reflect experiences of what interventions had been effective for similar clients before, generalized over clinicians. We reasoned that this would be appropriate since our participants were quite experienced and had seen many similar cases in their work. The fact that agreement about the effectiveness of interventions was not too high, especially not for Brian, indicates that participants might have used idiosyncratic rather than common experiences. These idiosyncratic experiences might result from different clients they have met, but also from subjective interpretations of causes for client problems and of the effectiveness of interventions (Haynes, Spain, & Oliveira, 1993; cf. Robins et al., 2004). Further, clinicians might refer to prototypical or stereotypical clients, and have different prototypes and stereotypes (Garb, 1996). Different judgments may also originate in differences in knowledge obtained from scientific literature, colleagues, or other sources.

To conclude, clinicians may use a variety of causal and non-causal strategies to derive judgments of effectiveness. Our results indicate that many clinicians derive their judgments from their idiosyncratic causal models.

Limitations

A limitation of this study is the two cases, which both presented problems of young boys. Also, all clinical psychologists participating in the study worked with children. Therefore, we cannot be sure to what extent these results can be generalized to other cases, and to clinicians working with adolescents or adults.

In our study we explicitly asked participants to draw a causal model of a client's problems. One may suspect that this procedure drew participants' attention to the importance of causal considerations for inferring the effectiveness of different interventions. Note however, that only half of the participants constructed a causal map before ranking interventions for effectiveness. The other participants first ranked interventions. We found no differences between these two groups in (i) the complexity of the generated models, (ii) the mean ranks assigned to the ten interventions, and (iii) correlations between rankings participants gave and rankings based on any of the causal and non-causal

strategies (except one in the opposite direction to what would be predicted, see Analyses and Results section). Thus it seems that we in fact tracked participants' causal assumptions.

Finally, we only asked participants to predict the effectiveness of different interventions and not to choose an intervention. However, as we mentioned in the beginning, in practice intervention choices will not only be determined by predicted effectiveness. Factors such as available time and budget, interventions available in an institution, and client-specific (e.g. motivation) or clinician-specific factors (e.g. experience) might also influence intervention choices (Beutler, Malik, Alimohamed, Harwood, Talebi, Noble, & Wong, 2004; Clarkin & Levy, 2004; Nelson & Steele, 2008). For a complete understanding of intervention choices, these other factors have to be taken into account as well.

Implications for Clinical Practice

We focused on the question whether psychotherapists base their intervention choices on causal client models. Another question is whether they should do so in practice. In the introduction we referred to handbooks which advocate basing intervention choices on case formulations which explain how clients' problems are caused and maintained (Carr, 2006; Eells, 2007; Groth-Marnat, 2003; Wenar & Kerig, 2006). However, it seems difficult for clinicians to make reliable causal inferences about the mechanisms underlying clients' problems. In addition, it is often unknown why precisely interventions are effective (De Los Reyes & Kazdin, 2006; Haynes & Williams, 2003). Clinicians might do better to choose empirically validated interventions based on classifications of clients' problems. But empirically validated interventions may not always be available, and classifications may not always be straightforward, as in cases with comorbidity of disorders (Nelson-Gray, 2003). Although evidence for better treatment outcomes following case formulations explicating causal mechanisms is scarce and not uniformly positive, some studies indicate that for complex cases and severe problems better outcomes have been associated with the identification of causal mechanisms (Beutler et al., 2004; Bieling & Kuyken, 2003; Eells, 2007; Nelson-Gray, 2003).

Our study indicates that clinicians' idiosyncratic causal models lead to different ratings of effectiveness for interventions. As a consequence they may lead to different intervention choices for the same client. We believe that clinicians should be cautious when making inferences about underlying causal mechanisms (see also Garb, 2005), and that they should put assumed mechanisms to stringent tests (e.g. Haynes, Spain, & Oliveira, 1993; Staines, 2007).

Acknowledgments

We thank Annemieke Lamers and Annemieke van Harten for collecting most of the data as part of their master thesis. We thank Rowella Kuijpers and Nicole Krol for revision of the case materials. We thank the members of Michael Waldmann's Lab and of the Diagnostic Decision Making group all for their helpful comments. And finally, this study would not have been possible without the enthusiastic participation of the psychotherapists, thanks!

Appendix A

Case description Thijs (translated from Dutch)

Thijs is eleven years old. Mother is an analyst, father is a sales representative. The parents are divorced. Thijs has a six year old sister. Thijs has been brought in by mother, on the advice of Thijs' teacher. The teacher was worried about Thijs and asked the pedagogical assistant in school, after having consulted mother, to have Thijs fill in the KDVK (Short Depression Questionnaire for Children). Thijs appeared to reach a score in the clinical range. Thijs works hard at school, but his achievements have detoriated since last year. The teacher thinks Thijs is a good and docile pupil, but he misses all initiative. He never raises his hand, and he will never ask anything. With creative subjects it takes long before he makes anything. With drawing for example he thinks a long time and keeps rubbing out, when working with clay he keeps moulding. With gym and playing outside he is withdrawn. It seems that he does not want to join others, but prefers to watch. He has one friend from another class. The teacher of the previous year says that Thijs used to be more actively involved in class and used to play a lot with other children. Mother was very surprised to hear that Thijs might be depressed. She does not have many problems with Thijs. He is always quiet, sweet, and obedient. He does however often complain about stomach ache. Thijs' sister on the contrary is very difficult to handle and hot-tempered. She demands all attention, because she refuses to eat or sleep. Mother says she finds it strange to be here for Thijs, she could use more help for his sister. Thijs always wants to help his mother with cooking and doing the dishes. Mother says she often finds that inconvenient. Thijs often helps his mother by looking after his sister, when she goes shopping for example. Thijs watches television a lot and often plays with his Gameboy. He does not want to train and has not joined any club. He did play soccer, but he stopped after the divorce, because father no longer took him to the training and competitions. Father left his family two years ago, he was having an affair. He didn't comply with the arrangement concerning parental visits. Often he did not come to pick up the children when this was agreed on. Thijs didn't like being at father's place, he always wanted to go back to his mother. This year Thijs has seen his father only at his grandmother's birthday. Mother has had great problems with the divorce. She felt rejected and abandoned. She still feels very lonely. She finds it hard to take care of the children all by herself. She has to do everything alone, while she works four days a week as well. During the interview Thijs does not speak much. If he is asked something, he looks at his shoes and often says: "I don't know". Thijs is cooperative during the tests, he seems to enjoy them. He says he would like to play with friends more often, but that this is often not possible because his mother cannot bring or fetch him. He finds it unpleasant that his sister is so difficult; he is sorry for his mum. He explains that his mother has to work hard, since she has to do everything alone. He says he used to miss his father a lot, but not anymore. What he misses most is that there are no fun activities anymore. With father they often used to go to the zoo, or swimming. He says he likes school, but often does not know what he has to do. Pregnancy and birth were normal. Thijs' development was fine. He has always been a sweet and easy child. When he was young, he suffered from frequent ear infections. After surgery at the age of three the infections disappeared. He has always been

careful with his sister. At school he achieved average results until this year. On the TRF (behaviour questionnaire filled in by the teacher) scores fall in the clinical range for withdrawn and anxious/depressive behaviour. On the CBCL (behaviour questionnaire filled in by mother) scores fall in the clinical subrange for somatic complaints. From the WISC-III (intelligence test) Thijs appears to have an IQ of 104, with a harmonic profile. The CBSK (Scale for children's perceived competences) shows low scores in the domains school performance, social acceptation, sports achievements, and self esteem. From the DVK (Depression questionnaire for children) it appears that Thijs thinks negatively of himself and his environment. Mother feels guilty that she didn't notice that Thijs felt so bad. She hopes that he will feel better soon and is prepared to do anything that is necessary. Father has been informed about the assessment, but he did not want to be involved.

Case description Brian (translated from Dutch)

Brian is ten years old. Mother is a housewife, father is a self-employed handyman. Brian is their only child. Brian has been brought in by his mother, because she can no longer influence Brian's behaviour, and on the advice of the teacher. Mother says it seems as if there's nothing but fights at home. She cannot ask or tell Brian anything. He gets angry at everything. He starts to shout, goes outside or to his room stamping his feet and slamming the doors. Brian has always been an active child and has always needed a lot of attention. He was always playing wild games, testing the limits and getting into mischief. He cannot sit still and always dominates the conversation. But now it seems there is nothing but fights. Mother says she's depressed and can no longer put up with these continuing fights. Father works a lot, also most Saturdays. Father recognises a lot in his son, he used to be very active too. He says he still cannot sit still. His father was very strict with him. He thinks Brian needs a firm approach too, which he does not get from his mother. When Brian doesn't listen to him, father sends him upstairs, and he isn't allowed to come back for the rest of the evening. Brian obeys his father more than his mother. Mother starts to cry when father criticises her approach to Brian. According to the teacher Brian has always been highly visible at school and has always been easily distracted. At the same time Brian was always a pleasant pupil, with a sense of humour and helpful to other children. He's repeating the sixth grade, because he is two years behind with reading, as appears from AVI and 1-minute tests (reading competence measures). The teacher thinks Brian has changed a lot since this last year. He has become very uninterested; he often sighs aloud and gives a lot of negative comments on everything he is asked to do. He doesn't comply with rules and agreements. When there's a conflict he walks out of the classroom. Usually his work is not finished, sloppy, and full of mistakes. In the schoolyard he's constantly engaged in fights. He swears, pushes, and kicks. If the teacher wants to talk, he says he doesn't care. Frequently he appears to be lying. Brian says he always does everything wrong. Things always go wrong and everyone always gets angry. He understands that everyone gets angry because he keeps doing stupid things. When he tries his best, he just doesn't manage. He just doesn't like school anymore. His best friend has moved in the summer holidays. Other children, who he calls babies, always start to fight and he always gets the blame. At home there isn't much to do. Usually he watches television or plays computer games. He also often goes outside. He

bikes around in the neighbourhood, but there isn't much to do there either. He does not often play with the children from his former class anymore. Pregnancy and birth were normal. Brian used to sleep little and to cry a lot when he was a baby. In his development he was quick with everything; he could walk with ten months of age. He has visited the emergency department more than once, because he was stunting on his bike, jumped off the roof of a shed, and was hit when he suddenly crossed the street. Apart from this, he has always been in good health. Brian has always had a lot of friends and his family was fond of him. He always used to make everybody laugh with his jokes. Three years ago, mother received help from the RIAGG (= mental health care institution) for half a year. After having had a miscarriage she became depressed. She stopped treatment there because she felt better. She says that she has again been suffering from depressive complaints for a while now, and wants to seek help again. Both the CBCL (behaviour questionnaire filled in by parents) and TRF (behaviour questionnaire filled in by the teacher) scores fall in the clinical ranges for aggressive behaviour and attention problems, and in the clinical subrange for anxious/depressive behaviour. On the AVL (ADHD-questionnaire) there are high scores in the domains hyperactivity, impulsivity, and attention problems. From the WISC-III (intelligence test) it appears that Brian has an average intelligence (IQ=97), he has low scores on the parts 'Substitution' and 'Digit Series'. The CBSK (Scale for children's perceived competences) shows low scores in the domains school performance, social acceptation, behaviour, and self esteem. From the DVK (Depression questionnaire for children) it appears that Brian thinks negatively of himself, his environment, and the future. Mother and father are shocked by the test results. They hope Brian learns to behave better at home and at school. They also hope he will feel better soon. Brian says he doesn't know, when he is asked what he wants.

Appendix B

General Coding. Causal strength ratings were re-coded following the proposal by Haynes and Williams (2003). Weak relations were set to .2, moderate relations to .5 and strong relations to .8. The most important problem was assigned a value of 8, the second most important problems a value of 4, the least important problems a value of 2, and variables not representing problems a value of 1. Modifiability ratings were re-coded as 1 for easily modifiable factors, .5 for moderately modifiable factors, and .1 for factors not being modifiable. All calculations described hereafter were performed for each causal model generated by participants. Formulae were programmed into an Excel file.

Calculation of predictions for Maximum Impact Heuristic. The basic idea of this heuristic is that people think of the impact each variable has on the client's situation. Interventions are chosen that influence the variables with the highest impacts. As the basic idea was adopted from Haynes and Williams (2003), the calculation also followed their proposed method.

First, the causal influence of each variable upon every other variable was computed by multiplying the strength of the causal relations connecting each cause to its effects, that is the causal strength relating variable V_i to variable V_j via m intermediate variables V_k was computed as $c_{ij} = c_{ik1} * c_{k1k2} * \ldots * c_{kmj}$. As causal relations were assigned values <1, the causal influence of variable V_i diminished the further apart cause and effect were. If a variable V_l was affected by two or more other variables V_k, the causal influence of these variables on their joint effect was calculated using the so-called noisy-or scheme, $c_l = 1 - \prod_{k=1}^{n}(1 - c_{kl})$ (Jensen, 1996). This rule ensures that causal influences do not sum up to values larger than 1, even if several variables strongly affect the same variable.

Next the impact of each variable V_i upon all other variables $V_{j \neq i}$ was computed by multiplying the importance of each variable (imp_j) with its modifiability (mod_j), and the strength of the causal relation relating variable V_i to all its effects c_{ij}, $impact_i = imp_i * mod_i + \sum_{i}^{f=n} imp_f * mod_f * c_{if}$. Note that the importance and modifiability of variable V_i is part of its impact. This captures the notion that the impact of a variable is its overall relevance for the current situation of the client.

Finally the causal impact of an intervention t was computed by summing up the causal impacts of the variables directly targeted by the intervention multiplied by the intervention's causal influence upon these variables, $impacttreatment_t = \sum_{i=1}^{n} c_{ti} * impact_i$. Based on the variable *impacttreatment* treatments T_1 to T_5 were rank ordered.

Calculation of predictions for Maximum Symptom Reduction Heuristic. The basic idea underlying this heuristic is that interventions reduce the presence of variables, which are either problems or causes of these problems. Two assumptions were made (i) each variable in the model was present and therefore received a starting value of $v_i = 1$, (ii) strong interventions reduce the presence of a variable that is directly affected by 80%, medium interventions by 50% and weak interventions by 20%. Assumption 2 was implemented by calculating the value of a variable V_i conditional on an intervention T with the causal strength c_{ti} as $v_{i|t} = v_i * (1 - c_{ti}) = 1 - c_{ti}$. This formula also implicitly assumes that an

intervention screens off the causal influences of other variables upon the variable being targeted by the intervention. This assumption is commonly made to model interventions in causal systems (see Pearl, 2000).

The causal influence of each intervention was calculated in a stepwise procedure. First the effect on the immediately affected variables was calculated using the formula above. Second the influence of interventions upon the variables that are direct effects of the variables directly affected were calculated, then the influence of the intervention upon the effects of these effects, and so on. The causal influence of an intervention T on variable V_j via variable V_i was computed as $v_{j|t} = v_{i|t} * c_{ij} = (1 - c_{ti}) * c_{ij}$. As V_j initially had a value of 1, the reduction is calculated by $reduction_j = 1 - v_{j|t}$. If variable j had several causes V_i, the reduced value of V_j was calculated by $reduction_j = 1 - \left[\sum_{i=1}^{i=n} v_{i|t} * c_{ij} / \sum_{i=1}^{i=n} v_i * c_{ij}\right]$. This formula computes the reduction of symptoms assuming that all other variables affecting V_j are still present and exert their influence. Note that variables not targeted by the intervention keep their initial value of $v_i = 1$ and therefore their complete influence. The formula also captures the intuition that if there are many alternative causes of a problem not targeted by an intervention, the intervention has only little influence on the problem.

Finally the reductions of all variables V_i directly or indirectly affected by an intervention T were summed, $\sum_{i=1}^{n} reduction_i$. Based on this sum treatments T_1 to T_5 were rank ordered. The intervention with the highest reduction received rank 1.

Chapter 5

Experience and the Causal Diversity Effect in Clinical Psychology

Abstract

Laypeople have been found to show a "causal diversity effect" in their reasoning about medical diseases: they perceive symptoms which emerge from a root cause via divergent causal paths as providing stronger evidence for the presence of a medical disease than symptoms which emerge from the root cause along a similar path (Kim & Keil, 2003). We investigated whether the causal structure of knowledge similarly influences experienced clinicians' diagnostic reasoning. In Study 1, we presented experienced clinical psychologists with recent theories for different disorders about how symptoms stem from a root cause. We asked them to choose which of two clients was most likely to have a disorder: a client with causally diverse or causally similar symptoms. The clinicians showed a causal diversity effect when they reasoned with the theories, but only when they could first use their own knowledge in a similar task. In Study 2, we elicited experienced clinical psychologists' causal assumptions about how symptoms stem from a root cause for different disorders. The clinicians didn't think clients with symptoms that were causally diverse according to their own assumptions were more likely to have a disorder than clients with causally similar symptoms. In Study 3 we elicited novices' causal assumptions. We found that novices did show a causal diversity effect. Causal diversity seems useful to infer the likelihood of a disorder for novices, while experienced clinical psychologists may use their knowledge of conditional probabilities obtained by experience to do so.

Clinicians often have to classify clients' symptoms and problems into a disorder category specified in the Diagnostic and Statistical Manual of Mental Disorders (APA, 2000). Kim and Ahn (2002) investigated such classifications. With a theory-drawing task they elicited clinicians' assumptions of how symptoms are causally related in different disorders. They found that experienced clinical psychologists and novices weight symptoms which receive equal weights in the DSM-IV differently, depending on the causal status of the symptoms in their theories. Thus, in a hypothetical client rating task clinicians were more likely to classify clients as having a hypothesized disorder when they had symptoms which were more causally central in their theories, in the sense that these symptoms caused more other symptoms. For example, in clinicians' theories for Anorexia Nervosa the symptom "refuses to maintain weight' causes several other symptoms of Anorexia Nervosa, while "absence of the period for more than 3 months" does not cause any other symptoms. The first more causally central symptom appeared more important for classifying a hypothetical client as having Anorexia Nervosa than the latter.

The studies we present here investigate whether the causal structure of knowledge may also influence clinicians' classifications in another way. Specifically, we aimed to test whether experienced clinical psychologists show a *causal diversity effect* in their diagnostic reasoning. Undergraduate psychology students have been found to show this effect when reasoning about medical diseases. They judge that patients are more likely to have a disease when they have symptoms that are causally diverse, that is: symptoms stemming from a root cause via divergent causal paths, than when they have symptoms that stem from the root cause via the same causal path (Kim & Keil, 2003).

Kim and Keil (2003) had presented undergraduate psychology students with information about the causal origins of four symptoms, for different medical diseases. The symptoms originated from a single common root cause, following a causal hierarchical structure, see Figure 1 for an example.

This figure illustrates how a root cause branches out in two causal paths, each again leading to two symptoms, of the disease "rheumatic fever." The undergraduates judged which of two patients was most likely to have a specific disease: a patient presenting with two symptoms originating from the same causal path ("causally proximal" symptoms), or a patient presenting with two symptoms originating from different paths ("causally diverse" symptoms). In the example of rheumatic fever, a patient presenting with joint inflammation and rapid heartbeat presents with proximal symptoms, and a patient with joint inflammation and skin rash around the neck presents with diverse symptoms. Most undergraduates judged patients with causally diverse symptoms as most likely to have the hypothesized disease. Also, undergraduates preferred to seek out causally diverse symptoms rather than causally proximal symptoms, in order to test whether a patient has a hypothesized disease. They did so even though they thought patients are more likely to have causally proximal symptoms than causally diverse symptoms (Kim, Yopchick, & De Kwaadsteniet, 2008).

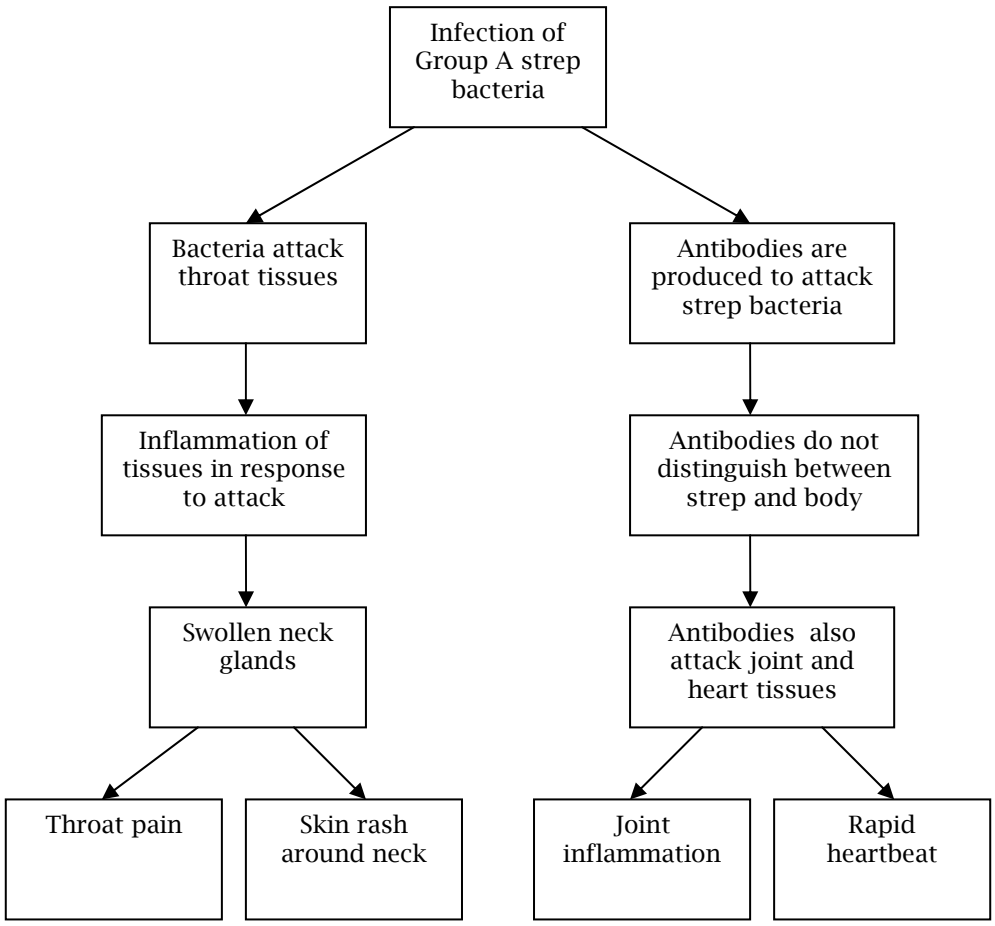


Figure 1. Causal diagram of rheumatic fever stimuli used by Kim & Keil (2003).

This causal diversity effect seems analogous to the “premise diversity effect” which occurs in novices’ category-based induction. In category-based induction tasks people are asked to infer properties of categories, using information about properties that other categories have. Undergraduates rate arguments consisting of premises leading to a conclusion as stronger if the categories in the premises to which the to-be inferred property is ascribed are more diverse. For example, Osherson, Smith, Wilkie, López, and Shafir (1990) found that most undergraduates rate an argument with premises ”Hippopotamuses have a higher sodium concentration in their blood than humans” and ”Hamsters have a higher sodium concentration in their blood than humans”, and conclusion ”All mammals have a higher sodium concentration in their blood than humans” as stronger than an argument having the same conclusion but premises ”Hippopotamuses have a higher sodium concentration in their blood than humans” and ”Rhinoceroses have a higher sodium concentration in their blood than humans”. Hippopotamuses and hamsters are perceived as more dissimilar or diverse than hippopotamuses and rhinoceroses, and therefore when they share a property that is stronger evidence for the conclusion that all mammals share that property (Osherson et al., 1990). Not only do undergraduates rate more diverse evidence stronger in category-based inferences, they prefer to seek out diverse information to test a

category-based inference too (López, 1995; see also Kincannon & Spellman, 2003; Spellman, López, & Smith, 1999).

More diverse evidence seems intuitively more convincing in category-based induction and causal diagnostic reasoning, and this has also been argued to be normative (see e.g. Heit, Hahn, & Feeney, 2005; Popper, 1963; Whewell, 1840/1999). Meehl (1954) argued that clinicians should try to find diverse evidence to corroborate their hypotheses: "Having once *conceived* a particular hypothesis concerning a patient, we must, if we are scientific (...), subject this hypothesis to the usual canons of inference. That is, we must see whether the hypothesis will entail (…) a greater range of or diversity of the known facts (…), and so on" (1954, p. 66).

People prefer explanations that are as simple as possible, and that account for as many data as possible (Read & Marcus-Newhall, 1993; Thagard, 1992). For causally diverse evidence the most simple, parsimonious explanation that unifies the observations is that the hypothesized root cause is present. For causally proximal evidence, the most economic explanation seems that the immediately preceding cause is present (see Figure 1: the cause on the last level but one, preceding two proximal symptoms) (Kim et al., 2008). In a similar vein, people may sooner discount alternative causes if diverse evidence is present, leaving the hypothesized root cause more likely, while it is easier to imagine alternative causes than the hypothesized root cause if proximal evidence is present (Kim & Keil, 2003).

Following Horwich (1982) and Kincannon and Spellman (2003), who gave Bayesian justifications for why diverse premises form stronger evidence in category-based induction, the reason why causally diverse symptoms form stronger evidence in diagnostic reasoning can also be formalized in a Bayesian way: The chance that your hypothesis is correct, given specific evidence - the chance that a specific root cause is present, given proximal/diverse evidence – is equal to the probability that the evidence occurs given the hypothesized cause ($P(e \mid c_1)$) multiplied by the probability of the hypothesized cause, independent of any evidence ($P(c_1)$), controlling for the prior probability that the evidence occurs, that is: the summed probabilities that the evidence occurs with all possible causes (hypothesized root cause and alternative causes). See (1):

$$(1) \qquad P(c_1 \mid e) = \frac{P(e \mid c_1)\, P(c_1)}{\Sigma c_{1\text{-}n}\, P(e \mid c)\, P(c)}$$

In the causal hierarchies used by Kim and Keil (2003) and Kim et al. (2008) the possible causes for proximal evidence may include the possible causes for the presence of diverse evidence, and more. Thus, the denominator in the equation gets larger for proximal evidence compared to diverse evidence, leading to a lower probability that with proximal evidence the hypothesized root cause is present. With fewer alternative causes, as with diverse evidence, the denominator gets smaller, leading to a higher probability of the hypothesized root cause.

Kim and Keil (2003) speculated that if experts have developed more pronounced causal theories "gaining expertise would also give rise to increasingly powerful diversity effects" (2003, p.165). If this is true, experienced clinical psychologists will show a stronger causal diversity effect in diagnostic reasoning than novices.

However, it is also possible that causal diversity effects will not occur in an expert sample. With categorical induction domain experts and undergraduate students do not display diversity effects equally, which shows that the availability of relevant background knowledge can make a big difference (Medin, Coley, Storms, & Hayes, 2003). While American undergraduates preferred more diverse bird categories to make general inductions about birds, US bird experts did not show this preference, but often seem to consider ecological relations that exist between birds of different categories (Bailenson, Shum, Atran, Medin, & Coley, 2002). Itza' Maja, who can be seen as expert biologists, also showed no diversity preference for category-based inferences about birds, nor about mammals, or palms, contrary to American undergraduate students. They often considered ecological relations too (Bailenson et al., 2002; López, Atran, Coley, Medin, & Smith, 1997). Similarly, tree experts showed no diversity effect when making category-based inductions of diseases affecting trees (Proffitt, Coley & Medin, 2000). A diversity-based strategy in categorical induction might thus be a default strategy that novices often use, and other strategies may be adopted when relevant background knowledge is available. Similarly, experts may not always show a preference for causally diverse evidence, but their relevant knowledge might trigger other reasoning strategies. We can imagine that expert clinicians might use their knowledge of the typicality or diagnosticity of symptoms (Klayman and Brown, 1993), or cue validity of symptoms (Sloman, Love, & Ahn, 1998), to make diagnostic decisions rather than knowledge of causal structure.

We tested whether experienced clinical psychologists show a causal diversity effect in their diagnostic reasoning. In our first study, with a design similar to the first study by Kim and Keil (2003), we presented experienced clinical psychologists with causal hierarchies for psychiatric disorders that explained how according to recent literature associated symptoms originate from a root cause. Then we asked them to choose which of two clients would most likely have a hypothesized disorder according to these theories: a client presenting with two causally proximal symptoms or a client presenting with two causally diverse symptoms.

In Study 2, we investigated whether a causal diversity effect would occur when clinical psychologists reasoned using their own causal theories of disorders. In daily practice, clinical psychologists will reason using their own theories (Kim & Ahn, 2002). We tried to extract clinicians' causal assumptions about how symptoms emerge in different disorders, and tested whether clinicians showed a preference for information that is causally diverse according to these assumptions. Study 3 was similar to Study 2, except that now novices in clinical psychology participated.

Study 1

Method

Participants. Participants were 19 clinical psychologists from the US and the Netherlands, recruited by phone and e-mail. We randomly assigned participants to condition 1 (n = 10, 5 American and 5 Dutch psychologists), or condition 2 (n = 9, 4 American and 5 Dutch). The American and Dutch samples did not differ significantly in experience, age, or gender. Mean experience of the whole sample was 19.6 years (SE = 2.4), 11 were female,

mean age was 47.5 (*SE* = 3.0). American participants were paid 40 dollars and Dutch participants received a 10 euro gift certificate for their participation.

Materials and procedure. We derived summary causal diagrams for five DSM-IV-TR (APA, 2000) psychiatric disorders from recently published theories. Each diagram had an underlying structure similar to the causal diagrams used by Kim & Keil (2003): a root cause branched out in two divergent causal paths, each containing two intermediate steps, leading to two pairs of symptoms. Although we did not intend to present complete scientific theories, as these usually consist of more than two causal pathways and contain more than four symptoms, we intended to correctly represent parts of the theories. In Figure 2 a sample diagram is presented, for autism, which we based on Belmonte, Allen, et al. (2004) and Belmonte, Cook, et al. (2004). Other psychiatric disorders for which we constructed causal diagrams were: anorexia nervosa (diagram based on Fairburn, Cooper, & Shafran, 2003, and Fairburn, Shafran, & Cooper, 1999), ADHD (based on Sonuga-Barke, 2002), panic disorder (based on Ninan & Dunlop 2005), and Borderline personality disorder (based on Mauchnik, Schmahl, & Bohus, 2005). In Appendix A the components of all diagrams are presented.

We discussed the plausibility of all diagrams with experienced clinical psychologists. From the diagrams we created sets of four separate chains, each showing how one symptom results from the root cause, via one of the branches with two intermediate steps. We used these sets of four chains per disorder instead of the summary diagrams to eliminate spatial distance cues (cf. Kim & Keil, 2003).

We sent two questionnaires to each participant, with one week in between. In one questionnaire, containing the causal task, participants were shown the sets of four chains for the disorders. Each chain was labelled with the name of the disorder and a reference to the publications used for the construction. The four chains for one disorder were presented in pseudo-randomized order on separate sheets, such that the same causal chain was never seen twice consecutively. After having read the four chains for the first disorder, participants were asked to write a summary paragraph of the information.

Next, they were asked to suppose that there were two clients: one client presenting with two symptoms that were shown to result from the same causal path, proximal symptoms, and the other client presenting with two symptoms that were shown to result from different causal paths, diverse symptoms. The two clients had one symptom in common. For example, for the disorder autism participants were presented with the statements: "*Client N.W. has impairment in social interactions and problems with communication*" and "*Client C.F. has a need for structure and rules and problems with communication*". As can be seen from Figure 2, symptoms of Client N.W. are proximal, and symptoms of Client C. F. are diverse. In order to ensure that participants would not simply discount the presented theories when they did not agree with their content, we asked them to determine which of two clients would be most likely, *according to the theory they had read*, to have the disorder. Continuing our example, we asked: "*Which client will be more likely, according to Belmonte et al. (2004a, 2004b), to have autism? Please choose one.* ". To avoid that some choices were far more plausible for reasons other than diversity (e.g. prevalence or salience in practice, mirroring of DSM-IV-TR criteria), we discussed the sets of

symptoms used in the questions with experienced clinical psychologists. We further asked participants to explain their choice, and to indicate whether they had ever heard of, read about, or studied the specific theory before.

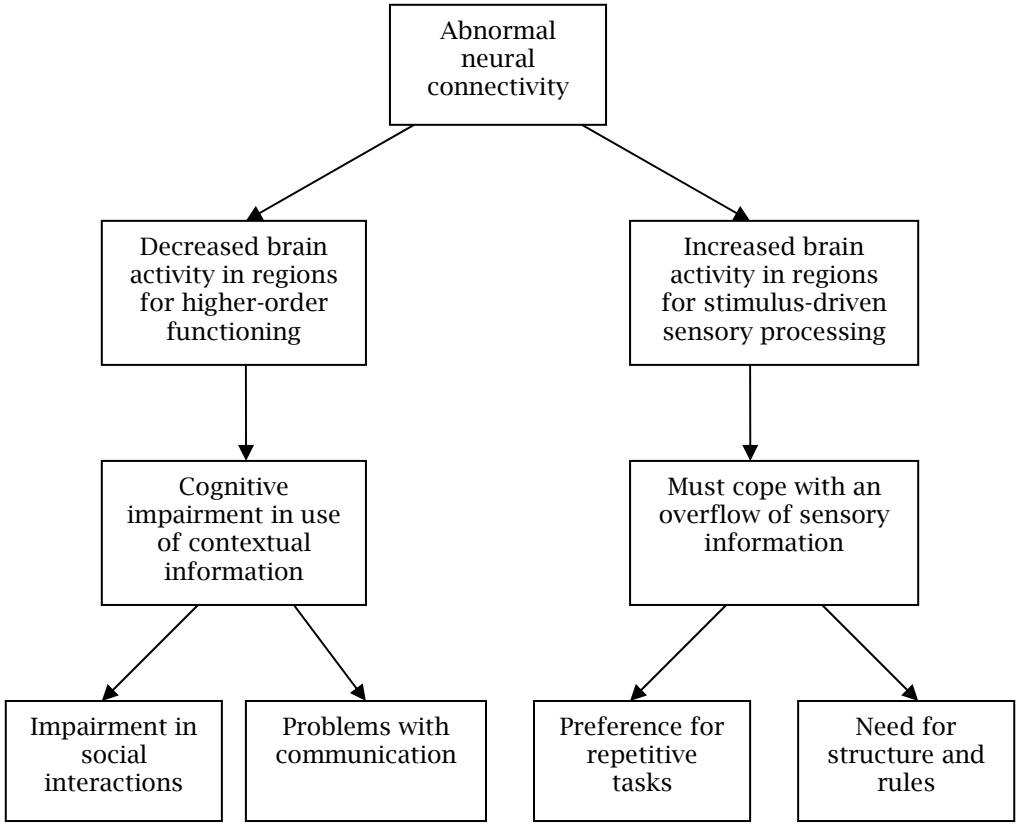


Figure 2. Causal diagram of autism stimuli used in Study 1.

The same procedure was followed for the other four disorders, except that for these disorders we did not ask to provide a written summary, but we asked participants to mentally summarize the information from the chains. We counterbalanced the sequence of disorders between participants and we counterbalanced which client, with proximal or diverse symptoms, was presented first. At the end participants filled in their age, gender, and experience, and we gave them complete references for the theories we had presented.

The other questionnaire contained a control task. We reasoned that experienced clinical psychologists would have developed strong opinions about psychiatric disorders (as in Kim & Ahn, 2002), with which we could compare their choices made after causal information was presented. In the control task, participants were presented with the same client statements, for the same disorders, and in the same sequence as in the causal task, but now we asked them to judge which of two clients would be most likely, in their own opinion, to have a disorder. For example, after presenting the same two client statements about *"Client N.W."* and *"Client C.F."*, we asked: "*Which client do you think is more likely to have autism? Please choose one*".

Participants in condition 1 completed the control task first. Participants in condition 2 completed the causal task first. For Dutch participants, all materials were translated in Dutch.

Results. A choice for the client presenting with diverse symptoms was coded as "1" and a choice for the client with proximal symptoms as "0". We collapsed the binary data across the five disorders for analysis. Results are shown in Figure 3.

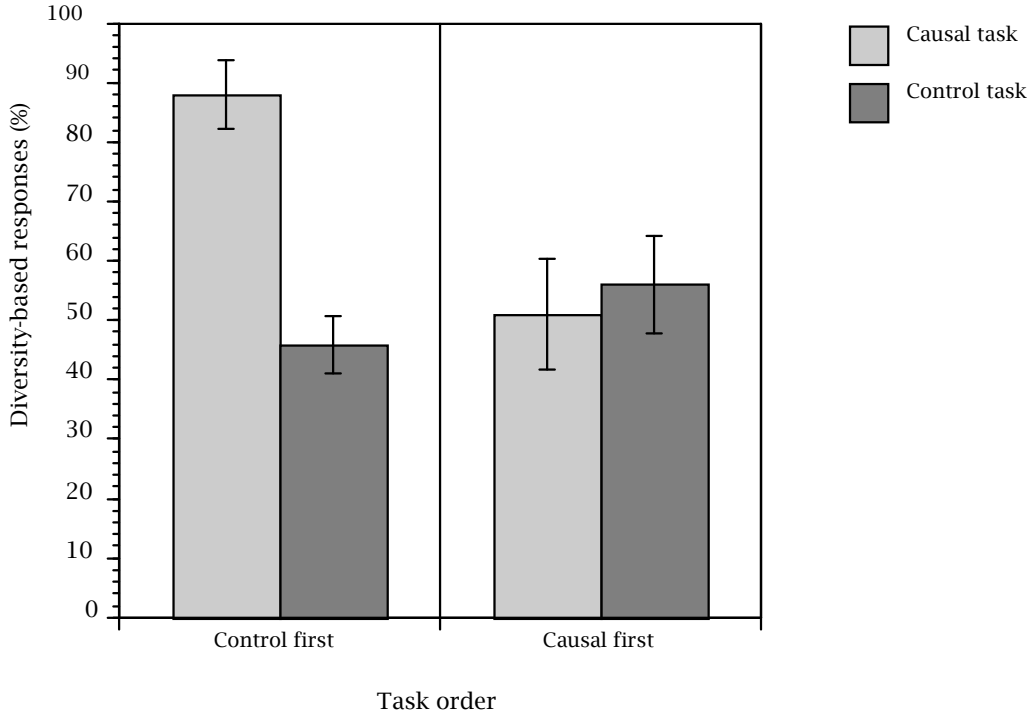


Figure 3. Results Study 1

We conducted all analyses at the $\alpha = .05$ level. A 2 within subjects (Task: control, causal) X 2 between subjects (Condition: control first, causal first) ANOVA revealed an interaction ($F(1,17) = 23.5$; $p < .001$; $\eta^2 = .58$), and a main effect of task ($F(1,17) = 15.4$; $p < .01$; $\eta^2 = .48$). Paired sample t-tests showed that participants who completed the control task first, chose the client presenting with diverse symptoms significantly more often in the causal task (M = 88%; SE = 5.7), than in the control task (M = 46%; SE = 4.8) ($t(9) = -5.2$; $p < .01$; Cohen's d = 2.0). The results were in the same direction for all five disorders, indicating that this difference was not caused by any single item. Participants who completed the causal task first did not more often prefer the client with diverse symptoms in the causal task (M = 51%; SE = 9.4) than in the control task (M = 56%; SE = 8.3) ($t(8) = 1.0$; $p > .3$).

Further, independent sample t-tests revealed no difference between frequency of diverse choices in tasks completed first, that is: between frequency of diverse choices in the control task for participants in the control-first condition and frequency of diverse choices in the causal task for participants in the causal-first condition ($t(12) = -0.3$; $p > .7$ (equal variances not assumed)). Also, frequency of diverse choices in the control task did not differ between conditions ($t(13) = -.7$; $p > .4$ (equal variances not assumed)). Frequency of diverse choices in the causal task did differ between conditions, with more diverse choices in the

control-first condition ($t(17) = 2.4$; $p < .05$; Cohen's $d = .97$). We found no differences between American and Dutch participants.

Justifications. Each justification that participants gave in the causal task was assigned to one of five coding categories by two coders, who were blind for the choices made and the conditions. We defined two categories for explanations in which participants referred to causal structure: Category A was for explanations according to which proximal symptoms provide the strongest evidence for the presence of the disorder, and category B was for explanations according to which diverse symptoms provide the strongest evidence for the presence of the disorder. Further, we defined category C for explanations referring to typicality or importance of symptoms. Category D was for explanations referring to distinction of the disorder from alternative disorders. Finally, category E was for explanations that did not fit any of the other categories. Cohen's kappa for interrater reliability was .73; differences were discussed until coders agreed.

For participants who did the control task first, 80% of the justifications in the causal task referred to causal structure. Of these all but two justifications referred to diverse symptoms as strongest evidence and explained a diverse choice. For participants who did the causal task first, 60% of the justifications referred to causal structure. Of these, 59% referred to diverse symptoms as strongest evidence and explained a diverse choice; the others referred to proximal choices, all but one to explain a proximal choice. In both conditions the remaining justifications most often referred to category C: typicality or importance of symptoms, to explain both diverse and proximal choices. Also, justifications assigned to categories D and E explained both diverse and proximal choices. So, most participants who chose the client with causally diverse symptoms appeared to do so because they considered causally diverse evidence to be stronger evidence for the presence of a disorder.

Discussion. We found a causal diversity effect with experienced clinical psychologists, similar to the causal diversity effect found with undergraduate psychology students reasoning about medical conditions (Kim & Keil, 2003). Participants who completed the control task first thought that in the causal task according to the presented theories a client with diverse symptoms was more likely to have the hypothesized psychiatric disorder than a client with proximal symptoms. In the control task no such effect was found. Although people may not always be conscious of their real reasons to choose a specific answer (Nisbett & Wilson, 1977), participants' justifications indicated that choices were indeed based on causal structure. Surprisingly, we found no diversity effect for participants who completed the causal task first. Also, we found no difference in frequency of diverse choices between tasks performed first.

Possibly, participants who completed the causal task first might have reasoned less according to the theories we presented, although they were instructed to do so, and made more use of their own insights, than participants who started with the control task and thus could express their own opinions first. To explore this explanation, we sent questionnaires to five more Dutch clinical psychologists (mean experience = 26.8 years, mean age = 50.2 years, two were female), with first the causal task and the control task one week later. We stressed in the causal task that participants could express their own opinions in the second

(control) task, by each time adding after the question which client was more likely to have a disorder according to the theory presented, this sentence between brackets: "In the second questionnaire we will ask for your personal opinion". These five participants chose the diverse client in 71 % of the cases in the causal task, and in 44 % of the cases in the control task. Of the justifications that these participants gave for their diverse choices in the causal task 89% referred to causal structure. Although the percentage of causal choices is smaller than it was for participants who did the control task first, the results are in the same direction, supporting our explanation.

Our instruction to reason according to presented theories constrains our findings, because in a sense the experienced clinical psychologists were novices. For all disorders, most participants answered that they had never heard about, read about, or studied the specific theory before. Results of participants who did the causal task first indicate that background knowledge may influence their diagnostic reasoning strategies. Therefore we investigated, in Study 2, whether experienced clinical psychologists would show a preference for evidence that was causally diverse in their own causal assumptions about a disorder.

Study 2

Method

Participants. Participants were 17 clinical psychologists recruited by phone and e-mail. Of these, 4 were American and 13 Dutch. Their mean experience was 22.4 years (*SE*=3.0), 7 were female, and their mean age was 48.6 (*SE*=2.6). American clinical psychologists were paid 30 dollars and Dutch clinical psychologists received a 10 euro gift certificate for their participation.

Materials and procedure. Participants received two tasks by mail, with a time lag of at least one week. In the first task we asked them to show how they thought four symptoms originated from a root cause via two divergent causal paths, for five psychiatric DSM-IV-TR disorders (APA, 2000). For each disorder, we asked them first to choose four symptoms from a list of 6 to 9 symptoms of that disorder according to DSM-IV-TR criteria (APA, 2000), and place these four symptoms in the bottom four boxes of an empty diagram. The diagrams had the same hierarchical structure as those used in Studies 1, see Figure 2. Next, we instructed participants to fill in the other boxes in the diagram explaining how, in their opinion, these four symptoms originate from a root cause, following the two paths in the diagram. They should take care that symptoms could not be easily switched between paths, meaning that symptoms explained with one pathway should not be equally plausibly be explained with the other pathway. As an example we showed a hierarchical diagram of Kim and Keil (2003) for radiation sickness. After having completed diagrams for the five disorders, we asked participants to rate for each disorder, on 7 point scales, how difficult it was to fill in the diagrams and how well they thought the diagrams represented how they thought the four symptoms emerged.

We picked the five disorders from the DSM-IV-TR (2000) based on the presence of lists with equally-weighted criteria: Major Depressive Episode, General Anxiety Disorder, Borderline Personality Disorder, Obsessive-Compulsive Personality Disorder and Schizotypical Personality Disorder. To adhere to the criteria of Major Depressive Episode at

least one of the symptoms 'depressed mood' or 'loss of interest/pleasure' should be present. We excluded these two symptoms from our list. For GAD we used the symptoms enumerated under criterion C: symptoms that are associated with excessive anxiety and worry. We excluded the symptoms excessive anxiety and worry that are separately mentioned in criterion A in the DMS-IV-TR. By letting participants choose between symptoms that in the DSM-IV-TR receive equal weights, we avoided that participants placed symptoms in separate paths in their diagrams based on a differential weighting in the DSM-IV-TR. Symptom lists from which participants could choose for each disorder are presented in Appendix B. For Dutch participants all materials were translated, using a DSM-IV-TR translation (Koster van Groos, 2003).

Based on their completed diagrams, we constructed a set of questions constituting the second task for each participant. For each disorder, we presented two clients: one client with two causally proximal symptoms and one client with two causally diverse symptoms. We asked participants to choose which of the two clients they thought was most likely to have the disorder, and to explain their choice in one sentence. We constructed the question sets with pre-defined symptom pairs, based on the position of symptoms in the causal hierarchies, using table 1.

Table 1

Pre-defined Symptom Sets used in Study 2 and Study 3

	Disorder				
	BPD	Depression	GAD	OCD	SPD
Diverse	S2-S4	S1-S3	S3-S1	S4-S1	S3-S2
Proximal	S2-S1	S1-S2	S3-S4	S4-S3	S3-S4

For example, for Borderline Personality Disorder, we presented each participant with client statements: "*Client E.V. shows S2 and S4*" and "*Client H.L. shows S2 and S1*", and asked *"Which client do you think is more likely to have Borderline Personality Disorder?"*. For S1 – S4 we filled in the symptoms that the participant had placed in the bottom boxes, with S1 being the symptom placed in the box most left and S4 the symptom in the box most right.

At the end, we asked participants for their experience, gender, and age. We sent the second task one week after we had received answers to the first task. We counterbalanced the order of client statements and options presented first, and randomized the sequence of disorders. The order in which the disorders were presented was the same in both tasks.

Results. A choice for the client presenting with diverse symptoms was coded as "1" and a choice for the client with proximal symptoms as "0". We collapsed the binary data across the five disorders for analysis. We conducted all analyses at the $\alpha = .05$ level. Participants chose the client with the diverse symptoms (with diversity based on the participants' diagrams) in 44% of the cases ($SE = 5.6$). The percentage of diverse choices did not significantly deviate from 50% ($t(80) = -1.0$; $p > .3$).

Because in the second task (the client choices) we did not remind participants of their diagrams, we did not expect participants to refer to causal structure in their justifications.

We analysed participants' justifications to see whether their choices were guesses. This may be expected if participants wanted to choose based on DSM-IV-TR-criteria, because symptom pairs of the hypothetical clients have equal weight in the DSM-IV-TR. Only two times did a participant explain the choice to be a pure guess. The lion's share of justifications referred to how discriminating symptoms were for a specific disorder.

It appeared difficult to fill in the diagrams: averaged over participants and disorders difficulty ratings were 4.5 (SE = 0.2) on a scale of 1 (not at all difficult) to 7 (very difficult), which deviated significantly from the midpoint 4 ($t(83) = 2.7$; $p < .01$). Participants will not always have been able to fit their causal beliefs into the fixed structures. Therefore we reanalyzed the data, now looking only at those choices for disorders for which participants had indicated that their diagrams were representative for how they thought. We assumed diagrams to be representative if participants had given them representativeness ratings of at least 4, the midpoint of the scale (with 1 meaning not at all representative and 7 highly representative). Two participants had not given any ratings higher than 3, the remaining 15 participants had rated 52 diagrams as representative. Results for representative diagrams were similar: in 40% of the cases the diverse client was chosen (SE = 6.9), which does not significantly deviate from chance level ($t(51) = -1.4$; $p > .1$).

Discussion. Participants didn't think that clients with symptoms that were causally diverse according to their own assumptions were more likely to have a disorder than clients with symptoms that were causally proximal. In Kim and Keil's study (2003) and in Study 1 the chains explaining how symptoms originate from a root cause were available to participants when they answered the diagnostic questions. Here there was a time lag of more than one week between the diagram drawing task and the questionnaire, and participants were in no specific ways reminded of their diagrams in the questionnaire. If their causal explanations as given in the diagrams were ad hoc constructions, they might have been forgotten after some time. Kim and Ahn (2002) however found implicit theories about how symptoms are causally related in DSM-IV-disorders to influence diagnostic decisions of experienced clinicians and novices two weeks after drawing these theories. So why did the causal diversity of symptoms in participants' theories not influence their choices here?

If clinicians relied on their knowledge of the DSM when making choices, a similar number of diverse choices could be expected, because the symptoms that clients presented with all have equal weight in the DSM-IV-TR. Participants didn't explain their choices as pure guesses however, but they thought symptoms were differently discriminating. This is in line with other studies showing that clinicians weight DSM-criteria differently, while in the DSM they are equally weighted (Evans, Herbert, Nelson-Gray, & Gaudiano, 2002; Kim & Ahn, 2002).

In the introduction we suggested that a preference for causally diverse evidence might be related to not having expertise in a domain. Results in Study 1 indicate that experienced clinical psychologists do perceive differential value of causally diverse versus proximal information, as they more often chose the client with diverse symptoms than the client with proximal symptoms in the causal task. However, the theories presented were mostly new to participants. Background knowledge appeared to interfere with a causal diversity preference, since we only found this preference when participants could express their personal opinions first, or knew they could later. In this second study, in which we

expected participants to refer to their own background knowledge, we found no causal diversity effect.

Domain novices may have a preference for causally diverse evidence because this allows them to make inferences about the probability of a root cause, when they have no further information or knowledge about conditional probabilities (cf. Horwich, 1982; Kim et al., 2008; Kincannon & Spellman, 2003). Experts may have other strategies available to infer the probability of a root cause given specific evidence (cf. Medin et al., 2003; Murphy, 2002). Specifically, experienced clinical psychologists may infer the probability of a disorder from knowledge obtained by experience about how likely it is that a client has a disorder given the presence of specific symptoms.

To see whether novices in clinical psychology would show a causal diversity effect, we replicated Study 2 with a group of beginning clinical psychologists. While participants in Study 2 had at least 10 years of experience, the novices sample in Study 3 had at most 3 years of experience. These novices will have domain knowledge obtained from education, so they should be familiar with the disorders and their symptoms. Kim and Ahn (2002) found that novices, and even lay people, have causal theories for disorders, which have elements in common with experts' theories. However, novices haven't seen many clients in practice, so we don't expect them to have obtained much experiential knowledge about how likely it is that a client has a disorder given that specific symptoms are present. Thus, we expected that these novices would judge clients with symptoms that are causally diverse according to their assumptions as more likely to have a disorder than clients with causally proximal symptoms.

Study 3

Method

Participants. Participants (N=14) were students in clinical psychology who were finishing their masters and had followed internships for several months, and clinical psychologists who had less than three years of experience. Together they had a mean experience of 1.0 years (SE=0.3), 10 were female, mean age was 24.7 (SE=0.5). Participants were recruited at the department of the author, and by phone and e-mail. All participants were Dutch. Students received 5 euros for their participation; clinical psychologists received 10 euro gift certificates.

Materials and procedure. The materials and the procedure were the same as in Study 2, except that students did the second task at the author's office, after which they received their rewards. Others did both tasks by mail.

Results. Participants chose the client with the diverse symptoms, with diversity based on the participants' own diagrams, in 61% of the cases (SE = 6.1). Results were similar for all five disorders. Participants' diverse choices did not significantly deviate from 50% ($t(63)$ = 1.8; p = .08).

Next, we looked only at the choices for those disorders for which participants had given representativeness ratings of at least 4. For 4 participants ratings were not higher than 3. The remaining novices rated 31 of their diagrams as representative. For these disorders, in

69% of the cases the diverse client was chosen (*SE* = 8.7), which does significantly deviate from 50% ($t(28) = 2.2$; $p < .05$).

Like the experienced clinical psychologists, the novices rarely explained their choices as being guesses. Justifications again referred mostly to how discriminating symptoms were for the hypothesized disorders.

Experts versus novices. Comparing experienced clinical psychologists with novices with a t-test for independent samples showed a significant effect of expertise when all diagrams are taken into account, regardless of representativeness: $t(143) = -2.0$; $p < .05$; Cohen's $d = .33$). Thus, the novices more often than experienced clinical psychologists judged clients with diverse symptoms, according to their own diagrams, to be most likely to have a disorder. Looking only at representative diagrams, with ratings of at least 4, we found an even stronger effect of expertise ($t(79) = -2.5$; $p < .05$; Cohen's $d = .57$).

Discussion. For representative diagrams, novices more often thought that clients who had symptoms that were causally diverse were more likely to have a disorder, than clients with causally proximal symptoms. Furthermore, they chose clients with causally diverse symptoms more often than experienced clinical psychologists did. Thus, novices showed a causal diversity effect. Relying on causal diversity might be a useful strategy for novices, because they do not have knowledge about conditional probabilities available as experts do, as we suggested above. Alternatively, they may have some statistical knowledge about conditional probabilities, but cannot easily apply this knowledge. In the domain of medical diagnostic reasoning it is assumed that experts not only have more experiential, associative knowledge, but also apply this knowledge more easily than advanced students, who will more often refer to their theoretical knowledge (see e.g. Norman, Eva, Brooks, & Hamstra, 2006; Van de Wiel, Boshuizen, & Schmidt, 2000).

General discussion

We aimed to find out whether experienced clinical psychologists would show a causal diversity effect in diagnostic reasoning tasks, like psychology undergraduates did when reasoning about medical conditions (Kim & Keil, 2003). In Study 1 we asked experienced clinical psychologists to decide, using recent theories about psychiatric conditions, which of two hypothetical clients was more likely to have a specific psychiatric condition. We found that participants more often chose the client with symptoms that were causally diverse than the client with symptoms with similar causal origins. However, they only did so when they could express their own opinions first, or when they knew they could later. Apparently participants' background knowledge interfered with the task.

In a sense the clinical psychologists in Study 1 were novices, since most had not heard about the theories presented before. Therefore, in Study 2 we elicited clinical psychologists' own causal assumptions about how symptoms originate from a root cause in a disorder, and tested whether they would show a preference for symptoms that were causally diverse rather than similar, according to their assumptions. They did not. In Study 3 novices in the same task did show a preference for symptoms that were causally diverse according to their own assumptions.

We asked clinical psychologists to fit their causal assumptions into the diagram structures, what might have been an unrealistic demand. However, with novices we did find a causal diversity effect for those diagrams that participants rated to be representative, while not with experienced clinical psychologists. Taken together, the results of the three studies indicate that relevant background knowledge may interfere with causal diversity. Specifically, we suggest that using information about causal diversity is a default strategy (cf. Kim et al., 2008), like relying on premise-diversity seems to be a default strategy in category-based reasoning (Medin et al., 2003; Murphy, 2002). This strategy may be useful when relevant knowledge about conditional probabilities of causes given evidence (see formula (1)) is lacking, or when it is not easily applied (see e.g. Van de Wiel et al., 2000). Experienced clinical psychologists will have experience-based knowledge about how likely clients are to have a disorder, given the presence of specific symptoms. They may use this knowledge about the diagnosticity of symptoms (Klayman & Brown, 1993) or cue validity (Sloman et al., 1998) to make inferences about the likelihood that a specific client has a specific disorder given a set of symptoms. The distinction between diversity and proximity in a causal structure is then no longer informative in this respect.

Justifications that both novices and experienced clinical psychologists gave referred mostly to the distinction of the hypothesized disorder from other disorders. We argued that novices will rely on causal structure, while experts will rely on their experiential knowledge of conditional probabilities. We had not expected participants to refer to causal structure in their justifications in Studies 2 and 3, because participants didn't have their diagrams available when answering the diagnostic questions. Reliance on causal diversity need not be a strategy that people are aware of. Perhaps, if we had enabled participants in Studies 2 and 3 to look at their diagrams when making their choices, as participants in Study 1 and in the studies by Kim and Keil (2003) and Kim et al. (2008) could look back at the causal information, participants would have given more diversity-based responses. We did not however, because not presenting the diagrams would be more similar to daily practice, in which clinicians will not have visualizations of their causal beliefs available when making decisions. It is therefore remarkable that we still found a causal diversity effect for novices in Study 3.

It is possible that novices based their choices not on causal structure diversity, but rather on feature diversity, choosing the clients with more dissimilar symptoms. We can also imagine that novices (and experts alike) constructed the causal diagrams using feature similarity: they may have proceeded by first choosing two dissimilar pairs of similar symptoms, and next thinking of different causes for these two pairs. Feature diversity and causal diversity need not necessarily be different concepts, however. Feature diversity might be a consequence of causal diversity, in the sense that divergent causal paths may lead to more different features while similar causal paths may lead to more similar features (cf. Murphy & Medin, 1985).

Our results are in line with those of Lo, Sides, Rozelle, and Osherson (2002) who concluded that people use probability information rather than information about diversity in category-based induction tasks. Lo et al. found that children and students adhere to the 'Premise Probability Principle', which states that arguments with less probable premises are

stronger, rather than to a diversity principle according to which arguments with more diverse premises are stronger. Our experienced participants may have reasoned using their knowledge about conditional probabilities rather than knowledge of the probability that symptoms occur, but this is not contradictory to the Premise Probability Principle. Symptoms that are more diagnostic for a disorder will occur less often in other disorders (see definition Klayman & Brown, 1993), and thus will be less probable since they have fewer alternative causes.

In this light, differences in beliefs about mental disorders between experienced clinicians and novices are interesting. Ahn, Flanagan, Marsh, and Sanislow (2006) found that novices seem to believe that mental disorders have a defining feature which causes symptoms associated with them, and that disorders will disappear if this defining feature is removed. Further, novices seem to believe that experts will have knowledge about what these defining features are. Experienced clinicians do not seem to believe these things. Thus, novices may assume there is one root cause for all symptoms associated with a disorder, while experienced clinicians may assume that disorders have different combined causes (Ahn et al., 2006). Also, experienced clinicians may see more different causes for symptoms than novices. Experienced clinicians' theories for different disorders may thus be overlapping, while novices' theories for different disorders may be more disjoint (see also Murphy & Wright, 1984). Such differences in knowledge representation may imply that experienced clinicians think of alternative causes for diverse symptoms sooner than novices, and that experienced clinicians prefer to look at symptoms which occur less often in other disorders when deciding whether a client has a specific disorder, while these symptoms may have similar causal origins.

We suggested that relying on causal diversity is a useful strategy for novices, and may no longer be used when more knowledge about conditional probabilities is available. This receives support in a new series of studies, in which we manipulate novices' knowledge. In the first of these studies, we presented novices with both causal structure information and information about conditional probabilities. We found that when novices are presented with both these kinds of information, they use statistical information rather than information about causal diversity to make diagnostic decisions (Kim & de Kwaadsteniet, in preparation). We finally note that knowledge about conditional probabilities and causal structure need not be separate, but will be integrated (cf. Rehder, 2007; Tenenbaum, Griffiths, & Niyogi, 2007).

To conclude, our results add to those showing that experts in a domain may use other reasoning strategies than novices in a domain (cf. Bailenson et al., 2002; López et al., 1997; Medin et al., 2003; Proffitt et al., 2000). A causal diversity effect might thus be specific to novices' diagnostic reasoning. Further research should establish whether experienced clinicians' strategies are indeed based on their experiential knowledge of conditional probabilities as we suggested, or on other strategies.

Acknowledgements

We are grateful to Ina van den Berckelaer-Onnes and Laurence Claes, and Nancy Snyder, the experts who helped us make balanced diagrams and client pairs for Study 1. We also thank our coders: Shane Lloyd, Daniel Paulus, and Chantal Berens. Further we want to thank colleagues who contributed to our studies with their comments.

Appendix A

Components of the Psychiatric Disorder Stimuli Used in Study 1

	Disorder			
Components	Anorexia Nervosa	ADHD	Panic Disorder	Borderline Personality Disorder
Root Cause	Too severe dietary restrictions due to an extreme need for self-control	Inhibitory dysfunction due to alterations in prefrontal control circuits	Specific trigger activates malfunctioning amygdala	Interaction of genetic predisposition and environmental stressors
Intermediate Causes 1	Breaking of (too) strict dietary rules; Perceived failure	Cognitive dysregulation; Poor planning and working memory functioning	Amygdala sends overly strong signal to lateral hypothalamus; Overactive sympathetic nervous symptom	Low serotonin activity; Lack of inhibitory influence over limbic system
Terminal Effects 1	Negative self-evaluation; Abuse of laxatives	Failure to complete tasks; Careless in daily activities	Chills; Sweating	Outbursts of anger; Impulsivity
Intermediate Causes 2	Severely low caloric intake; State of starvation	Behavioural dysregulation; Impulsiveness	Amygdala misdirects signal to the locus cereleus; Enhanced excitability	Increased amygdala activation; Neurochemically-based difficulty regulating emotions
Terminal Effects 2	Preoccupied with food and eating; Social withdrawal	Risky behaviour; Interrupts others frequently	Pounding heart; Chest pain	Fluctuating emotions; Stress-related transient paranoia
Patient Symptoms:				
Presenting	Abuse of laxatives	Careless in daily activities	Sweating	Stress-related transient paranoia
Diverse	Social withdrawal	Interrupts others frequently	Chest pain	Outbursts of anger
Proximal	Negative self-evaluation	Failure to complete tasks	Chills	Fluctuating emotions

Note. A 5th psychological disorder used in both studies, *Autism*, is depicted in Figure 2. These stimuli were presented to participants in diagram format, as described in Study 1.

Appendix B

Lists of Symptoms in the DSM-IV-TR (2000), used for the Psychiatric Disorder Stimuli in Study 2 and Study 3

Disorder	Symptoms to choose from
Borderline Personality Disorder	• Frantic efforts to avoid abandonment • Pattern of unstable and intense relationships • Identity disturbance • Impulsivity • Recurrent suicidal behaviour • Affective instability • Chronic feelings of emptiness • Difficulty controlling anger • Transient, stress-related paranoia
Major Depressive Episode	• Significant, unintentional change in weight • Daily insomnia or hypersomnia • Psychomotor agitation or retardation • Daily fatigue or loss of energy • Feelings of worthlessness • Diminished ability to think or indecisiveness • Recurrent thoughts of death/suicide
General Anxiety Disorder	• Restlessness or feeling keyed up or on edge • Easily fatigued • Difficulty concentrating or mind going blank • Irritability • Muscle tension • Difficulty falling or staying asleep, or restless unsatisfying sleep
Obsessive-Compulsive Personality Disorder	• Preoccupied with details to the extent that major point of activity is lost • Perfectionism that interferes with task completion • Excessively devoted to work and productivity • Scrupulous and inflexible about moral or ethical matters • Unable to discard worn-out or worthless objects • Reluctant to delegate tasks to others • Miserly spending style toward self and others • Rigidity and stubbornness

Schizotypical Personality Disorder	• Ideas of reference • Odd beliefs or magical thinking • Unusual perceptual experiences • Odd thinking and speech • Suspiciousness or paranoid ideation • Inappropriate or constricted affect • Behaviour or appearance that is odd, eccentric, or peculiar • Lack of close friends or confidants • Excessive social anxiety

Chapter 6

Summary of Findings and General Discussion

We introduced the studies described in chapters 2 to 5 as all searching for evidence for Meehl's idea of 'clinicians-reasoning-as-mechanics', or: for clinicians reasoning causally (Meehl, 1954). Causal reasoning is an essential capacity of people, making it possible to predict future events, but more importantly, to predict consequences of interventions (Beach, 1992; Gopnik & Schultz, 2007; Pearl, 2000; Sloman, 2005; Woodward, 2003). For clinicians, the ability to predict consequences of interventions is important, to effectively diminish complaints and problems of their clients. We proposed, in line with Meehl, that clinicians would reason causally when processing client information, and integrate client information to understand mechanisms underlying their clients' problems and to derive effective treatments. This is also how clinicians should proceed, according to handbooks. Clinicians should integrate client information in case formulations that explain how clients' problems are caused and maintained by interacting biological, psychological, and social factors, so interventions can be based on them (see e.g. Carr, 2006; De Bruyn, Ruijssenaars, Pameijer, & Van Aarle, 2003; Eells, 2007; Groth-Marnat, 2003; Kuyken, Fothergill, Musa, & Chadwick, 2005; Wenar & Kerig, 2006).

In several domains it has been shown that judgments and intervention decisions, of both novices and professionals, are based on causal mental models. But in the clinical domain there is hardly any evidence for causal mental models underlying judgments and decisions (see the introductory chapter). Here we will first briefly present the results of our studies as described in chapters 2 to 5. Then we will discuss these findings. We will end with implications of our findings and suggestions for further research.

Findings

In Chapter 2 and Chapter 3 we aimed to find evidence for clinicians' causal reasoning using the think aloud method (Ericsson, 2006; Ericsson & Simon, 1993). We presented clinicians with an intake video and test results of a boy with problematic behaviour at home and in school, and asked them to think aloud about these questions: "What do you think is the matter with this client; what would you propose as follow-up steps?"

Causality in clinicians' mental representations of client information

Chapter 2 reports our analysis of both the structure and the contents of clinicians' mental representations, as we extracted these from the verbalisations clinicians made after they had seen the intake video and had studied all test results. We extracted the structure of clinicians' mental representations in analogy with how Pennington and Hastie (1986) had analysed jurors' verbalisations to extract their mental representations. We distinguished statements about the client and follow-up steps that participants proposed. We identified causal and temporal links between client statements, and causal, temporal, and reason links between client statements and follow-up steps. We found that few client statements were causally linked, and that the causal representations as we reconstructed them often consisted of simple disjoint parts. Further, only one third of the follow-up steps that participants had proposed seemed to be related to these causal representations. Surprisingly, almost half of all follow-up steps seemed unrelated to any of their thoughts about the client.

To analyse the content of the clinicians' verbalisations we used the Case Formulation Content Coding Method of Eells, Kendjelic, Lucas, and Lombart (received from Tracy Eells by e-mail, October 2007, see Eells, Lombart, Kendjelic, Turner, & Lucas, 2005). We looked whether the content of participants' verbalisations was descriptive, classifying, explanatory, or mentioned additional research or intervention plans. For verbalisations with explanatory content we established what specific causal factors and what specific plans for further research and interventions participants mentioned. Consistent with the results of the structure analysis, this content analysis showed that clinicians' verbalisations were mostly descriptive rather than causal. In addition, clinicians appeared to disagree about the causal factors and about the follow-up steps.

Thus, contrary to our expectations, and contrary to findings with jurors, we did not find that clinicians constructed complex, coherent causal mental representations. Also, most of the follow-up steps proposed did not appear to be based on the causal representations we extracted. Our results are consistent however with findings that case formulations are mostly descriptive and lack specified mechanisms explaining a client's problems (Eells, Kendjelic, & Lucas, 1998; Kuyken et al., 2005).

Looking for causes? Causal reasoning in clinicians' information processing

In Chapter 3 we analysed clinicians' early verbalisations: the verbalisations they made while watching the intake video, before receiving test results. If indeed clinical reasoning is a comprehension process similar to text comprehension, in which information is integrated in coherent causal representations, we would expect clinicians similarly to readers to make mostly causal inferences while processing client information (cf. Trabasso & Magliano, 1996; Zwaan & Brown, 1996). Further, we would expect them to ask mostly for causal information in the questions and inquiry plans that they verbalise.

We identified which of the early verbalisations were repetitions of information from the video, and what verbalisations were inferences: additions to information from the video. Additionally, we identified questions and inquiry plans. For each inference, question, and inquiry plan we established whether it was causal, classifying, descriptive, or predicting. Further, we specified the specific contents of the causal and classifying verbalisations.

In line with findings reported in the previous chapter, we found that a minority instead of a majority of inferences, questions, and inquiry plans referred to or asked for causal factors. Content analysis further showed that different clinicians mentioned different causal factors. They agreed more about a classification. Thus, beliefs about possible causes of a client's problems already diverge from the beginning of the diagnostic process, that is: at the intake.

Causal client models in selecting effective interventions

Chapter 4 describes a study in which we used a cognitive mapping technique to elicit clinicians' causal mental representations, to see whether these may be the basis for clinicians' interventions choices. Cognitive mapping techniques aim to extract how people think about problems and how this may influence their decisions, by having people draw causal relations between variables that they believe to be relevant for their decision (Axelrod, 1976; Hodgkinson, Maule, & Bown, 2004). Thus, in contrast to the think aloud method we used before, in this study we explicitly asked clinicians for their causal client models.

Clinicians drew models of how they thought problems and causal factors influencing these problems were causally related for a client's case, and they rank ordered interventions for their expected effectiveness. We analysed whether we could predict which interventions participants would think of as most effective from the causal models that they had drawn. We also compared participants' rankings with an average ranking, which we based on how often all participants together had chosen an intervention to be one of the most effective. We took this average ranking as mirroring common beliefs about the effectiveness of the different interventions. For most participants, we could predict their effectiveness rankings from their causal models quite well, and better than from the average ranking.

We had counterbalanced the order of tasks: half of the participants constructed a causal map before they rank ordered interventions, the other half of the participants rank ordered interventions first. Results were similar for all participants, indicating that our findings are not the result of the specific demand to construct a causal map before ranking interventions for their expected effectiveness.

We also analysed agreement about the causal models and about the effectiveness of interventions. We found that clinicians used different variables in their models, and assumed different causal relations between variables. Further, agreement about the effectiveness of interventions was moderate to poor. Agreement about what variables are affected by different interventions was highest, though still only moderate.

Taking these findings together we concluded that clinicians may often base their expectations of how effective interventions will be on their individual, idiosyncratic causal models, rather than on common beliefs about the effectiveness of interventions.

Causal diversity effects in clinicians' reasoning about disorders

In the studies described in Chapter 5 we didn't look at mental causal representations that clinicians may have for individual clients, but we looked at how causal models for disorders may affect clinicians' reasoning. Kim and Keil (2003) had found that the structure of causal representations, indicating how for a medical disease symptoms originate from a root cause via different causal paths, affected students' reasoning about this disease. Psychology undergraduates rated symptoms originating from a root cause via different causal paths as stronger evidence that a patient has a disease, than symptoms originating from the same causal path. We investigated whether experienced clinicians would similarly show a "causal diversity effect", when reasoning about psychiatric disorders.

First, we presented experienced clinical psychologists with information about how according to recent scientific theories symptoms originate from a root cause via different causal paths, for a set of psychiatric disorders. When reasoning with these theories, clinicians rated clients who had causally diverse symptoms as more likely to have a hypothesized disorder than clients with causally proximal symptoms. However, they only did so when they could first do a similar hypothetical client task, using their own knowledge.

Next, we elicited experienced clinicians' own assumptions about how symptoms originate from a root cause for different psychiatric disorders. Clinicians indicated in diagrams how they thought symptoms of a disorder originate from a root cause via two different causal paths. Then we asked them to choose which of two clients was more likely to have a hypothesized disorder: one with causally diverse symptoms or one with causally

proximal symptoms. The clinicians didn't judge that clients with causally diverse symptoms were more likely to have the disorder than clients with causally proximal symptoms. When we similarly elicited novices' assumptions for how symptoms originate from a root cause for the same set of disorders, and asked them to do a similar client rating task, we found that novices did show a causal diversity effect. Thus, relying on causal diversity seems specific to novices' diagnostic reasoning.

Discussion

Causality in think aloud protocols

Results from the analyses of clinicians' think aloud protocols reported in the first two chapters indicate that clinicians do not engage in elaborate causal reasoning when they think about what is the matter with a client and what are appropriate follow-up steps. Although clinicians did mention several causal factors, their inferences were descriptive rather than causal, they searched for descriptive rather than causal information, and their mental representations as we extracted them were descriptive rather than coherent causal models. Most follow-up steps that clinicians proposed were not related to their causal representations, or even to any descriptive information they mentioned about the client. We discussed several explanations for these findings.

One explanation was that it may be more difficult for clinicians than for jurors to construct an appropriate causal representation. While jurors usually need to reconstruct why one event occurred, the crime, clinicians usually need to explain many problems in a client's life. Further, the structure of a causal representation of events leading to a crime might be simple: events unfold in a chronological order. Gaps in a linear story structure are easily perceived (Keil, 2003). Clients' problems usually result from several interacting biological, psychological, and environmental factors. Multidimensional mechanisms may be harder to identify. It is interesting in this context that Pennington and Hastie (1988) found that the ease with which a specific causal model could be constructed influenced jurors' verdicts. Jurors more often chose a verdict that fit with information that was presented in a chronologically and causally consistent order, than a verdict that fit with information that was presented in a scrambled order. Thus, the ease with which a coherent causal model can be constructed may indeed affect causal reasoning.

The difficulty inherent in clinicians' tasks may be further increased by the fact that scientific knowledge of what the specific causal factors are and how they interact in causing and maintaining psychological problems is incomplete (Kiesler, 1999; Cicchetti & Sroufe, 2000). Also, knowledge is lacking about how precisely interventions exert their effects in multidimensional mechanisms causing and maintaining clients' problems (Bieling & Kuyken, 2003; De Los Reyes & Kazdin, 2006). To make causal inferences in order to arrive at coherent causal mental models, relevant knowledge should be available and accessible (Frank, Koppen, Noordman, & Vonk, 2003; Leon & Perez, 2001; McKoon & Ratcliff, 1992; Pennington & Hastie, 2000; Trabasso & Magliano, 1996).

Apart from relevant knowledge to construct causal mental models, a specific goal that promotes understanding should be present too (Frank et al., 2003; McKoon & Ratcliff, 1992). In general, when the goal is to choose effective interventions, people should construct

causal models (Beach, 1992; Sloman, 2005; Sloman & Hagmayer, 2006). Specifically, clinicians are supposed to construct causal client models to decide on effective interventions (Carr, 2006; De Bruyn et al., 2003; Eells, 2007; Groth-Marnat, 2003; Haynes & Williams, 2003; Kuyken et al., 2005; Wenar & Kerig, 2006). However, in practice clinicians may not adhere to this prescription, but decide about interventions without having a clear causal understanding (see also Keil, 2006). They may for example classify a client's problems on the basis of criteria as defined in the DSM-IV and choose a treatment based on this classification (see e.g. Clarke, DeBar, & Lewinsohn, 2003; Kendall, Aschenbrand, & Hudson, 2003). Findings of Groenier, Pieters, Hulshof, Wilhelm, and Witteman (2008) indicate that in practice clinicians do not see it as their goal to causally understand their clients. Groenier et al. found that clinicians didn't think that a causal analysis of a client's problems was important, but they thought that a description of problems and intervention choice was more important.

Further, clinicians may think that classifications offer explanations. Clinicians mentioned classifications in terms of the DSM in their think aloud protocols, and sometimes suggested that these were causing specific behaviours or problems. According to Leon and Perez (2001) clinicians activate "clinical diagnosis inferences" when processing client information. Although these clinical diagnosis inferences look like classifications (e.g. kleptomania), Leon and Perez argue that clinicians assume an underlying psychopathology causing symptoms and behaviours. In contrast, the DSM-IV-TR (American Psychiatric Association, 2000) aims to describe disorders as clusters of co-occurring symptoms and behaviours in order to promote consensus in classification, while explicitly refraining from describing the aetiology. Mentioning classifications of symptoms as explaining the occurrence of these symptoms may be tautological, but clinicians may not realize this (cf. Keil, 2006).

Finally, an explanation for why we found so few causal links in clinicians' think aloud protocols may be that clinicians activate schemas or scripts when reasoning about clients (Hamm, 2003; cf. Pliske & Klein, 2003), and mention only elements of these in their protocols. Still, this would not explain our finding that many follow-up steps were also not related to any descriptive statements about the client.

Different methods

For methodological issues relating to sample size, case materials presented, and instructions we gave to participants, we refer to the discussions in Chapter 2 and 3. Here we focus on fundamental concerns about the think aloud method, and compare the think aloud method with the cognitive mapping method. We thus aim to clarify the discrepancy between our findings in the think aloud study and the cognitive mapping study. From our analyses in Chapter 4 of the causal maps that clinicians had generated, it appeared that these were more elaborate, coherent, and complex than the causal models that we had extracted from the think aloud protocols. Furthermore, with clinicians' causal maps we could predict their ratings of the effectiveness of interventions quite well, which indicates that clinicians may base their interventions on causal mental client models. Thus, the think aloud procedure we adopted may not have adequately tracked clinicians' reasoning. Alternatively, the causal maps drawn by clinicians may not have been representative for the mental representations that they normally construct when reasoning about a client.

The main assumption underlying the think aloud method is that people can reliably verbalise thoughts that pass through working memory while they are reasoning about information (Ericsson, 2006; Ericsson & Simon, 1993). Ericsson and Simon (1993) specified in detail the procedures that should be followed in a think aloud study in order to ensure that the protocols obtained are true reflections of spontaneous thought processes, and not, for example, memories of whatever thoughts may have occurred, or justifications for decisions. We adhered to this procedure closely, by making our task realistic, offering sufficient information, prompting with questions that seem to be most frequently asked in practice, and enabling clinicians to verbalise their thoughts at any moment. We thus meant to ensure that we would capture clinicians' thinking processes as these may occur spontaneously in practice when clinicians are thinking about a client.

Still, although thinking aloud may result in a reliable and rich data set of conscious thoughts that occur while people are engaged in a specific task, it will only deliver a subset of all thoughts (Einhorn, Kleinmuntz, & Kleinmuntz, 1979; Ericsson, 2006; Ericsson & Simon, 1993; Graesser, Millis, & Zwaan, 1997; Zwaan & Brown, 1996). We cannot rule out that causal assumptions are not verbalised when they are so evident that they are not attended to consciously, or conversely, when they are not clear enough to be verbalised. In our coding procedure causal links had to be verbalised explicitly, in analogy with Pennington and Hastie's coding procedure, which demanded explicit verbalisation of causal links. Implicit causal links would thus not be coded. Still, the content analysis gave very much the same results as the structure analysis. Also, from simply reading clinicians' think aloud protocols it was not obvious that clinicians made many implicit causal links.

With the cognitive mapping method we explicitly asked clinicians to indicate which variables they thought were important and how they thought these were causally related. Contrary to what may happen with the think aloud method, causal mapping may elicit causal inferences like the ones that occur when clinicians are reasoning in actual practice, plus causal inferences that they would not normally consider. It was clear that clinicians were not simply copying their mental causal models on paper in the cognitive mapping task. The construction of the causal maps took quite some time and effort. Clinicians thought about which variables to include and which causal relations to draw. They frequently changed their minds, often resulting in messy maps with many corrections. It seems rather unlikely that clinicians have mental representations as complex as the ones they drew, and that they take into account all directly and indirectly affected variables when they derive the effects of interventions from them.

Hodgkinson, Maule, and Bown (2004) compared two techniques to elicit the causal relations managers assumed to hold between variables in a strategic investment decision task. In one task managers constructed causal maps by drawing causal relations they believed to hold between a fixed set of variables, and rated the strengths of these relations. In the other task managers indicated for each pair of variables whether they thought there was a causal relation between them and, if so, how strong it was. Hodgkinson et al. found that with the pairwise comparison technique the managers indicated more and stronger relations than with the causal mapping technique. They suggested that with pairwise comparisons, people might add causal relations they would not normally consider, while

with causal mapping they might omit causal relations they did consider, but did not recall when drawing the map. We suggest that a similar explanation may apply to our findings. Clinicians may believe that there are many causal relations between variables, but they may not consider all these relations when reasoning about a client. However, with an explicit instruction to think of causal relations, clinicians may consider more causal relations than they would otherwise do. Furthermore, drawing causal maps will facilitate reasoning about how different factors interact in causing a client's problems, since causal maps can serve as a visual memory aid (cf. Morton, 2004; Nadkarni & Shenoy, 2004). When thinking aloud on the other hand, clinicians may not verbalise causal relations that they do consider.

Findings reconciled?

To reconcile our findings obtained in the think aloud study and in the cognitive mapping study we propose that clinicians do reason causally, but that in actual practice they will not do so as elaborately as in the cognitive mapping study. The causal maps that clinicians draw will contain the variables and the causal relations they consider most important. In a realistic situation clinicians may also consider these variables and relations, and they may verbalise them, although perhaps not all of them, when thinking aloud. The maps may further contain additional variables and especially causal relations between variables which clinicians would not normally consider, if they are not explicitly prompted to, and if they don't have variables visualised as a memory aid.

The variables that clinicians consider to be the most important may be those causal factors that affect most other factors: causally central or initiating variables (cf. Kim & Ahn, 2002; Rehder & Kim, 2006). Additional variables, which clinicians may not normally consider, may be those that are less central in the maps. When clinicians believe that the most effective interventions are those affecting the most causally central variables (cf. Haynes & Williams, 2003; Yopchick & Kim, 2008), the maps will predict clinicians' ratings of the effectiveness of interventions well, whether maps are drawn before or after rating the effectiveness of interventions.

Clinicians' knowledge about psychological problems may not consist of elaborate, coherent causal networks. Their knowledge base might resemble the physics knowledge base of physics students, which consists of loose fragments that are hardly connected in coherent causal theories (Di Sessa, Gillespie, & Esterly, 2004). Further, it may not normally be necessary for clinicians to provide detailed explanations of mechanisms underlying clients' problems. However, with coarse intuitions as to what causal factors are important and how these may exert their influence and with specific client information, clinicians might be able to construct complex mechanisms when they are prompted to do so (cf. Keil, 2003). It remains to be seen, however, how accurate such causal models are.

Disagreement about causal models

The analyses of clinicians' think aloud protocols and their cognitive maps showed that their causal client representations are highly idiosyncratic. This finding is not new. Reliability among clinicians appears to be generally poor (Garb, 1998), and specifically so for causal inferences (Garb, 1998, 2005; Kuyken et al., 2005; Persons & Bertagnolli, 1999). What is striking in our findings is that disagreement starts as early as at the intake interview.

The low agreement among clinicians about causal client models might be explained by a lack of common scientific knowledge of the complex aetiology of psychological problems (Cicchetti & Sroufe, 2000; De Los Reyes & Kazdin, 2006; Kiesler, 1999). Also, clinicians will not usually get adequate feedback about the accuracy of their judgments (Caspar, 1997; Garb, 2005; Lichtenberg, 1997). Further, biases such as the availability bias, the representativeness bias, or the hindsight bias may increase the idiosyncrasy of the clinicians' causal beliefs (Garb, 1998; Lichtenberg, 1997; Nezu & Nezu, 1993). Finally, differences in education and training, between institutions, and clients may contribute to the development of idiosyncratic causal beliefs.

Experience

In Chapter 5, we found that the causal structure of implicit theories affects reasoning of novices in clinical psychology. Novices rated symptoms with more diverse causal origins in their own theories as providing stronger evidence for a hypothesized disorder than symptoms with similar causal origins. This finding goes beyond those of Kim and Keil (2003) and Kim, Yopchick, and De Kwaadsteniet (2008), in two ways. First, our novices were beginners in their professional domain. They were close to graduating as a clinical psychologist, or had recently graduated. Participants in the studies of Kim and Keil (2003) and Kim et al. (2008) were psychology undergraduates who reasoned about real and artificial conditions in the medical domain, so they were lay people. Second, we assume that our participants used their own implicit theories, which were not directly available for inspection when doing the classification task. In contrast, in the studies of Kim and Keil (2003) and Kim et al. (2008) participants had studied causal structures immediately before the classification task, and had them available for inspection during it.

We did not find that experienced clinicians, who had seen clients for at least ten years, judged clients with symptoms that are causally diverse in their own theories as more likely to have a hypothesized disorder than clients with symptoms that are causally proximal. Our task demand to fit causal assumptions into the diagram structures might have been too artificial. Still, for those diagrams that clinicians rated to be representative (see Chapter 5) we found no causal diversity effect with experienced clinicians, while we did with novices. To explain this difference we suggested that relying on causal diversity might be a useful strategy for novices only. The experienced clinicians may rely on experiential knowledge about the diagnosticity of symptoms, that is: knowledge about how likely it is that a client has a specific disorder given the presence of specific symptoms, rather than on causal structure of their theories. Diversity in a causal structure might be informative precisely when this knowledge about diagnosticity of symptoms is still lacking or less accessible, as with novices. With no further information about probabilities, it can be inferred that the presence of a root cause is more likely given that causally diverse evidence is present than when evidence with similar causal origins is present (cf. Horwich, 1982; Kincannon & Spellman, 2003).

Findings of Ahn, Flanagan, Marsh, and Sanislow (2006) are interesting in this light. Novices seem to believe that mental disorders have a defining feature which causes the symptoms associated with them, and that disorders disappear if this defining feature is removed. They further appear to believe that experts will have knowledge about these

defining features. Experienced clinicians do not seem to share these beliefs. Thus, novices in our studies may assume there is one root cause of all symptoms associated with a disorder. Experienced clinicians on the other hand rather seem to believe that disorders can arise from different combinations of causes (Ahn et al., 2006), and that symptoms can have different causes. Experienced clinicians' theories for how symptoms are caused in different disorders will thus be overlapping, while novices' theories for different disorders will be more disjoint (cf. Murphy & Wright, 1984).

Experience may also be expected to affect causal beliefs about individual clients. In the think aloud study we found no effects of experience, but this might be due to the small sample size. It cannot be ruled out that more experienced clinicians reason with higher level concepts, and skip elaborate, detailed causal reasoning (cf. Schmidt & Boshuizen, 1993), or have developed scripts (Charlin, Tardif, & Boshuizen, 2000; Hamm, 2003). Research in the medical domain, however, shows that experienced doctors use both higher level concepts and detailed causal concepts in an integrated way (Rikers, Loyens, & Schmidt, 2004; Van de Wiel, Boshuizen, & Schmidt, 2000). Research on clinical case formulations shows that more experienced and higher qualified clinicians construct more complex, elaborate, and comprehensive client representations. Also, intervention choices seem to be more evidently linked to these representations (Eells & Lombart, 2003; Eells et al., 2005; Kuyken et al., 2005).

Finally, we note that in the clinical domain, experience is generally not associated with better performance (Dawes, 1994; Eells et al., 2005; Garb, 1998; Krol, De Bruyn, & Van den Bercken, 1992; Lichtenberg, 1997; Witteman & Van den Bercken, 2007). Experience is however associated with confidence: with more experience clinicians become more confident about their judgments (Krol et al., 1992; Lichtenberg, 1997; Oskamp, 1965; see also Chapter 2 and Chapter 4).

Causal versus probabilistic reasoning versus other reasoning strategies.

We tried to find evidence that clinicians reason causally, as Meehl (1954) proposed. Our findings indicate that clinicians do, although the extent to which they engage in causal reasoning seems to differ with different instructions, and with different clinicians. What about other reasoning strategies, such as probabilistic reasoning? In our discussion of the findings of our last series of studies we suggested that clinicians, at least experienced clinicians, may reason using probabilistic information, rather than causal information. Clinicians may also adopt other strategies, for example script-based reasoning (activating schemas, looking for schema-consistent information, and choosing interventions that are activated with the script, cf. Hamm, 2003), exemplar-based reasoning (activating memories of similar clients, and choosing the interventions that were effective for these clients), or rule-based reasoning ("if a specific set of symptoms is present, choose intervention X").

We have only looked for evidence for causal reasoning without taking into account alternative strategies. Therefore, we cannot make inferences about the proportion of causal reasoning compared to other reasoning strategies. It is likely that clinicians reason using different strategies, depending on the specific case, their experience, the task demands, individual reasoning styles, and so on (Brammer, 1997; Norman, Eva, Brooks, & Hamstra, 2006). Different strategies may also be combined. For example, in diagnostic reasoning

causal and probabilistic knowledge will be combined, since causes are known to usually generate effects with some uncertainty (Rehder, 2007; Tenenbaum, Griffiths, & Niyogi, 2007).

Implications

Perhaps the clearest finding from our studies was that clinicians' causal beliefs diverge (cf. Garb, 1998, 2005; Kuyken et al., 2005; Persons & Bertagnolli, 1999), and that they do so already at the intake. Further, we found that clinicians' idiosyncratic causal models predict their ratings of the effectiveness of interventions better than shared beliefs do. This finding suggests that clinicians may arrive at different intervention decisions for the same client, as a consequence of their idiosyncratic causal client representations.

Jensen-Doss and Weisz (2008) found that accuracy of clinicians' classifications predicted improved treatment processes and outcomes. It might be expected that inaccurate identifications of causal mechanisms also may lead to poorer treatment outcomes. Handbooks prescribe that clinicians should integrate client information in case formulations that explain how clients' problems are caused and maintained, so interventions can be based on them (e.g. Carr, 2006; De Bruyn, et al., 2003; Eells, 2007; Groth-Marnat, 2003; Wenar & Kerig, 2006). Nevertheless, one might argue that clinicians had better not draw causal inferences about the mechanisms underlying their clients' problems. They might rather proceed by describing or classifying a client's problems and, based on this description or classification, choose an empirically supported treatment, using available actuarial rules as Meehl and Grove (2006) proposed. For example, predictions about treatment success seem to be better when made with actuarial rules than when made by clinicians who use their own clinical methods (Ægisdóttir, et al., 2006; Grove, Zald, Lebow, Snitz, and Nelson, 2000; Meehl, 1954). Further, empirically supported treatments seem to be more effective than the interventions that are usually adopted in practice (Weisz, Jensen-Doss, & Hawley, 2006).

In practice clinicians often do not seem to follow actuarial rules (Ægisdóttir, et al., 2006; Caspar, Berger, & Hautle, 2004; Meehl & Grove, 2006). Clinicians also often do not seem to use empirically validated treatments, and empirically validated treatments often do not seem to be integrated extensively in training programs (Addis & Krasnow, 2000; Gotham, 2006; Weisz & Kazdin, 2003). Kendall and Beidas (2007) argued that inflexibility of treatment manuals may hamper the implementation of evidence-based practices (cf. Nelson & Steele, 2008). They believe that clinicians will be more ready to follow treatment manuals when these leave room for flexible application to individual clients. We suggest that clinicians might find it difficult to combine actuarial methods and rigid treatment manuals with how they usually reason (Ahn & Kim, 2008; Caspar et al., 2004).

It may be true that clinicians do not often choose empirically supported treatments based on descriptions or classifications of clients' symptoms, but we should note that this procedure isn't always as straightforward as it may seem. Although the number of empirically supported treatments increases, most treatments have not (yet) been tested empirically. Also, there are no empirically supported treatments for all disorders yet (Braet & Bögels, 2008; Weisz & Kazdin, 2003). Further, in the case of co-occurring disorders and when a treatment fails, treatment choice may become more difficult (Nelson-Gray, 2003). Finally, accurate description or classification of problems may not be straightforward in practice,

because of unclear symptom presentation in clients, ambiguous definitions, or subjective interpretations (cf. Jensen & Weisz, 2002; Widiger & Clark, 2000).

Although clinicians' causal reasoning seems to be unreliable, causal reasoning can in itself be perfectly rational and validated. In an ideal situation, when well-corroborated theories are available the relevant constructs of which can be measured accurately, clinicians may arrive at reliable and accurate causal explanations (Grove, 2005; Meehl & Grove, 2006). In the medical domain it was found that students recall causal knowledge better, and make more accurate diagnoses when they have learned about the causal structure of some conditions, than when they had learned about probabilities for these conditions (Woods, Brooks, & Norman, 2005). Thus, because (i) clinicians do seem to reason causally (although perhaps not always profoundly so), (ii) they may need to engage in causal reasoning in complex cases, or cases for which no empirically supported treatments are available, (iii) causal understanding may provide clinicians with high levels of confidence (Keil, 2006; Pearl, 2000) , and finally (iv) it seems difficult to convince clinicians to use other strategies than the ones they are used to (Ægisdóttir et al., 2006, Caspar, Berger, Hautle, 2004; Gotham, 2006; Kendall & Beidas, 2007; Meehl & Grove, 2006), we suggest two ways in which clinicians' performance might be improved, both focussing on clinicians' causal reasoning strategies.

First, clinicians may more often choose to use actuarial rules and empirically supported treatments when they are provided with their theoretical underpinnings. Clinicians may be especially interested in *why* treatments work, so they can use this knowledge in a causal reasoning process about what is the matter with a client and what intervention will be most effective (see also Kazdin, 2008). Knowing *why* an intervention is effective may convince clinicians to choose an intervention for a specific client sooner than just knowing that an intervention is often effective.

Second, decision aids and training of clinicians should focus on making causal assumptions explicit and on testing causal assumptions. In our cognitive mapping study, drawing a causal model before predicting the effectiveness of interventions did not appear to change participants' predictions. Also, the variation in the models that participants constructed was large. Still, we believe that techniques aimed at visualization of causal beliefs might be useful in clinical practice and in clinical research. For example, the theory-neutral 'causal modelling' method of Morton (2004; see also Krol, Morton, & De Bruyn, 2004) prompts clinicians to explicate all biological, psychological, behavioural, and environmental factors that are relevant for a client, and to indicate the causal relations between these factors. Such a visualisation of complex ideas shows where additional assumptions are needed to arrive at a coherent explanation, and which specific tests should be done to test important assumptions. Further, visualization of causal models may be useful to communicate causal hypotheses to clients and to colleagues, and thus serve as a basis for discussion and adaptation (Nadkarni & Shenoy, 2004). In particular, causal client models may be adapted by measuring treatment effects with individual clients (see e.g. Lambert, 2001). It is interesting in this light that generating explicit explanations may lead to increased causal knowledge (Chi, de Leeuw, Chiu, & LaVancher, 1994; Wellman & Liu, 2007).

In sum, we suggest that dissemination and implementation of empirically validated treatment and assessment methods may be facilitated if clinicians learn about their

underlying theories. Further, clinicians should be careful when generating hypotheses about mechanisms underlying clients' problems (Garb, 2005; Wood, Garb, Lilienfeld, & Nezworski, 2003). Causal modelling methods may aid the construction and testing of causal client models (Haynes, Spain, & Oliveira, 1993; Morton, 2004; Nadkarni & Shenoy, 2004).

Further research

From our studies it remains unclear to what extent clinicians engage in causal reasoning when processing client information: do clinicians construct mental causal client models that are complex and detailed, or rather simple and incomplete (cf. Keil, 2006)? To investigate this further, other methods could be used, for example measuring sentence reading times (cf. Cobos et al., 2008), measuring recall or recognition (cf. Ackerman, 1992), or priming tasks (cf. Leon & Perez, 2001) (see also Magliano, 1999). Also, we suggested that causal reasoning as it occurred in thinking aloud may be limited due to a lack of relevant causal knowledge. Thus, it would be interesting to see whether clinicians do indeed engage in more causal reasoning, and show this in think aloud protocols, if they are provided with relevant causal knowledge. Such a study may provide support for our findings of the think aloud study. Further, to learn more about clinicians' reasoning, a wider range of possible reasoning strategies should be investigated, also taking into account individual differences in the use of these strategies. It seems that definitive results on clinicians' thoughts will not appear soon, since they are so various and largely covert, and thus difficult to investigate (Caspar, 1997; Caspar et al., 2004).

Research on clinicians' intervention choices is scarce. We found that clinicians' effectiveness ratings of interventions depend on their causal assumptions. However, clinicians may base their intervention choices not only on effectiveness ratings, but on many other factors too, such as available time and budget, client preferences, experience with specific interventions, etcetera (Beutler, Malik, Alimohamed, Harwood, Talebi, Noble, & Wong, 2004; Clarkin & Levy, 2004; Nelson & Steele, 2008). To understand clinicians' actual choices, these other factors should also be investigated.

In line with a suggestion made above, it seems to be of direct relevance for practice to investigate whether actuarial rules and standardized treatments will be used sooner, and more strictly adhered to, if clinicians understand the theoretical underpinnings. If this turns out to be true, this would emphasize the need for research on mechanisms of change of interventions (see also Kazdin, 2008). Also, the usefulness of causal modelling methods for constructing and testing causal client representations should be investigated: do such models increase reliability and validity of causal client models, and enhance treatment outcomes?

Our studies in causal diversity effects might be followed up by investigating what strategies experienced clinicians use when they do not seem to rely on diversity. The structure of the causal knowledge of experienced clinicians may be investigated further to find out how it affects their reasoning. It is also of interest to find out how the structure of knowledge of experienced clinicians differs from that of novices, and whether differences in reasoning are related to such differences. Finally, we note that our diversity studies in Chapter 5 show that it is important to take clinicians' experience into account when studying

clinicians' reasoning. General conclusions about clinicians' performance from tasks in which only novices participate may not be valid.

To conclude

I end with the observation that it will be equally impossible for researchers to arrive at a complete description of what a clinician does and why, as it is for clinicians to arrive at a complete description of how a client behaves and why. Thus, clinicians are opaque boxes to researchers, as are clients for clinicians. Researchers may have idiosyncratic causal models of clinicians, like clinicians have of clients. I hope the studies described in this dissertation, with all their limitations, will be perceived as sensible studies to learn more about clinicians' reasoning. I also hope that their results are an increase, albeit small, in our understanding of clinicians and that they will inspire future studies, so that in the end people with psychological problems may profit.

References

Ackerman, B. P. (1992). The sources of children's source errors in judging causal inferences. *Journal of Experimental Child Psychology, 54*, 90-119.

Addis, M. E., & Krasnow, A. D. (2000). A national survey of practicing psychologists' attitudes toward psychotherapy treatment manuals. *Journal of Consulting and Clinical Psychology, 68*, 331-339.

Ahn, W. (1998). Why are different features central for natural kinds and artifacts?: The role of causal status in determining feature centrality. *Cognition, 69*, 135-178.

Ahn, W., Flanagan, E. H., Marsh, J. K., & Sanislow, C. A. (2006). Beliefs about essences and the reality of mental disorders. *Psychological Science, 17*, 759-766.

Ahn, W., Gelman, S. A., Amsterlaw, J. A., Hohenstein, J., & Kalish, C. W. (2000). Causal status effect in children's classification. *Cognition, 76*, B35-B43.

Ahn, W., & Kim, N. S. (2008). Causal theories of mental disorder concepts. *Psychological Science Agenda, 22*, 3-8.

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.

American Psychiatric Association (2000). *DSM-IV-TR: Diagnostic and Statistical Manual of Mental Disorders* (4th ed., rev.). Washington D.C.: Author.

Arocha, J. F., & Patel, V. L. (1995). Construction-Integration in theory and clinical reasoning. In C.A. Weaver, III, S. Mannes, & C.R. Fletcher (Eds.), *Discourse comprehension: Essays in honor of Walter Kintsch* (pp.359-381). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Axelrod, R. M. (Ed.) (1976). *Structure of Decision: Cognitive Maps of Political Elites*. Princeton, N.J.: Princeton University Press.

Ægisdóttir, S., White, M. J., Spengler, P. M., Maugherman, A. S., Anderson, L. A., Cook, R. S., et al. (2006). The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction. *The Counseling Psychologist, 34*, 341-382.

Bailenson, J. N., Shum, M. S., Atran, S., Medin, D. L., & Coley, J. D. (2002). A bird's eye view: Biological categorization and reasoning within and across cultures. *Cognition, 84*, 1-53.

Beach, L. R. (1992). Epistemic strategies: Causal thinking in expert and nonexpert judgment. In G. Wright & F. Bolger (Eds.), *Expertise and Decision Support* (pp. 107-127). New York: Plenum Press.

Belmonte, M. K., Allen, G., Beckel-Mitchener, A., Boulanger, L. M., Carper, R. A., & Webb, S. J. (2004). Autism and abnormal development of brain connectivity. *The Journal of Neuroscience, 24*, 9228-9231.

Belmonte, M. K., Cook Jr E. H., Anderson, G. M., Rubenstein, J. L. R., Greenough, W. T., Beckel-Mitchener, et al. (2004). Autism as a disorder of neural information processing: Directions for research and targets for therapy. *Molecular Psychiatry, 9*, 646-663.

Berman, R. A., & Slobin, D. I. (1994). Glossing and transcription conventions (Appendix 2). In R. A. Berman & D. I. Slobin (Eds.), *Relating Events in Narrative Across Linguistic Study* (pp.655-664). Hillsdale NJ: Erlbaum.

Beutler, L. E., Malik, M., Alimohamed, S., Harwood, T. M., Talebi, H., Noble, S., & Wong, E. (2004). Therapist variables. In M. J. Lambert (Ed.), *Bergin and Garfield's Handbook of*

Psychotherapy and Behavior Change (5th ed.) (pp. 227-306). New York: John Wiley & Sons, Inc.

Bieling, P. J., & Kuyken, W. (2003). Is cognitive case formulation science or fiction? *Clinical Psychology: Science and Practice, 10*, 52-69.

Böhm, G., & Pfister, H. R. (2000). Action tendencies and characteristics of environmental risks. *Acta Psychologica, 104*, 317-337.

Boshuizen, H. P. A. (2004). Does practice make perfect? A slow and discontinuous process. In H. P. A. Boshuizen, R. Bromme, & H. Gruber (Eds.), *Professional Learning: Gaps and Transitions on the way From Novice to Expert* (pp. 73-93). Dordrecht: Kluwer Academic Publishers.

Braet, C. & Bögels, S. M. (2008). *Protocollaire Behandelingen Voor Kinderen met Psychische Klachten* [Protol-based treatments for children with psychological complaints]. Amsterdam: Boom.

Brammer, R. (1997). Case conceptualization strategies: The relationship between psychologists' experience levels, academic training, and mode of clinical inquiry. *Educational Psychology Review, 9*, 333-351.

Carr, A. (2006). *The Handbook of Child and Adolescent Clinical Psychology: A Contextual Approach* (2nd ed.). East Sussex: Routledge.

Carroll, J. M., Maughan, B., Goodman, R., & Meltzer, H. (2005). Literacy difficulties and psychiatric disorders: Evidence for comorbidity. *Journal of Child Psychology and Psychiatry, 46*, 524-532.

Caspar, F. (1997). What goes on in a psychotherapist's mind? *Psychotherapy Research, 7*, 105-125.

Caspar, F., Berger, T., & Hautle, I. (2004). The right view of your patient: A computer-assisted, individualized module for psychotherapy training. *Psychotherapy: Theory, Research, Practice, Training, 41*, 125-135.

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.

Chi, M. T. H. (2006). Laboratory methods for assessing experts' and novices' knowledge. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge Handbook of Expertise and Expert Performance* (pp. 167-184). Cambridge: Cambridge University Press.

Chi, M. T. H., de Leeuw, N., Chiu, M.-H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*, 439-477.

Cicchetti, D., & Sroufe, A. L. (2000). Editorial: The past as prologue to the future: The times, they've been a-changin'. *Development and Psychopathology, 12*, 255-264.

Clarke, G. N., DeBar, L. L., & Lewinsohn, P. M. (2003). Cognitive-behavioral group treatment for adolescent depression. In A. E. Kazdin & J. R. Weisz (Eds.), *Evidence-Based Psychotherapies for Children and Adolescents* (pp. 120-134). New York: Guilford Press.

Clarkin, J. F., & Levy, K. N. (2004). The influence of client variables on psychotherapy. In: M. J. Lambert (Ed.), *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change* (5th ed.) (pp. 194-226). New York: John Wiley & Sons, Inc.

Cobos, P. L., Flores, A., & López, F. J. (2008, July). Causal reasoning in clinicians during reading of reports. In A. Godoy (Chair), *Clinical Reasoning*. Symposium conducted at the XXIX International Congress of Psychology, Berlin.

Dawes, R.M. (1994). *House of Cards: Psychology and Psychotherapy Built on Myth*. New York: Free Press.

Dawes, R. M. (1999). A message from psychologists to economists: Mere predictability doesn't matter like it should (without a good story appended to it). *Journal of Economic Behavior & Organization, 39*, 29-40.

Dawes, R. M. (2005). The ethical implications of Paul Meehl's work on comparing clinical versus actuarial prediction methods. *Journal of Clinical Psychology, 61*, 1245-1255.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science, 23*, 1668-1674.

De Bruyn, E., Ruijssenaars, A. J. J. M., Pameijer, N. K., & Van Aarle, E. J. M. (2003). *De diagnostische cyclus: Een praktijkleer* (The diagnostic cycle: Learning in practice). Leuven: Acco.

De Los Reyes, A., & Kazdin, A. E. (2006). Conceptualizing changes in behavior in intervention research: The range of possible changes model. *Psychological Review, 113*, 554-583.

De Wit, C. A. M. (2000). *Depressie bij Kinderen en Adolescenten. Theorie en Onderzoek, Diagnostiek en Behandeling* (Depression in children and adolescents. Theory and research, diagnosis and treatment). Houten/Diegem: Bohn Stafleu van Loghum.

Di Sessa, A. A., Gillespie, N. M., & Esterly, J. B. (2004). Coherence versus fragmentation in the development of the concept of force. *Cognitive Science, 28*, 843-900.

Drabick, D. A. G., Gadow, K. D., & Sprafkin, J. (2006). Co-occurence of conduct disorder and depression in a clinic-based sample of boys with ADHD. *Journal of Child Psychology and Psychiatry*, 47, 766-774.

Eells, T. D. (2007). *Handbook of Psychotherapy Case Formulation* (2nd Edition). New York: The Guilford Press.

Eells, T. D., Kendjelic, E. M., & Lucas, C. P. (1998). What's in a case formulation? Development and use of content coding manual. *Journal of Psychotherapy Practice and Research, 7*, 144-153.

Eells, T. D., & Lombart, K. G. (2003). Case formulation and treatment concepts among novice, experienced, and expert cognitive-behavioral and psychodynamic therapists. *Psychotherapy Research, 13*, 187-204.

Eells, T. D., Lombart, K. G., Kendjelec, E. M., Turner, L. C., & Lucas, C. P. (2005). The quality of psychotherapy case formulations: a comparison of expert, experienced, and novice cognitive-behavioral and psychodynamic therapists. *Journal of Consulting and Clinical Psychology, 73*, 579-589.

Einhorn, H. J., Kleinmuntz, D. N., & Kleinmuntz, B. (1979). Linear regression and process-tracing models of judgment. *Psychological Review, 86*, 465-485.

Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1978). *Medical Problem Solving: An Analysis of Clinical Reasoning*. Cambridge, MA: Harvard University.

Endsley, M. R. (1997). The role of situation awareness in naturalistic decision making. In C. E. Zsambok & G. Klein (Eds.), *Naturalistic Decision Making* (pp. 269-284). Mahwah, NJ: Lawrence Erlbaum Associates.

Ericsson, K. A. (2006). Protocol analysis and expert thought: concurrent verbalizations of thinking during experts' performance on representative tasks. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge Handbook of Expertise and Expert Performance* (pp. 223-241). Cambridge: Cambridge University Press.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol Analysis: Verbal Reports as Data*. Cambridge, MA: MIT Press.

Eva K. W., Brooks L. R., & Norman G. R. (2002). Forward reasoning as a hallmark of expertise in medicine: Logical, psychological, and phenomenological inconsistencies. In S. P. Shohov (Ed.), *Advances in Psychological Research* (Volume 8, pp.41-69). New York : Nova Science Publishers, Inc.

Evans, D. L., Herbert, J. D., Nelson-Gray, R. O., & Gaudiano, B. A. (2002). Determinants of diagnostic prototypicality judgments of the personality disorders. *Journal of Personality Disorders, 16*, 95-106.

Fairburn, C. G., Cooper, Z., & Shafran, R. (2003). Cognitive behaviour therapy for eating disorders: a "transdiagnostic" theory and treatment. *Behaviour Research and Therapy, 41,* 509-528.

Fairburn, C. G., Shafran, R., & Cooper, Z. (1999). A cognitive behavioural theory of anorexia nervosa. *Behaviour Research and Therapy, 37,* 1-13.

Fleiss, J. L. (1981). *Statistical Methods for Rates and Proportions* (2nd ed.). New York: Wiley.

Frank, S. L., Koppen, M., Noordman, L. G. M., & Vonk, W. (2003). Modeling knowledge-based inferences in story comprehension. *Cognitive Science, 27*, 875-910.

Furnham, A. (1995). Lay beliefs about phobia. *Journal of Clinical Psychology, 51*, 518-525.

Gaines, D. M., Brown, D. C., & Doyle, J. K. (1996). A computer simulation model of juror decision making. *Expert Systems With Applications, 11*, 13-28.

Galanter, C. A., & Patel, V. L. (2005). Medical decision making: A selective review for child psychiatrists and psychologists. *Journal of Child Psychology and Psychiatry, 46*, 675-689.

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

Garb, H. N. (1998). *Studying the Clinician: Judgment Research and Psychological Assessment*. Washington, DC: American Psychological Association.

Garb, H. N. (2005). Clinical judgment and decision making. *Annual Review of Clinical Psychology, 1*, 67-89.

Garcia-Retamero, R., & Hoffrage, U. (2006). How causal reasoning simplifies decision-making. *Minds & Machines, 16*, 365-380.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review, 103*, 650-669.

Goldstein, W. M., & Weber, E. U. (1995). Content and discontent: Indications and implications of domain specificity in preferential decision making. In J. Busmeyer, D. L. Medin, & R. Hastie (Eds.), *Decision making from a cognitive perspective* (Vol. 32. pp. 83-126). San Diego, CA: Academic Press.

Gopnik, A., & Schulz, L. (2007). *Causal Learning: Psychology, Philosophy, and Computation*. Oxford: Oxford University Press.

Gotham, H. J. (2006). Advancing the implementation of evidence-based practices into clinical practice: How do we get there from here? *Professional Psychology: Research and Practice, 37*, 606-613.

Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse comprehension. *Annual Review of Psychology, 48*, 163-189.

Graesser, A. C., Olde, B., & Klettke, B. (2002). How does the mind construct and represent stories? In M. C. Green, J. J. Strange, & T. C. Brock (Eds.), *Narrative Impact: Social and Cognitive Foundations* (pp. 231-263). Mahwah, NJ: Lawrence Erlbaum Associates.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative comprehension. *Psychological Review, 101*, 371-395.

Green, D. W., & McManus, I. C. (1995). Cognitive structural models: The perception of risk and prevention in coronary heart disease. *British Journal of Psychology, 86*, 321-336.

Groenier, M., Pieters, J. M., Hulshof, C. D., Wilhelm, P., & Witteman, C. L. M. (2008). Psychologists' judgements of diagnostic activities: Deviations from a theoretical model. *Clinical Psychology and Psychotherapy, 15*, 256-265.

Groth-Marnat, G. (2003). *Handbook of psychological assessment* (4th ed.). New York: John Wiley & Sons, Inc.

Grove, W. M. (2005). Clinical versus statistical prediction: The contribution of Paul Meehl. *Journal of Clinical Psychology, 61*, 1233-1243.

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment, 12*, 19-30.

Hagmayer, Y., & Sloman, S. A. (2005). Causal models in decision making: Choice as intervention. *Proceedings of the Twenty-Seventh Annual Conference of the Cognitive Science Society*. Stresa, Italy.

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

Hagmayer, Y., & Waldmann, M. R. (2002). How temporal assumptions influence causal judgments. *Memory & Cognition, 30*, 1128-1137.

Hamm, R. M. (2003). Medical decision scripts: Combining cognitive scripts and judgment strategies to account fully for medical decision making. In D. Hardman & L. Macchi (Eds.), *Thinking: Psychological Perspectives on Reasoning, Judgment and Decision Making* (pp. 315-345). Chichester: John Wiley & Sons, Inc.

Hastie, R., & Pennington, N. (2000). Explanation-based decision making. In T. Connolly, H. R. Arkes, & K. R. Hammond (Eds.), *Judgment and Decision Making: An Interdisciplinary Reader* (2nd ed.) (pp. 212-228). Cambridge, UK: Cambridge University Press.

Haynes, S. N. (1992). Models of causality in psychopathology: *Toward Dynamic, Synthetic and Nonlinear Models of Behavior Disorders*. New York: Macmillan.

Haynes, S. N., Spain, E. H., & Oliveira, J. (1993). Identifying causal relationships in clinical assessment. *Psychological Assessment, 5*, 281-291.

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.

Heit, E., Hahn, U., & Feeney, A. (2005). Defending diversity. In Ahn, W., Goldstone, R. L., Love, B. C., Markman, A. B., & Wolff, P. (Eds.), *Categorization Inside and Outside the Laboratory* (pp.87-99). Washington, DC: American Psychological Association.

Hodgkinson,G. P., Maule, A. J., & Bown, N. J. (2004). Causal cognitive mapping in the organizational strategy field: A comparison of alternative elicitation procedures. *Organizational Research Methods, 7*, 3-26.

Hong, Y., Chiu, C., Dweck, C. S., Lin, D. M., & Wan, W. (1999). Implicit theories, attributions, and coping: A meaning system approach. *Journal of Personality and Social Psychology, 77*, 588-599.

Horwich, P. (1982). *Probability and Evidence*. Cambridge: Cambridge University Press.

Huff, A. S., & Jenkins, M. (Eds.) (2002). *Mapping Strategic Knowledge*. London: SAGE Publications.

Huntley, J. E., & Costanzo, M. (2003). Sexual harassment stories: Testing a story-mediated model of juror decision-making in civil litigation. *Law and Human Behavior, 27*, 29-51.

Jensen, A., & Weisz, J. R. (2002). Assessing match and mismatch between practitioner-generated and standardized interview-generated diagnoses for clinic-referred children and adolescents. *Journal of Consulting and Clinical Psychology, 70*, 158-168.

Jensen, F. (1996). *An introduction to Bayesian networks*. London: UCL Press.

Kazdin, A. E. (2008). Evidence-based treatment and practice: New opportunities to bridge clinical research and practice, enhance the knowledge base, and improve patient care. *American Psychologist, 63*, 146-159.

Keil, F. C. (2003). Categorisation, causation, and the limits of understanding. *Language and Cognitive Processes, 18*, 663-692.

Keil, F. C. (2006). Explanation and understanding. *Annual Review of Psychology, 57*, 227-254.

Kendall, P. C., Aschenbrand, S. G., & Hudson, J. L. (2003). Child-focused treatment of anxiety. In A. E. Kazdin & J. R. Weisz (Eds.), *Evidence-Based Psychotherapies for Children and Adolescents* (pp. 81-100). New York: Guilford Press.

Kendall, P. C., & Beidas, R. S. (2007). Smoothing the trail for dissemination of evidence-based practices for youth: Flexibility within fidelity. *Professional Psychology: Research and Practice, 38*, 13-20.

Kendall, P. C., Holmbeck, G., Verduin, T. (2004). Methodology, Design, and Evaluation in Psychotherapy Research. In M. J. Lambert (Ed.), *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change* (5th ed.) (pp. 16-43). New York: John Wiley & Sons, Inc.

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

Kendjelic, E. M., & Eells, T. D. (2007). Generic psychotherapy case formulation training improves formulation quality. *Psychotherapy: Theory, Research, Practice, Training, 44*, 66-77.

Kiesler, D. J. (1999). *Beyond the Disease Model of Mental Disorders*. Westport: Praeger.

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.

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

Kim, N. S., Yopchick, J. E., & De Kwaadsteniet, L. (2008). Causal diversity effects in information seeking. *Psychonomic Bulletin & Review, 15*, 81-88.

Kincannon, A., & Spellman, B. A. (2003). The use of category and similarity information in limiting hypotheses. *Memory & Cognition, 31*, 114-132.

Klayman, J., & Brown, K. (1993). Debias the environment instead of the judge: An alternative approach to reducing error in diagnostic (and other) judgment. *Cognition, 49*, 97-122.

Klein, G. (1997). The recognition-primed decision (RPD) model: Looking back, looking forward. In C. E. Zsambok & G. Klein (Eds.), *Naturalistic decision making* (pp. 285-292). Mahwah, NJ: Lawrence Erlbaum Associates.

Koster van Groos, G. A. S. (2003). *Beknopte Handleiding bij de Diagnostische Criteria van de DSM-IV-TR.* Translated from *DSM-IV-TR: Diagnostic and Statistical Manual of Mental Disorders* (4th ed., rev.), American Psychiatric Association (2000). Lisse: Swets & Zeitlinger.

Krol, N. P., De Bruyn, E. E., & Van den Bercken, J. H. (1992). Diagnostic classification by experts. *Acta Psychologica, 81*, 23-37.

Krol, N. P., Morton, J., & De Bruyn, E. E. (2004). Theories of conduct disorder: A causal modelling approach. *Journal of Child Psychology and Psychiatry, 45*, 727-742.

Kuyken, W., Fothergill, C. D., Musa, M., & Chadwick, P. (2005). The reliability and quality of cognitive case formulation. *Behaviour Research and Therapy, 43*, 1187-1201.

Lambert, M. J. (2001). Psychotherapy outcome and quality improvement: Introduction to the special section on patient-focused research. *Journal of Consulting and Clinical Psychology, 69*, 147-149.

Lambert, M. J., Bergin, A. E., & Garfield, S. L. (2004). Introduction and historical overview. In M. J. Lambert (Ed.), *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change* (5th ed.) (pp. 3-15). New York: John Wiley & Sons, Inc.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159-174.

Langston, M., & Trabasso, T. (1999). Modelling causal integration and availability of information during comprehension of narrative texts. In H. van Oostendorp & S. R. Goldman (Eds.), *The Construction of Mental Representations During Reading* (pp.29-69). Mahwah, NJ: Lawrence Erlbaum Publishers, Inc.

Leon, J. A., & Perez, O. (2001). The influence of prior knowledge on the time course of clinical diagnosis inferences: a comparison of experts and novices. *Discourse Processes, 31*, 187-213.

Lichtenberg, J. W. (1997). Expertise in counseling psychology: A concept in search of support. *Educational Psychology Review, 9*, 221-238.

Lo, Y., Sides, A., Rozelle, J., & Osherson, D. (2002). Evidential diversity and premise probability in young children's inductive judgment. *Cognitive Science, 26*, 181-206.

López, A. (1995). The diversity principle in the testing of arguments. *Memory & Cognition, 23*, 374-382.

López, A., Atran, S., Coley, J. D., Medin, D. L., & Smith, E. E. (1997). The tree of life: Universal and cultural features of folkbiological taxonomies and inductions. *Cognitive Psychology, 32*, 251-295.

Lundberg, G. A. (1941). Case-studies vs. statistical methods: An issue based on misunderstanding. *Sociometry, 4*, 379-383.

Magliano, J. P. (1999). Revealing inference processes during text comprehension. In S. R. Goldman, A. C. Graesser, & P. Van den Broek (Eds.), *Narrative Comprehension, Causality, and Coherence: Essays in Honor of Tom Trabasso* (pp. 55-76). Mahwah, NJ: Lawrence Erlbaum Associates.

Mauchnik, J., Schmahl, C., & Bohus, M. (2005). New findings in the biology of borderline personality disorder. *Directions in Psychiatry, 25*, 197-215.

McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review, 99*, 440-466.

Medin, D. L., Coley, J. D., Storms, G., & Hayes, B. K. (2003). A relevance theory of induction. *Psychonomic Bulletin & Review, 10*, 517-532.

Medin, D. L., & Rips, L. J. (2005). Concepts and categories: Memory, meaning, and metaphysics. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge Handbook of Thinking and Reasoning* (pp.37-72). Cambridge: Cambridge University Press.

Meehl, P. E. (1954). *Clinical Versus Statistical Prediction*. Minneapolis, MN: University of Minneapolis Press.

Meehl, P. E., & Grove, W. M. (2006). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. In N. G. Waller, L. J. Yonce, W. M. Grove, D. Faust, & M. F. Lenzenweger (Eds.), *A Paul Meehl Reader: Essays on the Practice of Scientific Psychology* (pp.291-320). Mahwah, NJ: Lawrence Erlbaum Publishers, Inc.

Morton, J. (2004). *Understanding Developmental Disorders: A Causal Modelling Approach*. Malden, MA: Blackwell Publishing.

Murphy, G. L. (2002). *The Big Book of Concepts*. Cambridge, MA: MIT Press.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92*, 289-316.

Murphy, G. L., & Wright, J. C. (1984). Changes in conceptual structure with expertise: Differences between real-world experts and novices. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 10*, 144-155.

Nadkarni, S., & Shenoy, P. P. (2004). A causal mapping approach to constructing Bayesian networks. *Decision Support Systems, 38*, 259-281.

Nelson, T. D., & Steele, R. G. (2008). Influences on practitioner treatment selection: Best research evidence and other considerations. *The Journal of Behavioral Health Services & Research, 35*, 170-178.

Nelson-Gray, R. O. (2003). Treatment utility of psychological assessment. *Psychological Assessment, 15*, 521-531.

Nezu, A. M., & Nezu, C. M. (1993). Identifying and selecting target problems for clinical interventions: A problem solving model. *Psychological Assessment, 5*, 254-263.

Ninan, T. N., & Dunlop, B. W. (2005). Neurobiology and etiology of panic disorder. *Journal of Clinical Psychiatry, 66 (supplement 4),* 3-7.

Nisbett, R.E., & Wilson, T.D. (1977). Telling more than we can know: verbal reports on mental processes. *Psychological Review, 84*, 231-259.

Norman, G., Eva, K., Brooks, L., & Hamstra, S. (2006). Expertise in medicine and surgery. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge Handbook of Expertise and Expert Performance* (pp. 339-353). Cambridge: Cambridge University Press.

Nurcombe, B. (2000). Clinical decision making in psychiatry. In M. H. Ebert, P. T. Loosen, & B. Nurcombe (Eds.), *Current Diagnosis and Treatment in Psychiatry* (pp.115-120). Singapore: Lange Medical Books/McGraw-Hill.

Nurcombe, B., Drell, M. J., Leonard, H. L., & McDermott, J. F. (2002). Clinical problem solving: The case of Matthew, Part I. *Journal of the American Academy of Child and Adolescent Psychiatry, 41*, 92-97.

Nurcombe, B., Drell, M. J., Leonard, H. L., & McDermott, J. F. (2002). Clinical problem solving: The case of Matthew, Part III. *Journal of the American Academy of Child and Adolescent Psychiatry, 41*, 344-353.

Nurcombe, B., & Fitzhenry-Coor, I. (1982). How do psychiatrists think? Clinical reasoning in the psychiatric interview: A research and education project. *Australian and New Zealand Journal of Psychiatry, 16*, 13-24.

Nurcombe, B., & Gallagher, R. M. (1986). *The Clinical Process in Psychiatry*. Cambridge: Cambridge University Press.

Osherson, D. N., Wilkie, O., Smith, E., & López, A. (1990). Category-based induction. *Psychological Review, 97*, 185-200.

Oskamp, S. (1965). Overconfidence in case-study judgments. *Journal of Consulting Psychology, 29*, 261-265.

Ostrander, R., & Herman, K. (2006). Potential cognitive, parenting, and developmental mediators of the relationship between ADHD and depression. *Journal of Consulting and Clinical Psychology, 74*, 89-98.

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge, UK: Cambridge University Press.

Pennington, N., & Hastie, R. (1986). Evidence evaluation in complex decision making. *Journal of Personality and Social Psychology, 51*, 242-258.

Pennington, N., & Hastie, R. (1988). Explanation-based decision making: Effects of memory structure on judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 521-533.

Pennington, N., & Hastie, R. (1993). A theory of explanation-based decision making. In G.A. Klein, et al. (Eds.), *Decision making in action: models and methods* (pp. 188-201). Norwood, NJ: Ablex.

Pennington, N., & Hastie, R. (1994). Reasoning in explanation-based decision making. In P. N. Johnson-Laird & E. Shafir (Eds.), *Reasoning and Decision Making* (pp. 123-163). Cambridge: Blackwell Publishers.

Persons, J. B., & Bertagnolli, A. (1999). Inter-rater reliability of cognitive-behavioral case formulations of depression: A replication. *Cognitive Therapy and Research, 23*, 271-283.

Pliske, R., & Klein, G. A. (2003). The naturalistic decision-making perspective. In S. Schneider & J. Shanteau (Eds.), *Emerging Perspectives on Judgment and Decision Research* (pp. 165-200). Cambridge: Cambridge University Press.

Popper, K. R. (1963). *Conjectures and Refutations: The Growth of Scientific Knowledge*. London: Routledge.

Proctor, C. & Ahn, W. (2007). The effect of causal knowledge on judgments of the likelihood of unknown features. *Psychonomic Bulletin & Review, 14*, 635-639.

Proffitt, J. B., Coley, J. D., & Medin, D. L. (2000). Expertise and category-based induction. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 811-824.

Read, S. J., & Marcus-Newhall, A. R. (1993). Explanatory coherence in social explanations: A parallel distributed processing account. *Journal of Personality and Social Psychology, 65*, 429-447.

Rehder, B. (2007). Essentialism as a generative theory of classification. In A. Gopnik & L. Schulz (Eds.), *Causal Learning: Psychology, Philosophy, and Computation* (pp. 190-207). Oxford: Oxford University Press.

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

Rettinger, D. A., & Hastie, R. (2003). Comprehension and decision making. In S. Schneider & J. Shanteau (Eds.), *Emerging Perspectives on Judgment and Decision Research* (pp. 165-200). Cambridge: Cambridge University Press.

Rikers, R. M. J. P., Loyens, S. M. M., & Schmidt, H. G. (2004). The role of encapsulated knowledge in clinical case representations of medical students and family doctors. *Medical Education, 38*, 1035-1043.

Robins, R. W., Mendelsohn, G. A., Connell, J. B., & Kwan, V. S. Y. (2004). Do people agree about the causes of behavior? A social relations analysis of behavior ratings and causal attributions. *Journal of Personality and Social Psychology, 86*, 334-344.

Rozenblit, L., & Keil, F. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science, 26*, 521-562.

Rumelhart, D. E. (1975).Notes on a schema for stories. In D. G. Bobrow & A. Collins (Eds.), *Representation and Understanding: Studies in Cognitive Science* (pp. 211-236). New York: Academic Press.

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.

Sarbin, T. R. (1944). The logic of prediction in psychology. *Psychological Review, 51*, 210-228.

Schmidt, H. G., & Boshuizen, H. P. A. (1993). On the origin of intermediate effects in clinical case recall. *Memory & Cognition, 21*, 338-351.

Schulz, L., Kushnir, T., & Gopnik, A. (2007). Learning from doing: Intervention and causal inference. In A. Gopnik & L. Schulz (Eds.), *Causal Learning: Psychology, Philosophy, and Computation* (pp. 67-85). Oxford: Oxford University Press.

Scott, M. M., & Hatfield J. C. (1985). Problems of analyst and observer agreement in naturalistic narrative data. *Journal of Educational Measurement, 22*, 207-218.

Siegel, S., & Castellan, N. J. (1988). *Nonparametric Statistics for the Behavioral Sciences* (2nd ed.). New York: Mc Graw-Hill.

Sloman, S. A. (2005). *Causal Models: How People Think About the World and its Alternatives.* New York: Oxford University Press.

Sloman, S. A., & Hagmayer, Y. (2006). The causal psycho-logic of choice. *Trends in Cognitive Sciences,* 10, 407-412.

Sloman, S. A., Love, B. C., & Ahn, W. (1998). Feature centrality and conceptual coherence. *Cognitive Science, 22*, 189-228.

Sonuga-Barke, E. J. S. (2002). Psychological heterogenity in AD/HD – A dual pathway model of behaviour and cognition. *Behavioural Brain Research, 130,* 29-36.

Spellman, B. A., López, A., & Smith, E. E. (1999). Hypothesis testing: Strategy selection for generalising versus limiting hypotheses. *Thinking & Reasoning, 5*, 67-91.

Staines, G. L. (2007). Comparative outcome evaluations of psychotherapies: Guidelines for addressing eight limitations of the gold standard of causal inference. *Psychotherapy: Theory, Research, Practice, Training, 44*, 161-174.

Stinson, C. H., Milbrath, C., Reidbord, S. P., & Bucci, W. (1994). Thematic segmentation of psychotherapy transcripts for convergent analyses. *Psychotherapy, 31*, 36-48.

Strevens, M. (2007). Why represent causal relations? In A. Gopnik & L. Schulz (Eds.), *Causal Learning: Psychology, Philosophy, and Computation* (pp. 245-260). Oxford: Oxford University Press.

Tenenbaum, J. B., Griffiths, T. L., & Niyogi, S. (2007). Intuitive theories as grammars for causal inference. In A. Gopnik & L. Schulz (Eds.), *Causal Learning: Psychology, Philosophy, and Computation* (pp. 301-322). Oxford: Oxford University Press.

Thagard, P. (1992). *Conceptual Revolutions*. Princeton, NJ: Princeton University Press.

Trabasso, T., & Magliano, J. P. (1996). Conscious understanding during comprehension. *Discource Processes, 21*, 255-287.

Van den Broek, P. & Gustafson, M. (1999). Comprehension and memory for texts: Three generations of reading research. In S. R. Goldman, A. C. Graesser, & P. van den Broek (Eds.), *Narrative Comprehension, Causality, and Coherence: Essays in Honor of Tom Trabasso* (pp. 15-349). Mahwah, NJ: Lawrence Erlbaum Publishers, Inc.

Van de Wiel, M. W. J., Boshuizen, H. P. A., & Schmidt, H. G. (2000). Knowledge restructuring in expertise development: Evidence from pathophysiological representations of clinical cases by students and physicians. *European Journal of Cognitive Psychology, 12*, 323-355.

Van Someren, M. W., Barnard, Y. F., & Sandberg, J. A. C. (1994). *The Think Aloud Method: A Practical Guide to Modelling Cognitive Processes*. London: Academic Press.

Waldmann, M. R. (1996). Knowledge-based causal induction. In D. R. Shanks, D. L. Medin, & K. J. Holyoak (Eds.), *Causal Learning* (pp. 47-88). San Diego: Academic Press.

Weisz, J. R., Jensen-Doss, A., & Hawley, K. M. (2006). Evidence-based youth psychotherapies versus usual clinical care: A meta-analysis of direct comparisons. *American Psychologist, 61*, 671-689.

Weisz, J. R., & Kazdin, A. E. (2003). Concluding thoughts: Present and future of evidence-based psychotherapies for children and adolescents. In A. E. Kazdin & J. R. Weisz (Eds.), *Evidence-Based Psychotherapies for Children and Adolescents* (pp. 439-451). New York: Guilford Press.

Wellman, H. M., & Liu, D. (2007). Causal reasoning as informed by the early development of explanations. In A. Gopnik & L. Schulz (Eds.), *Causal Learning: Psychology, Philosophy, and Computation* (pp. 261-279). Oxford: Oxford University Press.

Wenar, C., & Kerig, P. (2006). *Developmental Psychopathology: From Infancy through Adolescence* (5th ed.). New York: McGraw-Hill, Inc.

Westen, D., & Weinberger, J. (2005). In praise of clinical judgment: Meehl's forgotten legacy. *Journal of Clinical Psychology, 61*, 1257-1276.

Whewell, W. (1840/1999). *The Philosophy of the Inductive Sciences, Founded Upon Their History*. Bristol: Thoemmes Press.

Whitten, S., & Graesser, A. C. (2003). Comprehension in problem solving. In J. E. Davidson, & R. J. Sternberg (Eds.), *The Psychology of Problem Solving* (pp. 207-229). Cambridge: Cambridge University Press.

Widiger, T. A., & Clark, L. A. (2000). Toward *DSM-V* and the classification of psychopathology. *Psychological Bulletin, 126*, 946-963.

Witteman, C. L. M., Harries, C., Bekker, H. L., & Van Aarle, E. J. M. (2007). Evaluating psychodiagnostic decisions. *Journal of Evaluation in Clinical Practice, 13*, 10-15.

Witteman, C. L. M., & Koele, P. (1999). Explaining treatment decisions. *Psychotherapy Research, 9*, 100-114.

Witteman, C. L. M., & Kunst, H. (1997). Planning the treatment of a depressed patient. *Clinical Psychology and Psychotherapy, 4*, 157-171.

Witteman, C. L. M., & Van den Bercken, J. H. L. (2007). Intermediate effects in psychodiagnostic classification. *European Journal of Psychological Assessment, 23*, 56-61.

Wood, J. M., Garb, H. N., Lilienfield, S. O., & Nesworski, M. T. (2002). Clinical assessment. *Annual Review of Psychology, 53*, 519-543.

Woods, N. N., Brooks, L. R., & Norman, G. R. (2005). The value of basic science in clinical diagnosis: creating coherence among signs and symptoms. *Medical Education, 39*, 107-112.

Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.

Woodward, J. (2007). Interventionist theories of causation in psychological perspective. In A. Gopnik & L. Schulz (Eds.), *Causal Learning: Psychology, Philosophy, and Computation* (pp. 19-36). Oxford: Oxford University Press.

Yopchick, J. E., & Kim, N. S. (2007). *The Influence of Causal Information on Treatment Choice*. Poster presented at the annual meeting of the Psychonomic Society, Long Beach, CA.

Zwaan, R. A. & Brown, C. M. (1996). The influence of language proficiency and comprehension skill on situation-model construction. *Discourse Processes, 21*, 289-327.

Samenvatting

Clinici als mecaniciens?
Causaal redeneren in klinisch oordelen en beslissen

Oordelen van clinici, zoals voorspellingen van gedrag en het toekennen van diagnoses, blijken vaak minder accuraat dan oordelen die tot stand komen door het combineren van informatie met behulp van statistische methoden (Ægisdóttir et al., 2006; Grove, Zald, Lebow, Snitz, & Nelson, 2000). Ook blijken oordelen van ervaren clinici niet beter te zijn dan die van beginnende clinici (Eells, Lombart, Kendjelic, Turner, & Lucas, 2005; Garb, 1998). Onderzoek naar klinische oordeelsvorming is dan ook relevant om na te gaan hoe dit proces verbeterd kan worden (Caspar, 1997; Witteman, Harries, Bekker, & Van Aarle, 2007). Dit proefschrift bevat een aantal studies waarin is gekeken naar causaal redeneren door clinici.

In 1954 gaf Meehl een beschrijving van hoe een ervaren clinicus tot oordelen komt. Meehl vergeleek de clinicus met een mecanicien, die met behulp van kennis van werkingsmechanismen accurate voorspellingen kan doen. Met deze vergelijking vatte Meehl het denken van de clinicus op als een causaal redeneerproces. Deze beschrijving komt overeen met wat klinische handboeken voorschrijven. Clinici dienen een compleet beeld te vormen van hoe verschillende biologische, psychologische en sociale factoren samen de problemen van een cliënt veroorzaken en in stand houden, en op basis daarvan interventies te kiezen (zie bijv. Carr, 2006; De Bruyn, Ruijssenaars, Pameijer, & Van Aarle, 2003; Eells, 2007; Groth-Marnat, 2003; Kuyken, Fothergill, Musa, & Chadwick, 2005; Wenar & Kerig, 2006).

Causaal redeneren wordt verondersteld fundamenteel te zijn in het denken van mensen. Door het construeren van causale representaties van informatie kunnen mensen niet alleen voorspellen wat er zal gebeuren, maar vooral ook welke acties het meest effectief zullen zijn om een gesteld doel te bereiken (Pearl, 2000; Sloman, 2005; Woodward, 2003).Verschillende onderzoeken hebben aangetoond dat mensen inderdaad informatie verwerken in causale representaties, bijvoorbeeld bij het lezen van teksten, of het verwerken van informatie in een rechtzaak om tot een oordeel te komen (zie bijv. Graesser, Olde, & Kletke, 2002; Hastie & Pennington, 2000). Ook is gebleken dat mensen op basis van causale representaties besluiten welke interventies effectief zullen zijn, bijvoorbeeld om ziekten of milieuvervuiling te voorkomen (zie bijv. Böhm & Pfister, 2000; Green & McManus, 1995; Pliske & Klein, 2003). Zelfs classificaties, indelingen van onderwerpen in een categorie, blijken mensen te baseren op causale representaties (zie bijv. Ahn, 1998).

Er is weinig onderzoek gedaan naar de wijze waarop clinici tot hun oordelen en beslissingen komen. Op basis van bovenstaande beschrijvingen en bevindingen veronderstellen we in dit proefschrift dat clinici door middel van causale representaties cliënten proberen te begrijpen en keuzen maken voor effectieve interventies. In hoofdstuk 2 tot en met 5 worden studies beschreven die met verschillende onderzoeksmethoden de rol van causale representaties in klinische oordeelsvorming onderzoeken.

Methoden en resultaten

In hoofdstuk 2 en 3 is beschreven hoe we causale denkprocessen van clinici hebben onderzocht door analyse van de uitingen van clinici die zij deden terwijl zij hardop nadachten over een cliënt. Hardop-denken is een methode om inzicht te krijgen in denkprocessen die zich voordoen wanneer mensen bijvoorbeeld een probleem moeten oplossen (Ericsson, 2006; Ericsson & Simon, 1993). Twintig clinici kregen een video van een intakegesprek van een negenjarige jongen met zijn ouders te zien en de resultaten van negen verschillende tests met observaties die daarbij zijn gedaan. Tijdens en na het bekijken van de video en het bestuderen van de testresultaten dachten de clinici hardop na over de vragen: “Wat denkt u dat er aan de hand is met deze jongen?” en “Wat zou u voorstellen als vervolgstappen?”.

Mentale representaties

In hoofdstuk 2 hebben we de uitingen geanalyseerd van clinici nadat ze de video hadden bekeken en de testresultaten hadden bestudeerd, om de structuur en de inhoud te achterhalen van de mentale representaties die de clinici hadden gevormd. We verwachtten dat clinici coherente causale representaties van cliëntinformatie zouden vormen en dat zij vervolgstappen zouden baseren op deze representaties. Om de structuur te achterhalen van de mentale representaties hebben we eerst onderscheid gemaakt tussen uitingen over de cliënt en uitingen over vervolgstappen. Vervolgens hebben we de causale en temporele verbindingen geïdentificeerd die clinici legden tussen uitingen over de cliënt en daarna de causale, temporele en redengevende verbindingen tussen uitingen over de cliënt en uitingen over vervolgstappen. Deze verbindingen hebben we tenslotte op grafische wijze weergegeven in causale modellen. Het bleek dat de meeste uitingen over de cliënt niet causaal met elkaar verbonden waren. De causale representaties die we hadden gereconstrueerd bestonden vaak uit verschillende losse delen en waren weinig complex. Verder bleek slechts een derde van de vervolgstappen die clinici voorstelden verbonden aan deze representaties. Opmerkelijk was dat bijna de helft van de vervolgstappen aan geen enkele uiting over de cliënt verbonden was.

Om de inhoud van de mentale representaties te analyseren hebben we de “Case Formulation Content Coding Method” gebruikt van Eells, Kendjelic, Lucas en Lombart (zie Eells, Lombart, Kendjelic, Turner, & Lucas, 2005). We hebben gekeken of de uitingen beschrijvende, classificerende, of verklarende informatie bevatten, of gingen over vervolgstappen. Vervolgens zijn we nagegaan welke verklarende factoren clinici noemden en welke vervolgstappen. De meeste uitingen bleken beschrijvend en niet verklarend te zijn. Verder vonden we weinig overeenkomst in de verklarende factoren die clinici noemden of in de vervolgstappen die zij voorstelden.

In tegenstelling tot onze verwachtingen, maar in overeenstemming met onderzoek naar de inhoud van case formulations (Eells, Kendjelic, & Lucas, 1998; Kuyken et al., 2005), lijken clinici cliëntinformatie dus slechts in beperkte mate causaal te representeren en zijn causale representaties die clinici vormen weinig coherent. Ook lijken ze hun vervolgstappen grotendeels niet te baseren op deze causale representaties.

Causaal redeneren tijdens informatieverwerking

In hoofdstuk 3 hebben we de uitingen geanalyseerd van clinici tijdens het bekijken van de intakevideo. Onze veronderstelling was dat indien clinici proberen te komen tot een coherent causaal model, zij tijdens het verwerken van informatie voornamelijk causale afleidingen zullen maken (cf. Trabasso & Magliano, 1996). Ook veronderstelden wij dat clinici vooral naar meer informatie zouden vragen over mogelijke verklaringen voor de problemen van een cliënt.

We zijn nagegaan in hoeverre clinici informatie uit de video herhaalden, in hoeverre zij nieuwe informatie afleidden (toevoegden) en in hoeverre zij vragen stelden of plannen formuleerden voor verder onderzoek. Voor de afleidingen die de clinici maakten hebben we vastgesteld of deze beschrijvend, classificerend, verklarend of voorspellend waren. Voor alle vragen en onderzoeksplannen hebben we vastgesteld of deze om beschrijvende, classificerende, of verklarende of informatie vroegen of naar verder verloop. Verder hebben we gekeken naar de specifieke inhoud van de classificerende en verklarende uitingen, met behulp van de CFCCM (Eells et al., 2005).

In tegenstelling met onze verwachtingen maar in overeenstemming met de bevindingen in hoofdstuk 2 vonden we dat de meeste afleidingen die clinici maakten niet verklarend maar beschrijvend waren. Ook werd meestal niet gevraagd naar verklarende factoren, maar naar meer beschrijvende informatie. Verder vonden we dat verschillende clinici al bij het intakegesprek verschillende causale factoren naar voren brachten.

De rol van causale representaties bij de selectie van effectieve interventies

In hoofdstuk 4 hebben we met een andere methode, “cognitive mapping”, onderzocht of clinici hun interventiekeuzen baseren op causale cliëntrepresentaties. Cognitive mapping is een methode om te achterhalen hoe mensen een probleem oplossen, door ze te laten tekenen hoe zij denken dat belangrijke factoren voor hun beslissing causaal gerelateerd zijn (Axelrod, 1976; Hodgkinson, Maule, & Bown, 2004). We vroegen veertig clinici om twee cliëntbeschrijvingen te lezen en voor iedere cliënt een causaal model (“cognitive map”) te tekenen waarin ze aangaven hoe zij dachten dat de problemen van de cliënt werden veroorzaakt en in stand gehouden. Ook vroegen we ze om voor iedere cliënt interventies te rangordenen op verwachte effectiviteit. Om voor mogelijke volgorde-effecten te controleren vroegen we de helft van de clinici modellen tekenen voordat zij interventies rangordenden, de andere helft deed deze taken in omgekeerde volgorde. We hebben berekend hoe effectief de interventies waren volgens de causale modellen van de clinici en we hebben deze uitkomsten vergeleken met de rangordeningen die de clinici zelf hadden gegeven. Ook hebben we de rangordeningen van de clinici vergeleken met rangordeningen gebaseerd op de effectiviteitbeoordelingen van alle deelnemers samen.

De rangordeningen voor verwachte effectiviteit van interventies die door clinici zelf waren gegeven bleken redelijk goed overeen te komen met de effectiviteit van de interventies zoals wij die hadden berekend uit de individuele causale modellen die de clinici hadden getekend. De door clinici gegeven rangordeningen hingen bovendien meer samen met de effectiviteit van interventies afgeleid uit hun persoonlijke causale modellen dan met de rangordeningen gebaseerd op de gezamenlijke beoordelingen. Uit analyse van de inhoud van de causale modellen bleek verder dat clinici verschillende factoren in hun modellen

gebruikten en verschillende causale relaties tekenden tussen factoren die ze gemeenschappelijk hadden. Ook bleek de overeenstemming over de effectiviteit van interventies laag tot matig.

Clinici lijken dus de verwachte effectiviteit van interventies voor een cliënt af te leiden uit causale representaties die zij vormen voor die cliënt. Deze causale representaties blijken verschillend voor verschillende clinici en kunnen verschillen in oordelen over de effectiviteit van interventies verklaren.

Causale diversiteitseffecten in het redeneren over stoornissen

In de studies in hoofdstuk 5 hebben we gekeken naar de invloed van causale representaties van stoornissen op het redeneren van clinici. We hebben onderzocht of clinici een "causaal diversiteitseffect" laten zien wanneer zij redeneren over psychiatrische stoornissen. Leken op medisch gebied blijken meer waarde te hechten aan symptomen die op verschillende wijze ontstaan uit een hoofdoorzaak voor een medische aandoening, dan aan symptomen die op gelijke wijze ontstaan uit die hoofdoorzaak, wanneer zij moeten beoordelen hoe waarschijnlijk een patiënt een medische aandoening heeft (Kim & Keil, 2003: zie ook Kim, Yopchick, & De Kwaadsteniet, 2008).

In de eerste studie van hoofdstuk 5 hebben we 19 ervaren clinici informatie gegeven over hoe volgens recente wetenschappelijke theorieën symptomen op verschillende wijzen ontstaan uit een hoofdoorzaak, voor verschillende psychiatrische stoornissen. Vervolgens vroegen we hen om aan de hand van de gepresenteerde theorieën te kiezen wie van twee cliënten het meest waarschijnlijk een stoornis had. Eén cliënt toonde twee symptomen die op gelijke wijze ontstaan uit de hoofdoorzaak, de andere cliënt toonde twee "causaal diverse symptomen": symptomen die op verschillende manieren uit de hoofdoorzaak ontstaan. De deelnemers bleken vaker te kiezen voor cliënten met causaal diverse symptomen, maar alleen wanneer zij eerst tussen dezelfde cliënten konden kiezen zonder gebruik te maken van de gepresenteerde theorieën, of als ze wisten dat ze dat later zouden kunnen. Blijkbaar beïnvloedde hun eigen kennis hun keuzen.

In de tweede studie hebben we 17 ervaren clinici gevraagd om voor verschillende stoornissen te tekenen hoe zij dachten dat verschillende symptomen uit één hoofdoorzaak ontstaan. Op basis van deze individuele causale modellen hebben we vragen opgesteld met cliëntparen waarbij één cliënt symptomen had die volgens de clinicus op gelijke wijze uit de hoofdoorzaak ontstaan en één cliënt symptomen die volgens de clinicus op verschillende wijzen uit de hoofdoorzaak ontstaan voor een stoornis. We vroegen de clinici te kiezen welke cliënt het meest waarschijnlijk de stoornis had. Zij bleken niet vaker te kiezen voor de cliënt met causaal diverse symptomen. Deze studie hebben we herhaald met masterstudenten en clinici met maximaal drie jaar ervaring. Deze beginners bleken wel vaker te kiezen voor een cliënt met symptomen die in hun individuele model causaal divers waren.

Informatie over verschillen in causale oorsprong van symptomen kan gebruikt worden om af te leiden hoe waarschijnlijk een stoornis is, als kennis over hoe waarschijnlijk een stoornis aanwezig is gegeven de aanwezigheid van specifieke symptomen ontbreekt. Deze kennis zullen beginners in tegenstelling tot ervaren clinici niet hebben (cf. Horwich, 1982; Kincannon & Spellman, 2003). Verschillen in causale oorsprong van symptomen zullen daardoor informatiever zijn voor beginners dan voor ervaren clinici.

Discussie

De resultaten uit de hardop-denkenstudie en uit de cognitive-mappingstudie lijken niet direct verenigbaar. Hoewel we met de causale modellen die clinici tekenden konden voorspellen welke interventies zij het meest effectief zouden vinden, bleek in de hardop-denkprotocollen de mate van causaal redeneren beperkt te zijn. Ook bleken in deze protocollen vervolgstappen die clinici voorstelden vaak niet gebaseerd op causale representaties van cliëntinformatie. Onze resultaten over de complexiteit van causale representaties die clinici construeren en de mate waarin clinici vervolgstappen op deze representaties baseren lijken afhankelijk te zijn van de gebruikte onderzoeksmethoden.

Mogelijk leiden de methoden in verschillende mate tot het expliciteren van mentale causale representaties. Misschien vinden we in hardop-denkprotocollen minder aanwijzingen voor causaal redeneren dan er daadwerkelijk plaatsvindt, doordat clinici mogelijk niet alle gedachten verwoorden die zij hebben over causale relaties, bijvoorbeeld wanneer verklaringen automatisch gegenereerd worden (Ericsson & Simon, 1993). Bij cognitive mapping wordt de uitgebreidheid van causaal redeneren misschien juist overschat, door de expliciete instructie (cf. Hodgkinson, Maule & Bown, 2004). Ook kan het op papier zetten van causale representaties het causaal redeneren faciliteren (cf. Morton, 2004).

Een gemeenschappelijke bevinding in de hoofdstukken 2, 3 en 4 is de grote verscheidenheid in de causale representaties van clinici voor dezelfde cliënt. Deze verscheidenheid bleek al aanwezig tijdens een intakegesprek. Een mogelijke verklaring hiervoor is de incompleetheid van kennis over hoe biologische, psychologische en sociale factoren samen kunnen leiden tot psychologische problemen en gedragsproblemen (Kiesler, 1999; Cicchetti & Sroufe, 2000). Ook ontbreekt kennis over wat behandelingen precies effectief maakt (Bieling & Kuyken, 2003; De Los Reyes & Kazdin, 2006). Daarnaast kunnen verschillen in opleiding, werkervaring en theoretische voorkeur leiden tot verschillende opvattingen over de causale factoren die een rol spelen bij de problemen van clienten.

Naast dat kennis incompleet is, zal de veelheid aan factoren die in interactie het ontstaan en in stand houden van problemen beïnvloeden (De Los Reyes & Kazdin, 2006; Haynes, 1992; Morton, 2004) het moeilijk maken voor clinici om tot een compleet en accuraat beeld te komen van hoe problemen van cliënten ontstaan en in stand worden gehouden. Accurate causale cliëntrepresentaties zullen complex zijn. Verder is het de vraag in hoeverre clinici in de praktijk coherente, complete causale representaties nodig hebben om te kunnen besluiten over interventies. Clinici maken interventiekeuzen mogelijk vooral op basis van beschrijvingen of classificaties (zie bijvoorbeeld Clarke, DeBar, & Lewinsohn, 2003; Kendall, Aschenbrand, & Hudson, 2003). Groenier, Pieters, Hulshof, Wilhelm en Witteman (2008) vonden dat clinici het zoeken naar verklaringen voor problemen minder belangrijk vinden dan het beschrijven van problemen en kiezen van interventies.

Uit de studies in hoofdstuk 5 blijkt dat de invloed van causale representaties van stoornissen op het redeneren van clinici verschillend is voor clinici met meer en minder ervaring. Met meer ervaring herkennen clinici wellicht patronen en gebruiken zij in mindere mate causale representaties om tot oordelen te komen. Het is echter niet duidelijk of de ervaren clinici in de studies beschreven in hoofdstuk 5 werkelijk in mindere mate afgaan op causale representaties van stoornissen dan beginners, of dat hun causale kennis over

stoornissen op een andere wijze is georganiseerd dan bij beginners. Clinici tekenden een geïsoleerd deel van hun theorieën in een vooraf gegeven vorm, wat dus slechts een beperkt beeld kan geven van de causale structuur van hun kennis. We vonden in de andere studies geen verschillen in de mate van causaal redeneren tussen meer en minder ervaren clinici. Wel is uit onderzoek naar case formulations gebleken dat meer ervaren en beter gekwalificeerde clinici meer complexe en omvattende case formulations vormen dan beginners en dat hun interventiekeuzen daar beter op aansluiten (Eells & Lombart, 2003; Eells et al., 2005; Kuyken et al., 2005).

Implicaties en verder onderzoek

Het is niet duidelijk geworden in welke mate clinici causaal redeneren. Daarom is verder onderzoek nodig naar causale redeneerprocessen van clinici, bijvoorbeeld met meer experimentele methoden (Einhorn et al., 1979; cf. Gaines, Brown, & Doyle, 1996; Magliano, 1999). Daarnaast hebben we niet gekeken naar andere denkwijzen, zoals probabilistisch redeneren of redeneren aan de hand van schema's of scripts (zie bijv. Hamm, 2003). Clinici zullen verschillende denkwijzen toepassen (Brammer, 1997; Norman, Eva, Brooks, & Hamstra, 2006). Zo zullen clinici causale and probabilistische kennis combineren, omdat oorzaken gewoonlijk met een mate van onzekerheid effecten genereren (Rehder, 2007; Tenenbaum, Griffiths, & Niyogi, 2007). Interessant is te onderzoeken welk effect het verstrekken van relevante causale kennis heeft op causaal redeneren van clinici. Mogelijk gaan clinici meer causaal redeneren, vormen zij meer coherente en complexe causale representaties en vertonen zij meer overeenkomst in hun causale representaties.

Ook is nog niet duidelijk in welke mate interventiekeuzen die clinici maken afhangen van effectiviteitoordelen gebaseerd op causale representaties, of op andere factoren zoals beschikbaarheid van interventies, tijd of kosten (cf. Nelson & Steele, 2008). Wanneer causale representaties inderdaad belangrijk zijn voor interventiekeuzen, kan onderzocht worden of interventiekeuzen beïnvloed kunnen worden door het verstrekken van informatie over de werkingsmechanismen van effectieve interventies. Mogelijk worden interventies wanneer kennis over de werking ervan aanwezig is makkelijker geïntegreerd in causale cliëntrepresentaties dan wanneer niet bekend is waardoor interventies effectief zijn. Door dit te onderzoeken kan wellicht informatie verkregen worden over hoe de implementatie van effectief gebleken interventies gestimuleerd kan worden (cf. Kazdin, 2008).

De vraag is of coherente causale cliëntrepresentaties leiden tot betere interventiekeuzen en uiteindelijk tot betere interventie-uitkomsten. Het identificeren van causale mechanismen lijkt soms maar niet altijd tot betere uitkomsten te leiden (Nelson-Gray, 2003). In eenvoudige gevallen kunnen effectieve interventies vaak worden geselecteerd op basis van een classificatie. In meer complexe gevallen, bijvoorbeeld bij comorbiditeit, of wanneer eerdere behandelingen niet effectief zijn gebleken, zal een clinicus toch moeten proberen om tot een coherent beeld te komen van de oorzaken en in stand houdende factoren voor de problemen van cliënt (Nelson-Gray, 2003). Jensen-Doss en Weisz (2008) vonden dat accuraatheid van classificaties de uitkomst van behandelingen verbeterde. Mogelijk wordt de uitkomst van behandelingen ook beïnvloed door de accuraatheid van causale cliëntrepresentaties van clinici.

Gezien de verscheidenheid in causale representaties die verschillende clinici vormen voor een cliënt en de mogelijke invloed daarvan op de behandeling die clinici kiezen, zou in opleidingen en trainingen aandacht kunnen worden besteed aan het expliciteren en toetsen van veronderstellingen over causale factoren en verbanden. Specifieke methoden kunnen gebruikt worden om deze representaties te visualiseren, zoals “causal modelling” (Morton, 2004). Hoewel visualisatie op zich niet tot betrouwbare en valide modellen leidt, kan visualisatie duidelijk maken waarin veronderstellingen verschillen en overeenkomen en hoe deze het beste getoetst kunnen worden (zie bijv. Haynes, Spain, & Oliveira, 1993; Staines, 2007).

Tenslotte: zo moeilijk als het is voor clinici is om de denkprocessen van cliënten te achterhalen (Meehl, 1954), zo moeilijk zal het voor onderzoekers zijn om de denkprocessen van clinici te achterhalen.

Dankwoord

Dit proefschrift was niet tot stand gekomen zonder de investeringen van mijn begeleiders. Cilia, Nicole en John: ik dank jullie voor het vertrouwen dat jullie mij hebben gegeven, voor al jullie hulp en voor alles wat jullie mij geleerd hebben. En bovenal wil ik jullie bedanken voor de plezierige samenwerking! Cilia, jou wil ik in het bijzonder bedanken omdat je er altijd voor me bent. Nicole, je kunt werkelijk de mooiste complimenten geven. John, altijd al hielp je me, en aan het eind nog eens extra.

Nancy, I felt so happy about our collaboration, and I still do. You taught me so much. Best is your friendship! I will never forget the lovely day with you, Ed and Sophie. We went to the Boston aquarium, to see all the marvellous fish, and to the fish restaurant, to eat one, and above all: we ate cookies looking over Boston from the top floor of a skyscraper. Thanks so much! And Jenelle, thank you too!

York, you seem to know everything of anything. You told me about the history of Starbucks, while ordering our coffee. Funny: York as in New York, on your cup. I really enjoyed our project with the Annemiekes. Thanks a lot, and I hope we will continue our collaboration!

Rowella, bedankt voor al je hulp bij mijn onderzoek, en in het bijzonder ook voor jouw ervaringskennis. Edward, Marieke, Marleen, Chantal, Monique, Judith, José, Martin, Koen en Lieke, dankjewel voor al jullie constructieve bijdragen. En awel Laurence, jij hebt me ook geholpen dit varkentje te wassen!

Zonder deelnemers geen onderzoek. Ik dank alle ervaren diagnostici en studenten voor de gedachten die zij met mij deelden, in het onderzoek en over het onderzoek.

Scriptiestudenten Annemieke, Annemieke, Callista, Emmy, Heleen, Jolin, Karine, Mieke, Miranda, Mirjam, Nancy, en Noortje, dank voor jullie hulp en vele goede ideeën.

Dan zijn er al mijn collega's, met name ook de secretaresses!, mijn familieleden en vrienden, die ik vooral heel graag uitnodig om straks, als alles goed is gegaan, mijn promotie met mij te vieren!

William, hartelijk dank voor de mooie foto op de omslag, van de doos die Jeroen zo mooi heeft gemaakt, met knopjes en lichtjes van Technica.

Monique, zonder jou was ik niet eens begonnen aan dit proefschrift. En hoe minder proefschriftzorgen, hoe groter ons familiegeluk! Ik hoop dat we er zoveel mogelijk van genieten. Harriëtte, ik droom 's nachts over hoe goed je voor onze poezen zorgt en daarmee ook voor ons. Grappig he? Ik ben blij dat ook jij mijn paranimf bent.

En tenslotte wil ik dit graag voorop stellen: Frijke, Sil, Marit, Ulla en Jarne, jullie zijn mijn grote trots! Ik zal niet meer zoveel werken. En Jeroen, zonder jou was dit proefschrift vast veel dikker geworden of juist niet gelukkig maar hahaha.

Dankjulliewel!
Leontien

Curriculum Vitae

Leontien de Kwaadsteniet, born in 1974, finished VWO at the Pius-X-College in Bladel, in 1992. She studied philosophy at the Radboud Universiteit Nijmegen and graduated in logic and philosophy of language, in 2000. Then she worked for two years as junior staff officer ('rijkstrainee') at the Ministry of Agriculture, Nature, and Fishery, at the department for animal welfare.

In 2003 she started her PhD at the Radboud University Nijmegen, Faculty of Social Sciences, in the research group Diagnostic Decision Making. During this project, she collaborated with Nancy Kim at the Northeastern University in Boston, MA, and with York Hagmayer from Georg-August University in Göttingen, Germany. She presented her studies at international conferences (Egproc, International Congress of Psychology, SPUDM, workshops Causal Reasoning in Clinical Decision Making). She also presented her studies to practicing clinicians on several occasions. She supervised master theses of students in Special Education.

Currently she's still working at the same faculty, doing research on clinical decision making and teaching courses on psychometrics and academic skills, and supervising students' master theses.

In 1994 Leontien met the love of her life, Jeroen, and together they have five wonderful children.