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

Leontien de Kwaadsteniet
Radboud University Nijmegen, the Netherlands

York Hagmayer
Georg-August University Goettingen, Germany

Nicole P. C. M. Krol and Cilia L. M. Witteman
Radboud University Nijmegen, the Netherlands

An important reason to choose an intervention to treat psychological problems of clients is the expectation that the intervention will be effective in alleviating the problems. The authors investigated whether clinicians base their ratings of the effectiveness of interventions on models that they construct representing the factors causing and maintaining a client’s problems. Forty clinical child psychologists drew causal models and rank ordered interventions according to their expected effectiveness for 2 cases. The authors found that different clinicians constructed different causal models for the same client. Also, the authors found low to moderate agreement about the effectiveness of different interventions. Nevertheless, the authors could predict clinicians’ ratings of effectiveness from their individual causal models.

Keywords: causal reasoning, clinical decision making, cognitive mapping, treatment planning

How do clinicians decide on interventions for their clients? Hundreds of different interventions exist that clinicians might choose to use, and this number keeps growing (Lambert, Bergin, & Garfield, 2004; Wenar & Kerig, 2006). A specific intervention will be chosen in 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 systemic-oriented therapy to reduce a child’s aggressive behavior, will depend on whether this aggressive behavior is caused by a lack of social skills, whether poor parenting reinforces the aggressive behavior, 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 causal system underlying a client’s problems such that problems diminish, we expect that clinicians base their predictions of the effectiveness of interventions on client-specific models that specify these causal systems. Such models allow predictions of the consequences of different interventions (Beach, 1992; Sloman, 2005; Sloman & Hagmayer, 2006). This is also what is recommended: Clinicians should integrate client information in coherent case formulations. A case formulation should give a comprehensive description of the client and causally explain the behaviors and symptoms, referring to past events and to interacting physical, psychological, and social factors, such that interventions can be based on them (see, e.g., handbooks of Carr, 2006; Eells, 2007; Groth-Marnat, 2003; Wenar & Kerig, 2006; see also Kuyken, Fothergill, Musa, & Chadwick, 2005).

Causal Models in Intervention Selection

Research in other domains provides converging evidence that people use models representing causal structures to decide on effective interventions (Hagmayer & Sloman, 2005, 2009; Sloman & Hagmayer, 2006). Individual causal models of coronary heart disease were found to be strongly correlated with how nonmedical and medical students rate the effectiveness of different preventive interventions (Green & McManus, 1995). Similarly, a strong relation was 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, Ahn, Proctor, and Flanagan (2009) showed that clinicians distinguish between biological and psycho-
logical or environmental causes for disorders. Furthermore, clinicians believe that medication will be more effective for disorders that have a biological main cause, whereas psychotherapy will be more effective for disorders with a psychological or environmental main cause. Similarly, 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. Yopchick and Kim (2009) investigated how the structure of causal models influenced students’ judgments of the effectiveness of interventions for artificial mental disorders. Students judged interventions that affect initiating causes in the models to be more effective than interventions that affect effects. These findings indicate that clinicians and lay people take causal theories of disorders into account when judging the effectiveness of interventions for disorders.

We know of no evidence, however, that clinicians’ causal client models influence their intervention choices for individual clients. It seems likely that clinicians take into account specific causal factors for specific clients, beyond those for their disorders, because clients diagnosed with the same disorder may have different presentations. Also, “a diagnosis does not carry any necessary implications regarding the causes of the individual’s mental disorder or its associated impairments” (American Psychological Association [APA], 2000, p.xxxiii). According to the APA Presidential Task Force on Evidence-Based Practice

  psychologists must attend to the individual person to make the complex choices necessary to conceptualize, prioritize, and treat multiple symptoms. It is important to know the person who has the disorder in addition to knowing the disorder the person has (APA Presidential Task Force on Evidence-Based Practice, 2006, p. 279).

Thus, it is an interesting question whether clinicians base their treatment judgments for individual clients on individual causal client models.

Several researchers found that clinicians make different causal inferences for the same client (Garb, 2005; Kuyken et al., 2005; Persons & Bertagnoli, 1999). But there is little evidence on how these inferences are related to treatment choices. Witteman and Kunst (1997) found no associations between clinicians’ interpretations of client data and proposed interventions. Clinicians who chose the same data and intervention choices did not share different causal assumptions. Several researchers have found that clinicians differ in their causal beliefs, at least on the level of disorders. In the present study, we aimed to see whether clinicians base their expectations for the effectiveness of different interventions for an individual client on causal models that they construct to explain the problems of this individual client.

We asked clinicians to draw causal client models for two case vignettes, using a cognitive mapping technique. Cognitive mapping aims to extract how people think about a problem and arrive at a decision by asking them to draw the causal relations they perceive between variables relevant for their decision (Axelrod, 1976; Hodgkinson, Maule, & Bown, 2004; see also Kim & Ahn, 2002, for a similar method used with clinicians). We also asked clinicians to rate the effectiveness of various interventions for each of these two case vignettes. We varied the order in which clinicians performed tasks (i.e., model drawing first vs. effectiveness rating first). We analyzed whether clinicians chose different models and the agreement about the effectiveness of the different interventions. Next, we analyzed whether we could predict which interventions clinicians believed to be most effective from their individual causal client models. We hypothesized that different clinicians would construct different causal models and that ratings of effectiveness could be predicted from the individual models.

Method

Participants. Forty Dutch therapists participated who had worked for at least 2 years with children of primary school age. Participants were randomly assigned to one of two groups, which differed in the order of tasks (model drawing first vs. rating of effectiveness first). Table 1 shows participants’ characteristics, for each group. Statistical analyses showed that randomization was successful: There were no significant differences between groups with respect to experience, age, or number of children seen in the past 3 months. That most participants are women is representative for psychotherapists working with children in the Netherlands. Participants who did not work in a youth care institution or an educational setting worked in a private practice or in a medical setting. Participants who said to adhere to one specific theory mentioned different theories (e.g., psychodynamic, learning theory).

Materials. Two case descriptions were constructed on the basis of descriptions of depressed children from De Wit (2000). Each description was one page long (810 words). One case description was about a boy named Thijs, 10 years old, presenting with depressive symptoms from which he suffered since his parents divorced a year ago and he lost contact with his father. The other case description was about a boy named Brian, 11

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Mapping first (n = 20)</th>
<th>Ranking first (n = 20)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Years of experience (M, SE)</td>
<td>11.5 (2.3)</td>
<td>16.4 (2.4)</td>
</tr>
<tr>
<td>Age (M, SE)</td>
<td>38.3 (2.8)</td>
<td>43.4 (2.6)</td>
</tr>
<tr>
<td>Gender (n women)</td>
<td>17</td>
<td>16</td>
</tr>
<tr>
<td>Children seen in last 3 months (M)</td>
<td>19</td>
<td>20</td>
</tr>
<tr>
<td>Working in youth care (n)</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>Working in educational setting (n)</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>Adherence to one specific theory (n)</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 1
Participants’ Characteristics for Each Task Order Group
years old, who also presents with depressive complaints that he had developed about a year ago, when he repeated sixth grade and lost contact with his best friend. Complexity in Brian’s case was increased by adding information indicating comorbidity of attention-deficit/hyperactivity disorder (ADHD), oppositional defiant disorder, and a reading problem (this specific combination was based on findings indicating that these disorders often co-occur; see Carroll, Maughan, Goodman, & Meltzer, 2005; Drabick, Gadow, & Sprafkin, 2006; Ostrander & Herman, 2006). Similar lengths of text were provided 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 represented with both cases, the Wechsler Intelligence Scale for Children-III-NL (intelligence test; Wechsler, 2002), Teacher Report Form (behavior questionnaire filled in by teacher; Achenbach, 1991), Child Behavior Checklist (behavior 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.** The drawing instructions were based on Haynes and Williams’ (2003) functional analytic clinical case formulations, which specify the problems and causal factors that are most important for a client, and the multiple causal relations between these. Specifically, we instructed participants to draw how they understood the case in four separate steps, following cognitive mapping techniques (see Hodgkinson, Bown, Maule, Glaister, & Pearman, 1999; Kim & Ahn, 2002). An example of what such a drawing might look like was presented. The example consisted of empty cells and causal relations between them (see Figure 1). With each step, participants received new instructions.

In the first step, participants wrote down on index cards the problems they identified. Additionally, they wrote down on index cards the causal factors that they believed affected (i.e., initiated, increased, maintained, decreased, or prevented) these problems. Participants could use as many cards as they thought appropriate, increased, maintained, decreased, or prevented) these problems. They were instructed that their 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 was rated as 1, the one or two problems that were the next most important problems as 2, and the remaining problems as 3. In the third step, participants arranged the cards with the problems and causal factors on A3-sized paper and indicated causal relations between them by arrows. Participants were allowed to draw bidirectional relations and loops. Participants also indicated, with a plus or minus sign, for each relation whether it was initiating/increasing or preventing/decreasing. Also, they rated the strength of the relations as weak (1), moderate (2), or strong (3). Finally, in the fourth step participants rated for each problem or causal factor how modifiable they thought it was. We explained that the modifiability of a factor could depend on many factors, for example, resistance to change or availability of effective treatments (cf. Haynes & Williams, 2003). Easily modifiable factors were rated as 2, moderately modifiable as 1, and nonmodifiable as 0.

**Interventions rating task.** In the intervention rating task, participants were presented with a list of 10 possible interventions. Interventions were taken from the literature (Carr, 2006; De Wit, 2000) and discussed with the clinicians who had also commented on the case descriptions. The interventions were (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. For each vignette separately, participants chose and ranked the five most effective interventions, that is, the five interventions they expected to result in in 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 how confident they were that intervention they rated as most effective would indeed be the most effective for the client described in the vignette.

**Linking interventions to variables in the causal models.** For each case, participants indicated which factors or problems in their maps were immediately affected by the five interventions they had chosen. They wrote down the interventions in their maps and drew causal relations between the interventions and the variables affected. Again, participants rated these causal relations as weak (1), moderate (2), or strong (3). Figure 2 shows an example of a participant’s causal map with interventions linked to variables in it.

**Questionnaires.** For both cases, participants completed a short questionnaire, on which they indicated how many similar cases they had seen before, how difficult and how complex they thought the present case was in relation to other cases, whether the description contained sufficient information to rate the effectiveness of interventions, and whether their maps were representative of 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 chose interventions. All instructions were typed; however, participants were encouraged to ask for explanations at any time. The order of

---

1 In a pretest in which we had not limited the number of cards, most participants used eight or nine index cards. We thus decided to constrain the number of cards to 10 in order to avoid participants’ construction of very detailed models, which would have caused difficulties in the comparison of their content. None of the participants indicated they had a problem with this constraint.
the tasks was manipulated. In the mapping-first group, participants first constructed a causal map for a case. Finished maps were put back at their intervention ranking while constructing their model, then they proceeded to the causal mapping task for the first case. Participants could not look back to their maps while making their rankings. Next, participants indicated how the five most effective interventions directly 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, participants began with the intervention ranking task for the first case, followed by the second case. Then they proceeded to the causal mapping task, again starting with the first case. Participants could not look back at their intervention ranking while constructing their model, until they were asked to link the five most effective interventions to variables in their maps. After linking the interventions, they filled in the questionnaire. Then, they repeated these same tasks for the second case.

In both groups, the order of cases was counterbalanced. Upon task completion, participants in both groups answered questions about their age, gender, experience, place of employment and position, theoretical adherence, and they were encouraged to provide comments. Participants were thanked and received a gift certificate of 25 € (about $30 U. S.). The task took, on average, one and a half hours. Two research assistants administered the task at participants’ workplaces, with the exception of two participants who came to the office of the first author.

Analyses and Results

First, we describe characteristics of the causal maps. Then, we present analyses of clinicians’ agreement about causal client models and interventions. Next, we analyze how well clinicians’ individual causal client models predict their rankings of the effectiveness of interventions.

Characteristics of the causal maps. 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. Table 2 depicts the number of variables (i.e., problems and causal factors) and causal relations in participants’ causal models.

As Table 2 shows, participants tended to construct more elaborate causal models for Brian than for Thijs. Paired samples t tests proved 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. These findings conform to the fact that the case description of Brian, but not Thijs, indicated comorbidity of disorders. Participants indicated that their maps were representative of how they thought about the cases, that the amount of information was sufficient, and their confidence that the interventions would be most effective for the case at hand, as expected.

Agreement about causal models and effectiveness of interventions. First, we describe how we recoded variables that participants included in their models so we could assess whether clinicians had constructed similar models. Then, we report the analyses we ran to assess the agreement among clinicians. We analyzed how much models of different clinicians had in common by calculating kappa values separately for the recoded variables in the models and for the causal relations drawn. We also assessed agreement about what interventions clinicians expected to be most effective by calculating kappa values for the effectiveness rankings.

In order to compare the content of participants’ individual causal models, variables had to be coded into the same categories. Three coders (the first author and two research assistants) compared the variables used in 10 randomly selected maps and assigned variables with similar meanings to the same content cate-

![Example of a causal map generated by a participant for Thijs. Variables are depicted as rectangles; interventions as ovals. Straight arrows indicate causal relations among variables; dashed arrows indicate causal effects of interventions. All causal relations in this model were positive. Variables are depicted as rectangles; interventions as ovals. Straight arrows indicate causal relations among variables; dashed arrows indicate causal effects of interventions.](image-url)

### Table 2: Overview of the Number of Variables and Causal Relations

<table>
<thead>
<tr>
<th>Causal Maps</th>
<th>Variables</th>
<th>Causal Relations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thijs</td>
<td>25</td>
<td>36</td>
</tr>
<tr>
<td>Brian</td>
<td>29</td>
<td>42</td>
</tr>
</tbody>
</table>

2 Interestingly, the difference between the models is not reflected by participants’ subjective assessments. The mean difficulty rating for Thijs was 3.6 (SE = 0.2) on a scale ranging from 1 (not at all) to 7 (very); for Brian the mean rating was 3.3 (SE = 0.2). Mean complexity rating for Thijs was 3.6 (SE = 0.2); for Brian 3.6 (SE = 0.2). Paired t tests showed no differences on these measures for the cases, t(37) = 1.1, p > .2; t(37) = 0.2, p > .8. Perhaps the fact that participants had seen more cases like Brian than Thijs (Mdn = 40 for Brian; Mdn = 20 for Thijs) explains why clinicians did not perceive Brian’s case to be more complex or difficult, although they created more elaborate client causal models for Brian.

3 Participants indicated that their maps were representative of how they thought about the case. Means on a scale ranging 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. Participants judged the amount of information on a scale ranging from 1 (insufficient) to 7 (too much) (participants may have judged there was too much information if they believed that the case description included information that was irrelevant to judge the effectiveness of interventions). 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 a midpoint 4, and thus seem to indicate that participants thought that the amount of information was sufficient, t(39) = −1.6, p > .1; t(38) = −1.2, p > .2. Participants also indicated how confident they were that their highest rated intervention would indeed be the most effective for the boy in the case description. On a scale ranging from 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; both scores deviate significantly from a midpoint 4, t(38) = 8.5, p < .001; t(38) = 5.9, p < .001.
gory (e.g., “poor self image” and “thinks negatively about himself” were both recoded into “low self-esteem”). 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 content categories, adding one each time they thought the content was not adequately reflected by any of the categories created before. In the end, 15 content categories for Thijs and 23 content categories for Brian were defined. All three coders agreed for 90.8% used for the case of Thijs; for another 8.9%, two of the three coders agreed. For the case of Brian, all three coders agreed for 83.2% of the variables; for another 15.4%, two of the three coders agreed. Table 3 shows for both cases the relative frequencies of the content categories and the mean importance ratings participants had assigned to the original variables indicating problems. Importance 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 problem, and 1 point to variables not representing a problem.4 This was done separately for each model.

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).

Causal relations were retained in coding. Whenever coding resulted in more than one causal relation between two categorical variables, we copied the strongest relation. Agreement among participants about causal relations was calculated by computing kappa using pairs of categories mentioned by at least five participants. Agreement was again poor for both Thijs (κ = .29) and Brian (κ = .26).

Agreement about the effectiveness of the 10 possible interventions for Thijs and Brian is depicted in Table 4, which shows the relative frequencies of participants choosing each of the interventions as one of the five most effective. On the basis of these relative frequencies, again kappa values were calculated, yielding a fair agreement for Thijs (κ = .43), but a really poor agreement for Brian (κ = .12). Independent samples t tests comparing the mean ranks between task order groups yielded no significant differences. Thus, participants did not choose different interventions as most effective depending on whether they first constructed a causal model or did so after they made their rankings.

To sum up, participants showed rather low agreement about factors and causal relations in the models for the two cases. Thus, different clinicians constructed different causal models for the same client, as we had expected. In addition, there was little to moderate agreement about the effectiveness of different interventions.

Prediction of clinicians’ effectiveness rankings from their individual causal models. To see whether we could predict clinicians’ effectiveness ratings from the individual causal models, we calculated the effects of interventions in participants’ causal models. Because people may derive the effects of interventions from causal models in different ways, we used two different strategies. Each strategy specifies what information from a causal map is used and how this information is integrated to predict the effects of possible interventions (see the next paragraph and Appendix B for details). With each strategy, we calculated the rank orders of the interventions that would be most effective given the individual causal client models. Then we compared these rank orders of interventions with the actual rankings made by participants. However, we do not claim to know which of these two possible causal strategies—or any others—clinicians are more likely to use.

The first strategy we call the maximum impact strategy. It is based on the functional analysis approach proposed by Haynes and Williams (2003), and its basic idea is that first the impact of each variable on 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 the impacts of the variables directly affected by the intervention (see Appendix B for details).

The second strategy we call the maximum reduction strategy. It is based on 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. The effect of an intervention is calculated as the direct and indirect reductions of the problems and causal factors resulting from the intervention (see Appendix B for details).

Using both strategies, we calculated the effects of the interventions in participants’ individual causal models—for the five interventions that were considered most effective and that were thus included in the map. The resulting predictions (impact values in the case of the maximum impact strategy and reduction values in the case of the maximum reduction strategy) were transformed into
ranks. We correlated these ranks with the rankings given by each participant, using Spearman’s rho.5

We averaged these rank-order correlations 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’s z values before averaging. In consequence, the reported means, shown in Table 5, underestimate the actual mean correlations.

In order to test whether these mean correlations were significantly different from chance, we could not use standard test procedures because mean correlations were not based on Fisher’s z-transformed values. Instead, we created a theoretically derived test distribution (cf. Efron & Tibshirani, 1993). First, we created all possible combinations of two rank orders with Ranks 1–5, allowing for double ranks. Next, we correlated each combination of rank orders using Spearman’s rho. Then, we randomly drew 3,000 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 5 are significantly above chance, which means that with both strategies we are able to predict participants’ effectiveness rankings from their individual causal client models.

To test for possible task-order effects, we compared the mean correlations for each strategy and each case between groups (causal mapping first vs. rating effectiveness first) using independent samples t tests. To correct for the accumulation of Type I error, the significance level was set to α = .05/4 = .013. None of the four tests reached significance: −2.5 < t(36) < 1.0; ps ≥ .02. Thus, there were no differences between the two task orders.

Taken together, these findings again are in line with our expectations. Despite the low agreement among clinicians on the causal factors underlying the cases and the effectiveness of different interventions, individual causal client models allowed predictions of the actual rankings well above chance.

---

5 One participant created for both Thijs and Brian a causal model that implied that all interventions would be equally effective. Therefore, these two causal models had to be excluded from the analyses reported hereafter.
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 the extent to which it changes factors that cause and maintain a client’s problems, we expected clinicians to base their predictions of effectiveness on causal client models representing these causal factors. We elicited clinicians’ causal models for two cases and asked them to rate intervention effectiveness. We hypothesized that different clinicians would construct different causal models for the same client and that from the individual causal models, we could predict clinicians’ effectiveness ratings.

Analyses of the agreement about the contents of the causal models showed that participants focused on different variables and assumed different causal relations. Other studies have also found that different clinicians make different causal inferences for the same cases (Garb, 2005; Kuyken et al., 2005; Persons & Bertagnolli, 1999). Furthermore, agreement about intervention effectiveness was moderate to poor in the present study. We could, however, predict clinicians’ effectiveness rankings from their individual causal models. Taken together, it appears that, despite the differences in the causal client models, participants’ effectiveness judgments were associated with these models, supporting our hypotheses.

Similar findings have been obtained for students’ effectiveness ratings of preventive coronary heart disease measures (Green & McManus, 1995), or to reduce environmental risks (Böhm & Pfister, 2000). Our findings are also consistent with previous studies showing that clinicians’ causal assumptions affect their diagnoses (Kim & Ahn, 2002) and inferences about unobserved symptoms (Proctor & Ahn, 2007). Furthermore, based on these data, we may now explain why Witteman and Kunst (1997) and Witteman and Koele (1999) found no association between clinicians’ interpretations of client data and their treatment choices. We observed that when different clinicians consider the same factors to explain a client’s problems, they often disagree on the causal relations between these factors. As a consequence, an intervention may have different effects, depending on how these factors are thought to be causally related. If clinicians take causal structure into account, as our results indicate, they may, while considering the same aspects, still arrive at different treatment choices.

Strategies for predicting intervention effectiveness. We could predict most participants’ effectiveness ratings quite well with both of the causal strategies we defined, and for some participants even perfectly so. Hence, our findings suggest that clinicians consider their own causal models when judging the effectiveness of different interventions. Still, we do not know which particular strategy each clinician used. Theoretically, one may derive effectiveness of interventions from causal models in many different ways, and different clinicians may use different strategies. We do not have any claims about which of these two, or any other possible causal strategy, is more likely to be used. Despite the different assumptions about the information taken into account, it turned out that in most cases both strategies led to similar predictions. In order to differentiate between causal strategies, a more controlled experimental research paradigm would be necessary.

Looking at the complexity of the causal models, it seems that clinicians may not derive estimates of effectiveness from their models by taking into account all variables directly and indirectly affected by the interventions, 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, clinicians may refrain from inferring distant but weak effects of an intervention, because the effect of an intervention rapidly fades, the farther away a problem is from the point of intervention. Predictions from this simpler causal strategy will still closely resemble those from strategies that do include distant effects like the two considered here.

An alternative strategy is that clinicians base their expectations of intervention effectiveness on a general causal model of behavior (i.e., a psychological theory of mind and behavior) rather than on specific causal models constructed for individual clients. Robins, Mendelsohn, Connell, and Kwan (2004) found that people disagree about what causes behavior. However, individuals tend to arrive at similar explanations for other people’s and their own behavior. Thus, Robins et al. suggest that people have a personal theory about what generally causes behavior in themselves and others and that different people have different theories. Similarly, clinicians in our study may have personal theories about what generally causes (problematic) behaviors. They may use these theories to infer the effectiveness of interventions (cf. Hong, Chiu, Dweck, Lin, & Wan, 1999), or, they may have disorder-specific causal models (cf. Kim & Ahn, 2002). Inspection of the contents of the causal models shows that participants in our study used different variables for the two cases (see Table 3). Also, for categories that are present in models for both Thijs and Brian, participants

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Thijs M</th>
<th>Thijs SE</th>
<th>Brian M</th>
<th>Brian SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum impact</td>
<td>0.59</td>
<td>.11</td>
<td>0.45</td>
<td>.08</td>
</tr>
<tr>
<td>Maximum symptom reduction</td>
<td>0.62</td>
<td>.10</td>
<td>0.33</td>
<td>.10</td>
</tr>
</tbody>
</table>
often showed different beliefs about causal relations. For example, for both boys depressive complaints were identified as a category, but more participants believed that depressive complaints lead to fewer social contacts, low self-esteem, and low school performance for Thijs than for Brian. Thus, from the comparison of the models for the two cases, it seems that participants constructed client-specific models rather than general models.

For about one fourth of the participants’ correlations were low between the rankings they gave and the rankings we had derived from their models, with either strategy. At this point, we can only speculate about the strategies these participants might have used. They may, for example, have ranked interventions on the basis of their practical experiences with the effectiveness of different interventions without considering the causal underpinnings in each case. Clinicians participating in the study were quite experienced and had seen many similar cases in their work. Because participants’ agreement about the effectiveness of interventions was quite low, especially for Brian, these experiences must have been different. A second possibility is that these clinicians referred to different prototypical or stereotypical clients (Garb, 1996). By comparing the two boys with different prototypical or stereotypical clients, they may have reached different estimates of effectiveness. Finally, clinicians may have used knowledge about effective interventions obtained from different sources. To conclude, although clinicians probably use a variety of causal and noncausal strategies to derive judgments of effectiveness, our results indicate that many clinicians derive their judgments from their individual causal models.

Limitations and further research. The causal mapping task may have affected some properties of the models generated for each client. Specifically, clinicians’ drawings may be more complex than their actual mental models, given limitations of working memory. Drawing may lead to the construction of more complex models, as visualization facilitates consideration of more variables and relations at the same time (cf. Morton, 2004). However, a comparison of different methods to elicit people’s causal beliefs indicates that causal models that are elicited with a cognitive mapping method are less complex than people’s actual mental models (Hodgkinson et al., 2004). Whether the causal models that clinicians constructed following our specific instructions are more or less complex than clinicians’ actual mental client models is a matter for further research.

An additional possible limitation of this study is that the participant-rated interventions proposed types of treatment rather than implying specified mechanisms of change. Advancing on the findings of Ahn et al. (2009) that beliefs about the modalities of main causes for disorders (biological vs. psychological) affect treatment effectiveness ratings at a high level (medication vs. psychotherapy), we showed that causal beliefs affect clinicians’ ratings of the effectiveness of interventions at a more fine-grained level. Still, interventions used in this study were not at the level of unique, specific techniques. We chose to present broader intervention types to reduce clinicians’ tendency not to consider interventions with which they are unfamiliar. This broader level of interventions seems to match the level of detail we found in clinicians’ causal client models. The fact that clinicians’ causal maps do not depict detailed mechanisms might be a consequence of the amount and type of information offered in the case descriptions (although participants did not rate it as insufficient to make effectiveness judgments for interventions). In addition, it should be noted that specific working mechanisms of interventions are often unknown (Bieling & Kuyken, 2003; De Los Reyes & Kazdin, 2006; Haynes & Williams, 2003). It may be further investigated whether clinicians construct more detailed models if they have sufficiently detailed information to do so, and whether they then aim to target more specific mechanisms of change with specific interventions.

Finally, we only asked participants to predict the effectiveness of different interventions. As we mentioned earlier in the present article, intervention choices in actual practice will not be determined by expected effectiveness only. Factors such as available time and budget, interventions available in an institution, and client-specific (e.g., motivation) or clinician-specific (e.g., experience) factors might also influence intervention decisions (Beutler et al., 2004; Clarkin & Levy, 2004; Nelson & Steele, 2008). For a complete understanding of intervention decisions, these other factors have to be investigated as well.

Implications for clinical practice. We focused on the question of whether clinicians base their intervention effectiveness judgments on causal client models. It is another question whether they should do so in practice. In the introduction, we referred to handbooks suggesting that clinicians derive the most effective interventions from case formulations that 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 inferences about the causes underlying clients’ problems. In addition, it is often unknown precisely how interventions exert their influence (De Los Reyes & Kazdin, 2006; Haynes & Williams, 2003). Clinicians might do better to choose empirically validated interventions based on classifications of clients’ problems. However, empirically validated interventions are not always available, and classifications are not always straightforward, as in cases with multiple disorders or severe problems (Nelson-Gray, 2003). Until now, it is unclear whether the identification of causal mechanisms leads to better treatment outcomes, although some studies indicate that it does with complex cases and severe problems (Beutler et al., 2004; Bieling & Kuyken, 2003; Eells, 2007; Nelson-Gray, 2003).

Our study indicates that different causal models lead to different ratings of intervention effectiveness, and as a consequence may lead to different intervention choices by different clinicians for the same client. Relying on individual causal client models might thus result in bias, of which clinicians should be aware (cf. APA Presidential Task Force on Evidence-Based Practice, 2006). We therefore believe that clinicians should be cautious when making inferences about underlying causal mechanisms (see also Garb, 2005) and that they should apply stringent tests to assumed mechanisms (cf. Haynes, Spain, & Oliveira, 1993; Staines, 2007).

References

Case description Thijs (translated from Dutch). Thijs is eleven years old. His mother is an analyst, his father a sales representative. The parents are divorced. Thijs has a six-year-old sister. Thijs comes in with his mother, who was advised to seek help by Thijs’ teacher, who was worried about him. The teacher had asked the school’s pedagogical assistant, after having consulted mother, to have Thijs fill in the KDVK (a short depression questionnaire for children). Thijs scored within the clinical range.

He works hard at school, but his achievements have deteriorated since last year. The teacher thinks Thijs is a good and docile pupil, but he lacks all initiative. He never raises his hand, and he never asks anything. With creative subjects, it takes long before he gets started. When asked to draw, for example, he thinks for a long time and keeps rubbing things out, when working with clay, he keeps moulding. He keeps to himself with PE and when playing outside. He seems to prefer watching to joining in. He has one friend, who is in another form. The teacher of the preceding school year says Thijs used to be more actively involved in class and 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 of stomachache. Thijs’ sister, on the contrary, is very difficult to handle and hot-tempered. She demands all attention, refusing to eat or sleep. Mother says it feels strange to be here for Thijs; she could use more help with his sister. Thijs always wants to help her in the kitchen, cooking and doing the dishes, which she often finds inconvenient. He does often help her by looking after his sister, for example, when mother goes shopping. Thijs watches television a lot and often plays with his Gameboy. He does not want to engage in sport or join a club. He used to play football, but he stopped after the divorce, because his father no longer took him to the training and matches.

Father left his family two years ago; he was having an affair. He didn’t comply with the arrangement concerning parental access; he often did not show up when it had been agreed that he would pick up the children. Thijs didn’t like being at his father’s, he always wanted to go back to his mother. This year, Thijs has seen his father only at his grandmother’s birthday. Mother has suffered greatly from the divorce. She felt rejected and abandoned. She still feels very lonely. She finds it hard to take care of the children on her own. She has to do everything alone, while she works four days a week as well.

During the interview, Thijs does not talk much. When 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 dislikes the fact that his sister is so difficult; he is sorry for his mother. 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 the fun activities. With his father they used to go to the zoo, or swimming. He says he likes school, but often does not know what he is expected to do.

Pregnancy and birth were normal, as was Thijs’ development. He has always been a sweet and easy child. When he was younger, he frequently suffered from ear infections. After surgery at the age of three, the infections disappeared. He has always been caring with his sister. At school, he achieved average results, until this year.

Scores on the TRF (behaviour questionnaire filled in by the teacher) fall in the clinical range on the scales “withdrawn” and “anxious/depressed behaviour.” Scores on the CBCL (Child Behaviour Checklist filled in by mother) fall in the subclinical range on the scale “somatic complaints.” From the WISC-III (intelligence test), Thijs appears to have an IQ of 104, with a harmonic profile. The CBSK (a scale for children’s perceived competences) shows low scores in the domains “school performance,” “social acceptance,” “sports achievements,” and “self-esteem.” From the DVK (a depression questionnaire for children), it appears that Thijs evaluates himself and his environment negatively.

Mother feels guilty that she had not noticed that Thijs felt so bad. She hopes that he will feel better soon and is prepared to do anything she can. Father was informed about the consultation, but he did not want to be involved.
Case description Brian (translated from Dutch). Brian is ten years old. His mother is a housewife, his father a self-employed handyman. Brian is their only child. Mother comes in with Brian because she can no longer control his behaviour, and on the advice of Brian’s teacher. Mother says it seems as if there is nothing but fights at home. She cannot ask or tell Brian anything. Everything makes him angry. Then he starts to shout, goes outside or to his room stamping his feet and slamming the doors.

Brian has always been a lively 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 it. Father works a lot, most Saturdays too. Father recognises a lot in his son; he used to be very active himself. 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 is not 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 how she handles Brian.

According to the teacher, Brian has always been noticeably present 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 for other children. He is repeating the sixth grade, because he is well behind with language/reading. From the AVI and 1-minute test (reading competence measures), it appears he is two years behind in reading. The teacher thinks Brian has changed a lot this last year. He has become very uninterested; he doesn’t care. Frequently he seems to be lying.

Brian says he always does everything wrong. Things always go wrong and everybody always gets angry. He understands that, because he keeps doing stupid things. Even when he tries his best, it just doesn’t work out. He just doesn’t like school anymore. His best friend has moved away 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 old form anymore.

Pregnancy and birth were normal. Brian used to sleep little and to cry a lot when he was a baby. He developed rapidly; he could walk when he was ten months. He has visited the emergency department more than once, because he had been stunting on his bike, had 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 the treatment 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.

On both the CBCL (Child Behaviour Checklist filled in by the parents) and the TRF (behaviour questionnaire filled in by the teacher), scores fall in the clinical ranges on the scales “aggressive behaviour” and “attention problems,” and in the subclinical range on the scale “anxious/depressive behaviour.” On the AVL (an ADHD questionnaire), Brian has high scores in the domains of hyperactivity, impulsivity, and attention problems. The WISC-III (intelligence test) shows that Brian has an average intelligence (IQ = 97); he has low scores on the parts “substitution” and “digit series.” On the CBSK (a scale for children’s perceived competences), Brian has low scores on “school performance,” “social acceptance,” “behaviour,” and “self-esteem.” From the DVK (depression questionnaire for children), it appears that Brian evaluates himself, his environment, and the future negatively.

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. When Brian is asked what he wants, he says he doesn’t know.

Appendix B

Calculation of the Effect of Interventions in Participants’ Causal Models

General coding. Causal strength ratings were recoded 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 recoded 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 envision the impact each variable has on the client’s situation. Interventions are chosen to 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 on 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 \( c_{ij} = c_{iki} \times c_{klk2} \times \ldots \times c_{kmj} \). 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 \( V_i \) 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, \( c_j = 1 - \prod_{i=1}^{n} (1 - c_{ij}) \) (Jensen, 1996). This rule ensures that causal influences do not sum up to values larger than 1, even if several variables strongly affect one other 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 \( \text{imp}_i \) with its modifiability \( \text{mod}_j \), and the strength of the causal relation relating variable \( V_i \) to all its effects \( c_{ij} \), \( \text{impact}_i = \text{imp}_i \times \text{mod}_j + \sum_{j=1}^{m} \text{imp}_j \times \text{mod}_j \times c_{ij} \). Note that the importance and modifiability of variable \( V_k \) 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 over the causal impacts of the variables directly targeted by the intervention multiplied by the treatment’s causal influence on these variables, \( \text{impacttreatment}_t = \sum_{i=1}^{n} \text{impact}_i \times c_{ij} \). On the basis of the variable \( \text{impacttreatment} \) treatments, \( T_1-T_5 \) 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: (a) Each variable in the model was present and therefore received a starting value of \( v_i = 1 \) and (b) strong interventions reduce the presence of a variable being directly affected by 80%, medium interventions by 50%, and weak interventions by 20%. The second assumption was implemented by calculating the value of a variable \( V_i \) conditional on an intervention \( T \) having the causal strength \( c_{ij} \) as \( v_{ij} = v_i \times (1 - c_{ij}) = 1 - c_{ij} \). This formula also implicitly assumes that an intervention screens off the causal influences of other variables on 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 immediate effect variables was calculated using the above formula. Second, the influence of interventions on the variables that are direct effects of the those directly affected were calculated, followed by the influence of the intervention on the effects of these effects, 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. The causal influence of an intervention \( T \) on variable \( V_j \) via variable \( V_i \) was computed as \( v_{ij} = v_i \times (1 - c_{ij}) \times c_{ij} \). As \( V_i \) initially had a value of 1, the reduction is calculated by \( \text{reduction}_i = 1 - v_{ij} \). If variable \( j \) had several causes \( V_i \), the reduced value of \( V_j \) was calculated by \( \text{reduction}_j = 1 - \left[ \sum_{i=1}^{n} v_{ij} \times c_{ij} \right] / \sum_{i=1}^{m} v_{ij} \). This formula computes the decline 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 supports the notion 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, \( \sum_{i=1}^{n} \text{reduction}_i \). On the basis of this sum treatment, \( T_1-T_5 \) were rank ordered. The intervention with the highest reduction received Rank 1.

Received July 2, 2009
Revision received January 26, 2010
Accepted March 9, 2010