In: Kokinov, B., Karmiloff-Smith, A., Nersessian, N. J. (eds.) ***European Perspectives on Cognitive Science***.
© New Bulgarian University Press, 2011 ISBN 978-954-535-660-5

Does the Structure of Causal Models Predict Information Search?

Ana Sofia Morais (morais@mpib-berlin.mpg.de)
Henrik Olsson (olsson@mpib-berlin.mpg.de)
Lael J. Schooler (schooler@mpib-berlin.mpg.de)
Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany

Abstract

This paper investigates whether the structure of people's knowledge of causal relations between the features of categories predicts how they search for information in a categorization task. Participants were asked to draw a causal model that described how the symptoms of depression are causally related to one another, and to estimate the strengths of those relationships. Additionally, they were asked to categorize a series of patients as suffering from depression or not, after searching their symptoms. The results showed that the structurally more important a symptom was in a causal model, the more frequently and the earlier in search it was inspected. Also, a measure of feature importance that ignored causal strengths accounted for search behavior at least as well as the weighted version of the same measure.

Keywords: Causal models; information search.

Introduction

Most instances of real-world categorization involve information search. Police officers categorize consumer goods as authentic or counterfeit, after examining features such as the spelling of brand names or the purchase prices. Doctors categorize patients as being or not at high risk of having a disease, after inspecting features such as the patients' chief complaints. In situations such as these, people need to determine in what order of importance to inspect features or cues in the environment[1]. Gigerenzer, Hoffrage and Goldstein (2008) reviewed three ways cue importance ranks can be established: evolutionary learning, social learning (involving explicit instruction or imitation, for example), and learning by individual experience. In this paper, we explore another route through which the importance of cues can be determined. Specifically, we investigate whether people rely on their causal models of real-world categories to determine in what order of importance to inspect cues in the environment.

A category's causal model can be defined as a person's knowledge of how the category's features are causally related to one another. Causal models are described as a directed network, where the nodes represent the category's features and the links denote the direction of the causal relationships between them; the node at a link's tail represents a cause, and the node at a link's head denotes an effect. For instance, research looking into the representation of mental disorders has shown that laypeople hold beliefs about causal relations between symptoms (Kim & Ahn, 2002). The standard way of eliciting people's causal models is to present participants with a category's features, such as the symptoms of depression, and ask them to draw directed links between the features, pointing from cause to effect, and to estimate the strengths of those relations (e.g., Kim & Ahn, 2002; Kim & Park, 2009; Sloman, Love, & Ahn, 1998). Figure 1 illustrates the causal model of depression drawn by one participant.

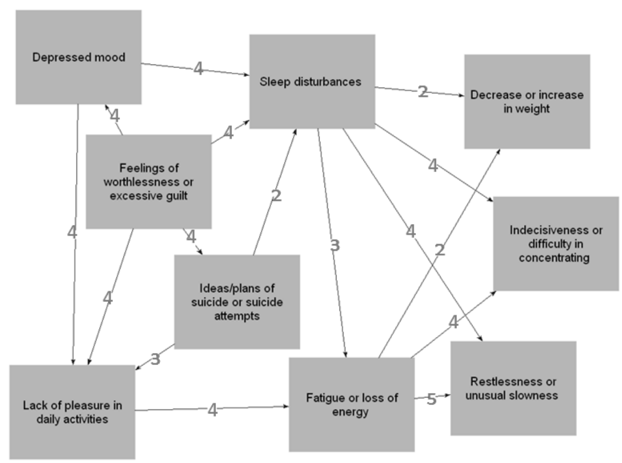

Figure 1: The causal model of depression of one participant.

Past studies have shown that the structure of causal models assessed in this manner affects the importance, or weight, that individual features carry in categorization decisions (e.g., Ahn, Kim, Lassaline, & Dennis, 2000; Rehder, 2003; Rehder & Hastie, 2001). When real-world categories such as depression are tested, researchers first assess the causal models that participants hold for the category and then test how the structure of the models affects the feature weights in categorization decisions (Kim & Ahn, 2002; Sloman, et al., 1998). To evaluate how the elicited causal models affect the importance of individual features in categorization, participants are typically presented with objects (e.g., patients) verbally described as possessing all features (e.g., symptoms) of the target category (e.g., depression) except one, and are asked to rate how likely it is that the object belongs to the category. Differences among the ratings of these objects are interpreted as indicating how the relative importance of missing features varies as a function of their role in the causal models. For example, if a patient without symptom *X* is rated as being less likely to suffer from depression than a patient without symptom *Y*, this is taken to mean that *X* was more important, or carried more weight, than *Y* for

[1] We use the terms "feature" and "cue" interchangeably, where a cue is a feature of object that is used to make an inference.

establishing category membership. However, studies employing the missing feature task tell us little about whether causal models play a role in information search. One key characteristic of the missing feature task is that participants are provided with the cue values, that is, they are told whether each feature is present or absent in a given object. Thus, these studies are silent regarding the role that causal models play in more realistic situations where cue values have to be actively sought in the environment.

In the present study, we assess participants' causal models of a real-life category and examine whether the elicited models influence sequential search in a categorization task where participants were required to search for information. In the experiment, participants were asked to draw a causal model that described how they thought that the symptoms of depression are causally related to one another, and to estimate the strengths of those relationships. In addition, participants performed an information search task in which they were asked to diagnose a series of patients as suffering from depression or not, after searching the same symptoms used in the causal model task. Whether each symptom was present or absent in a given patient, was unknown until the participant selected the cue. Thus, participants were not provided with all the information for categorizing an object, but instead were required to actively search for cues values in the task environment.

What features in a causal model should be examined more often and earlier in search?

Although existing theoretical accounts in causal-based categorization overlook information search, these accounts can be used to generate hypotheses about how the structure of causal models may influence search behavior. Three accounts have been proposed: the *generative theory of categorization* (Rehder, 2003), the *relational centrality account* (Ahn et al., 2000; Rehder & Hastie, 2001), and the *weighted prestige centrality account* (Sloman et al., 1998). These accounts propose alternative rules that assign a feature a measure of its importance as a function of its position in a causal model. Next, we generate alternative hypotheses from these accounts about what cues are important in a causal model, being therefore examined more often and earlier in search. Note, however, that the key question is not whether important cues are examined more often and earlier in search, but rather what makes a cue important in the first place.

Generative Theory of Categorization. According to the generative theory of categorization, causal relations between features are represented as probabilistic causal mechanisms that induce subjective beliefs about how frequently the features are believed to occur in category members (Rehder, 2003). The more frequently a feature is perceived to occur in category members, the more important the feature is to category membership, and so the more often and the earlier in search it should be examined. The quantitative formalization of the theory allows for deriving the feature probabilities directly from causal models as long as the models do not contain causal cycles (Rehder, 2003). Yet when laypeople are asked to report their causal theories of concepts, most participants spontaneously report causal cycles (see Kim, Luhmann, Pierce, & Ryan, 2009). Past work has pointed out possible ways of developing the generative theory in order enable it to handle cyclic structures (Kim et al., 2009), but the suggested changes have not yet been developed. For this reason, the feature weight predictions of the generative theory are not tested in the present work.

Relational Centrality Account. The relational centrality account proposes that features are weighted more, or are more important to category membership, to the extent that they enter into a high number of direct causal relationships, regardless of the strength or direction of those relationships (Ahn et al., 2000; Rehder & Hastie, 2001). In other words, a feature's importance is just the sum of the number of incoming and outgoing links that it has in the category's network of causal relations. According to the relational centrality account, cues involved in many direct relations are given stronger weights in categorization decisions compared with cues involved in fewer relations. Thus, features that enter into a high number of relations should be examined more often and earlier in search compared with features that enter into very few relations.

Weighted Prestige Centrality Account. According to the weighted prestige centrality account, a feature's importance to category membership increases with its number of dependents and the strengths of the causal links with those dependents (Sloman et al., 1998). A feature's number of dependents includes both a feature's direct effects (i.e., the features it causes) and its indirect effects (i.e., features that are, in turn, caused by the features that it causes). Studies have shown that, for many causal structures, the more dependents a feature has, the more important to category membership people consider it to be (Ahn et al., 2000). It has been suggested that people consider cause features to be more important in categorization perhaps because they tend to believe that a cause feature has more predictive power than an effect feature (Ahn et al., 2000).

Sloman and colleagues proposed an iterative equation for determining a features' importance from a category's causal model that corresponds to what are generally known as prestige measures of centrality. In these measures, the importance or centrality of nodes in a network is recursively related to the centralities of the nodes to which they are connected. Nonetheless, since the Sloman et al. (1998) equation cannot be meaningfully applied to many natural and artificial causal structures of theoretical interest, researchers have turned to the PageRank algorithm (Brin & Page, 1998) to derive prestige centrality predictions from causal models (Kim et al., 2009).

In its weighted version, PageRank instantiates the idea that the centrality or importance of a node in a network (i.e., a page on the web, or a feature in a causal model) is a function of the number of links it receives from other nodes, the importance of those nodes, and the strengths of the links.

For weighted PageRank to generate predictions that are comparable to the Sloman et al. equation, it has to be applied to a category's causal network whose links have been inverted, so that the importance ranks assigned by the algorithm depend on the outgoing links (i.e., effects) rather than on the incoming links (i.e., causes). The resulting algorithm – called weighted inverse PageRank – produces prestige centralities that are qualitatively comparable to those produced by the Sloman et al. (1998) equation, with a feature's importance score increasing with the number of direct and indirect effects depending on it, and the strengths of the causal links with those dependents. Therefore, according to the weighted prestige centrality account, the higher the number of features depending strongly on a feature, the more often and the earlier in search this feature should be inspected.

Unweighted Prestige Centrality Account. We also explore how well a naïve version of prestige centrality that does not take causal strengths into account describes sequential search behavior. Causal induction research has shown that models that estimate causal strength do not provide a full account of people's judgments because they are often sensitive to the purely structural question of whether or not a causal relationship exists (Griffiths & Tenenbaum, 2005). Moreover, it need not be the case that people's representations of categories embody causal relationships of various strengths; it is indeed possible that these representations entail relationships with, for instance, equal weights, being therefore coarser than it is often assumed. Finally, several researchers have shown that the choice of weights in a linear model, provided the sign is correct, does often not matter much for minimizing the difference between predicted and true criterion values (Dawes, 1979).

Motivated by the evidence that weights often do no help much, this work examines whether the unweighted version of PageRank, applied to the transposed causal network models, predicts sequential search better than the more complex, weighted version of the algorithm. When measured by unweighted inverse PageRank, a feature's prestige centrality is simply a function of its number of dependents, regardless of the strengths of the links with those dependents. So the higher a feature's number of dependents, the more often and the earlier in search the feature should be inspected.

Summary of Hypotheses. We have derived three alternative hypotheses concerning what cues are important in a causal model, being therefore examined more often and earlier in search. The relational centrality account proposes that a feature's importance to category membership is given by the number of causal relations a feature enters into, regardless of the strength or direction of those relationships. The weighted notion of prestige centrality, in contrast, proposes that a feature's importance is best measured by its number of direct and indirect effects, and the strengths of the links with those effects. Alternatively, according to a more naïve, unweighted notion of prestige centrality, a feature's importance is given solely by the number of direct and indirect effects depending on it.

Method

Participants and Design

Thirty-nine undergraduate students, with an average age of 26, participated in the study. Half of the participants were male. The experimental instructions informed participants that they would receive 5 Euro for their participation, plus 50% of what they would earn in the information search task. On average, participants received a total of 6.71 Euro.

Materials

Two real-life mental disorders were used as categories: anorexia nervosa was used as a training category and depression as a test category. Additionally, the symptoms listed in Table 1 for each of these mental disorders were used as features or cues. These symptoms were compiled based on the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR; Saß, Wittchen, Zaudig, & Houben, 2003).

Procedure

The experiment entailed three tasks: a causal model task, an information search task, and an estimation task. The order of the first two tasks was counterbalanced across participants,

Table 1: Symptoms used for each mental disorder.

Anorexia Nervosa	Depression
Refusal to maintain minimal body weight	Depressed mood
Fear of being fat even when underweighted	Lack of pleasure in daily activities
Disturbed experience of body shape	Sleep disturbances
Absence of period for three menstrual cycles	Fatigue or loss of energy
Binge-eating or purging behavior	Indecisiveness or difficulty in concentrating
Social retraction	Feelings of worthlessness or excessive guilt
Perfectionism	Decrease or increase in weight
Obsessive-compulsive behavior about food	Restlessness or unusual slowness
Strong urges to control one's environment	Ideas/plans of suicide or suicide attempts

with the estimation task being always last.

Causal Model Task. The causal model task assessed the participants' causal models as described earlier in the paper, by means of the software tool ConceptBuilder (Kim & Park, 2009). Participants were asked to draw diagrams of how the symptoms of mental disorders relate to one another. First, participants practiced drawing a diagram of their beliefs about how the symptoms of anorexia nervosa are related to each other, among women aged between 18 and 65 years. After training, participants were asked to draw a diagram of their beliefs about how the symptoms of depression, the test disorder, are related to each other, among women in the same age range.

The symptoms presented in Table 1 appeared all together on the screen for one disorder at a time. The symptoms were presented inside small boxes, organized randomly into two rows. For every symptom X that the participants thought caused a symptom Y, they were asked to draw a directed link between the two symptoms pointing from cause to effect (e.g., $X \rightarrow Y$). After a link was drawn, participants were prompted to give it a weight on a scale of 1 to 5 (where 1 = very weak and 5 = very strong), indicating how likely one symptom was to cause the other. The more likely they thought one symptom was to cause the other, the higher should the weight of the relationship be. In addition, participants were told that, if they thought that a symptom did not cause or was not caused by any other symptom, that symptom should be left in isolation (i.e., without any incoming or outgoing links). Participants were given as much time as needed to complete the task carefully.

Information Search Task. The search task employed an information board. Participants were asked to diagnose a series of 100 patients as suffering from depression or not. Participants were given the same reference class as in the causal model task, that is, all patients to be diagnosed were female, aged between 18 and 65 years. In addition, participants were informed via instructions that the proportion of depression cases in the task reflected the proportion of cases in Germany: 15% of the patients to be diagnosed in the task suffered in fact from depression, while 85% did not (see Robert Koch Institute, 2008).

To make a diagnosis, participants could inspect up to nine symptoms, that is, the nine cues that comprised the profile of the patient. The depression symptoms, which could either be present or absent, were those presented in Table 1. The symptom names were presented inside boxes on the computer screen, with their presentation order randomized across patients. To know whether a symptom was present or not in a given patient, the participant had to click on the box containing that symptom using the mouse. The cue values "present" or "absent" would then appear inside the box, indicating that the symptom was either present or absent in that patient. Once participants moved the mouse out of the box in order to click on the next symptom, the first box was closed by replacing its cue value with the symptom name. This way, only one cue value was visible at a time. At least one symptom had to be looked up to make a decision. For each symptom looked up, 1 cent was subtracted from the participants' payoffs. The balance of their account was not visible on the computer screen until the end of the task. When participants thought they possessed enough information for making a diagnosis, they gave their answer by pressing one of two buttons: one saying "The patient suffers from depression", and the other "The patient does not suffer from depression". For a correct decision, participants earned 7 cents (minus what they spent on looking up cues). After making a diagnosis, participants moved on to next patient with no feedback being provided on the accuracy of the response. In order to ensure that participants felt comfortable with the program, they were asked to first practice diagnosing four patients as suffering from anorexia nervosa or not, after searching the symptoms listed in Table 1 for this disorder. These four practice trials did not contribute to the participants' payoff. The final payoff was displayed on the computer screen at the end of the task.

The order of presentation of the 100 cases was randomized across participants. In the absence of real-world statistics concerning the occurrence of depression symptoms in the population, we constructed the cue profiles of the 100 cases based on the DSM-IV-TR (Saß et al., 2003) guideline for diagnosing depression. According to the manual, for a patient to be diagnosed with depression, at least five of the depression symptoms presented in Table 1 have to be present and at least one of them has to be either "depressed mood" or "lack of pleasure in daily activities". Following this guideline, the no-depression cue profiles contained a maximum number of four symptoms present. In equal proportions, these profiles could either have all symptoms absent, or one, two, three or four of the symptoms present. The symptom(s) that were present in the profile of each non-depressed patient were selected randomly. The depression cue profiles had at least five symptoms present. Three of these profiles had all symptoms present, whereas the remaining 12 had, in equal proportions, five, six, or eight of the symptoms present, with at least one of them being either "depressed mood" or "lack of pleasure in daily activities". Except for the latter two symptoms, all the other present symptoms were selected randomly.

Estimation Task. At the end of the experiment, participants were asked to estimate how many out of 100 women in Germany aged between 18 and 65 years, who have been diagnosed with depression, have each of the depression symptoms presented in Table 1. The results of this task are not analyzed in the present paper.

Results

A first analysis involved calculating sensitivity and response bias measures for the classification data according to the prescriptions of signal detection theory. Since participants were informed about the proportion of depression cases in the task, and the payoff structure of the task involved no cost for a misclassification, it was important to examine whether participants made their decisions on the basis of probability alone. The analysis suggested that this was not likely to be the case. Participants were fairly good at discriminating between the cases in which depression was present and those in which it was absent (mean d'= 2.05, SD = .39), and were only slightly biased to classify patients as not suffering from depression (mean β = 1.55, SD = 3.65).

In the following, we examine whether the structure of people's causal models predicts search behavior, by testing the predictions derived earlier from the relational centrality account and the prestige centrality account (in its weighted and unweighted versions). Different measures were calculated based on the causal model and the search data of each participant. For each feature in the causal model, we calculated its number of relations (i.e., the sum of its number of causes and effects), as well as its weighted and unweighted prestige centrality. The latter two measures were calculated by applying the weighted and unweighted versions of PageRank to the participant's transposed causal model. Moreover, two measures were derived from each participant's search data. One measure was the frequency of cue search, that is, the number of times that each of the nine cues was inspected by the participant across the 100 trials. The second measure was the mean rank order of search, calculated for each participant by taking the average of the rank positions at which each of the nine cues was searched across all trials. Cues that were not examined on a given trial were assigned the median of the missing rank positions. Note that high frequency values signify that a cue was searched often, and high mean ranks imply that the cue tended to be examined late in search.

A nonparametric correlation analysis was performed at the individual level. For each participant, we calculated Spearman's correlation coefficient for the relation between each measure of cue importance and the frequency and order of search. The median correlations across participants were in the expected direction for all feature importance measures: the more important a feature was the more frequently and the earlier in search it was examined. In addition, most of the observed correlations indicated clear, yet moderate effects (Cohen, 1977). First, a feature's number of relations was moderately correlated with search frequency (median r = .31) and with the mean rank order of search (median r = -.29). Weighted prestige centrality exhibited weak correlations with both frequency (median r = .25), and mean rank order of search (median r = -.19). The unweighted version of prestige centrality yielded comparatively stronger, moderate correlations with both frequency (median r = .36) and mean rank order of search (median r = -.35). These results are consistent with those held by a linear regression analysis.

A complementary correlation analysis involved classifying each participant according to the feature importance measure that best described his or her search behavior, as indicated by a comparatively stronger correlation with the two search variables. For 37 out of 39 participants, the measure that held a stronger correlation with search frequency also described the mean rank order better. A cue's number of relations displayed stronger correlations for nine participants' frequency of search and for ten participants' mean rank order. Weighted prestige centrality exhibited comparatively higher correlation coefficients for 13 participants' frequency of cue search and 12 participants' mean rank order of search. The unweighted version of prestige centrality displayed higher correlation coefficients for 18 participants' frequency of cue search and 17 participants' mean rank order of search.

Discussion

The present study provides the first empirical demonstration that the structure of people's causal models predicts how they search for information on which to rely on when making categorization decisions. The structure of people's prior causal knowledge about a category accounted for their search behavior in the absence of any free parameters. Moreover, the study provides evidence that a naïve measure of prestige centrality that ignores causal strengths does not inevitably provide a poorer account of people's search behavior compared to a more complex, weighted prestige centrality measure. Specifically, prestige centrality did not yield higher correlations with the two search measures by taking causal strengths into account, nor did it provide a better description for the search behavior of a larger number of participants. This evidence is reminiscent of the flat maximum effect – the finding that the weights in a linear model can be varied across a broad range without decreasing predictive accuracy (Lovie & Lovie, 1986; von Winterfeldt & Edwards, 1982). In addition, the result suggests that search behavior appears to be sensitive to the purely structural question of whether a causal relationship exists between two features. On the whole, these findings contribute to research on causal-based categorization. Past work has typically overlooked information search and called upon causal strengths when deriving prestige centrality predictions from people's causal models (Sloman et al., 1998; Kim & Ahn, 2002).

Even though clear evidence has been provided showing that causal model structure predicts information search, the theoretical models used to derive predictions could have provided a better account of people's search behavior. It is possible that causal model structure actually had a strong influence on information search behavior in the present experiment, but the different measures used to derive the cue orderings from the causal models were unsuitable for fully uncovering that influence. Thus far, we have mainly considered information search as external – examining cues in the task environment. However, the working hypothesis that causal models influence sequential search in the external environment presumes that information search is also internal. That is, it presumes that a process of memory search and retrieval operates on the structure of a causal model to come to a rank order of cues for external search, with the cues being inspected in the environment in order of their retrieval.

Having established that causal models influence search, the next step is to specify the cognitive processes that may operate on the causal structure of semantic representations. In the spirit of research on semantic networks, one could assume that activation spreads in a causal network model from feature to feature along the links that represent the causal relationships between them. The cue with the highest activation, would be the most likely to be retrieved first, and inspected first in the external environment. Note that, for the memory process to predict external search well, it would probably be sufficient if the first few cues to be retrieved from a causal model would be those that participants considered to be the most important. Supplementary analyses not included in this paper showed that our participants examined few cues and used a rule for stopping search, regardless of how complex their knowledge of causal relations between the cues was.

Another important direction for future research will be to examine whether this process may simplify categorization by retrieving cues in order of their statistical usefulness for making correct inferences. To investigate this question, it will be necessary to disentangle the effects of causal models and environmental statistics in inference tasks. The finding that the structure of people's causal models may simplify categorization is relevant to research concerned with people's use of simple heuristics like fast and frugal decision trees (e.g., Gigerenzer, Todd, and the ABC Research Group, 1999; Martignon, Katsikopolous, & Woike, 2008). One approach for constructing fast and frugal trees has been to order the cues in order of their validity, that is, the probability that each cue alone leads to a correct inference. Nevertheless, studies have shown that participants tend to experience difficulties in computing cue validities and ordering the cues accordingly (e.g., Newell, Rakow, Weston, & Shanks, 2004). Yet the prospective finding that a memory process may retrieve useful cue orders from causal models, would suggest that people do not actually need to compute cue validities before applying a fast and frugal tree to the decision problem. A memory process operating on people's causal models could do the trick.

References

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.

Brin, S., & Page, L. (1998). *The anatomy of a large-scale hypertextual web search engine.* Proceedings of the 7th World-Wide Web Conference, Brisbane, Australia.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences.* New York: Academic Press.

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist, 34,* 571– 582.

Gigerenzer, G., Hoffrage, U., & Goldstein, D. G. (2008). Fast and frugal heuristics are plausible models of cognition: Reply to Dougherty, Franco-Watkins, and Thomas. *Psychological Review*, *115*, 230-239.

Gigerenzer, G., Todd, P. M., & The ABC Research Group. (1999). *Simple heuristics that make us smart.* New York: Oxford University Press.

Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology, 51*, 354–384.

Kim, N. S., & Ahn, W. (2002). The influence of naive causal theories on lay concepts of mental illness. *American Journal of Psychology, 115,* 33–65.

Kim, N. S., Luhmann, C. C., Pierce, M. L., & Ryan, M. M. (2009). Causal cycles in categorization. *Memory & Cognition, 37,* 744–758.

Kim, N. S., & Park, E. Y. (2009). Concept- Builder: An open-source software tool for measuring, depicting, and quantifying causal models. *Behavior Research Methods, 41,* 128–136.

Lovie, A. D., & Lovie, P. (1986). The flat maximum effect and linear scoring models for prediction. *Journal of Forecasting,* 5, 159-168.

Martignon, L, Katsikopoulos, K. & Woike, J. (2008). Categorization with limited resources: a family of simple heuristics. *Journal of Mathematical Psychology, 52,* 352-361.

Newell, B. R., Rakow, T., Weston, N. J., & Shanks, D. R. (2004). Search strategies in decision making: The success of ''success''. *Journal of Behavioral Decision Making, 17*, 117–137.

Rehder, B. (2003). A causal-model theory of conceptual representation and categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 1141–1159.

Rehder, B., & Hastie, R. (2001). Causal knowledge and categories: The effects of causal beliefs on categorization, induction, and similarity. *Journal of Experimental Psychology: General, 130,* 323–360.

Robert Koch Institute (2008). *Health in Germany.* Berlin: Federal Health Reporting.

Saß, H., Wittchen, H.-U., Zaudig, M., & Houben, I. (2003). *Diagnostisches und statistisches Manual psychischer Störungen — Textrevision — DSM-IV-TR*. Göttingen: Hogrefe.

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

Winterfeldt, D. von, & Edwards, W. (1982). Costs and payoffs in perceptual research. *Psychological Bulletin, 91,* 609–622.