Opinion TRENDS in Cognitive Sciences Vol.10 No.9

The causal psycho-logic of choice

Steven A. Sloman[1] and York Hagmayer[2]

[1] Cognitive and Linguistic Sciences, Brown University, Box 1978, Providence, Rhode Island 02912, USA
[2] Department of Psychology, University of Göttingen, Gosslerstr. 14, 37073 Göttingen, Germany

Choices do not merely identify one option among a set of possibilities; choosing is an intervention, an action that changes the world. As a result, good decision making generally requires a model specifying how actions are causally related to outcomes. Interventions license different inferences than observations because an event whose state has been determined by intervention is not diagnostic of the normal causes of that event. We integrate these ideas into a causal framework for decision making based on causal Bayes nets theory, and suggest that deliberate decision making is based on simplified causal models and imaginary interventions. The framework is consistent with what we know so far about how people make decisions.

Introduction

The circumference of our waistlines and the amount of grey in our hair are strongly correlated. Unfortunately however, coloring our hair will not reduce our girth. The correlation arises from a common cause – age – so manipulating our hair color does not solve the real problem. Good decision making generally depends on how actions are causally related to outcomes. Surprisingly, the canonical normative model of decision making, expected utility theory, is not directly concerned with causal relatedness. Make choices, it says, to maximize the probability of achieving the most valued outcomes, but the theory is silent about how to determine those probabilities. The counsel to maximize expected utility would, on the face of it, suggest that all you need to know is the likelihood of events and your own preferences to make rational decisions. But outcome probabilities cannot always be determined by the distribution of previously observed outcomes. The probabilities that are relevant to decision making must reflect the likelihood of outcomes given that the relevant options are actively chosen and not merely observed. Determining these probabilities in general requires a causal model; a probability distribution is not enough.

We argue here that people are sensitive to causal structure: people employ causal models to make decisions and treat choices appropriately, usually as interventions. We present a conceptual framework based on causal Bayes nets to describe how people use causal models to make decisions; we review some evidence to support our position. The importance of causality in decision making was first noticed by philosophers who tried to specify the logic of causal inference, a logic that we believe is captured most fully and elegantly by causal Bayes nets. Although our focus is cognitive, we start with a very brief introduction to the logic of causal intervention.

Causal decision making

Philosophers have long been aware of the importance of causal considerations in decision making [1,2], and several specifications of expected utility theory have developed in response [3–5]. For example, Nozick introduced the distinction between *evidential expected utilities* computed on the basis of merely statistical relations and *causal expected utilities* computed using probabilities derived from causal relations [4]. However, until very recently no formal criteria for differentiating causal and evidential probabilities were available. Causal Bayes net theory [6–9] has closed this gap. Causal Bayes nets constitute a formal framework for representing and reasoning about causal systems using graphical representations called causal models. Causal models represent the causal mechanisms that generate a probability distribution, which in turn reflects the relative frequency of events that we observe (Figure 1a).

A causal relation is more than a correlation (i.e. an evidential relation) in disguise. For example, having fur is correlated with being able to run fast, but this doesn't mean that putting on a fur coat will make you run faster. Whereas causal relations allow for effective interventions, evidential relations only enable inferences. Thus, causal models are causal in two senses: (i) links do not correspond to correlations but to the mechanisms that produce those correlations; and (ii) causal models support a distinction between observation and intervention, a crucial distinction for decision making. In Pearl's system [7], observations, like the probability that a body could run fast if I observed that it had fur, are represented using a conventional conditional probability, P(can run fast | it has fur). Having fur has some diagnostic value for running fast, so this probability is higher than P(can run fast | it does not have fur). But representing an intervention requires a new operator, the *do* operator, because interventions temporarily change the structure of the relevant causal model. An intervention renders the variable intervened on independent of its normal causes because the variable's normal causes are no longer setting it – the intervention is. Therefore, causal links into the variable being intervened on are severed within the causal model representing the causal system, a representational device we call *undoing* [10]; others refer to it as *edge-breaking*. So an intervention, like the probability a body could run fast if I gave it fur, would be calculated from the causal model in which undoing has been implemented (Figure 1b) and written down as P(can run fast | *do*(it has fur)). This interventional probability

Corresponding author: Sloman, S.A. (Steven_Sloman@brown.edu).
Available online 8 August 2006.

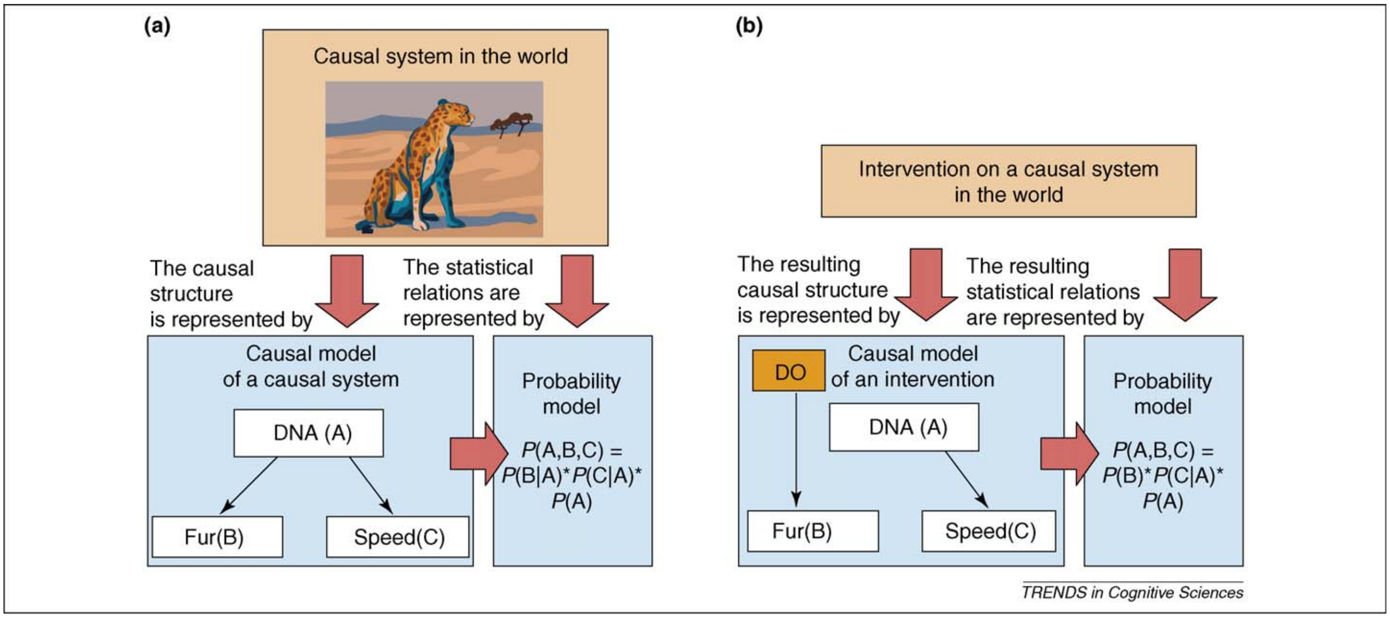

Figure 1. **(a)** Representation of a causal system in the world by a causal model. The small black arrows within the causal model mean 'a causal mechanism exists in this direction'. The graph represents the structure of the causal system. The factorized probabilistic model captures the observable statistical relations by representing the impact of the causal relations and the base rates of root causes. **(b)** Causal model of an intervention setting the value of variable B (e.g. putting on a fur coat). Note that the intervention on B renders B independent of its usual cause A. B is no longer diagnostic of A.

would have the same value as P(can run fast | *do*(it does not have fur)) because the effect of undoing is to render having fur independent of the other variables in the model.

We now have a glimpse of the formal tools Bayes net theory provides to distinguish evidential from causal expected utilities in terms of the probabilities used to determine expectations. Evidential expected utilities are calculated on the basis of observational probabilities; causal ones use interventional probabilities.

Whereas most theorists agree that decisions should be based on causal expected utilities in most cases, debate about rational choice continues in cases like Newcomb's paradox, where causal and evidential expected utilities conflict [1,4]. As the focus of the remainder of this paper is descriptive, not normative, we will not deal with this topic here. Our concern is how people actually *do* make decisions, not how they *should*.

Causal reasoning in decision making

People care about causal relations. They frequently do not settle for observational data but instead seek explanatory mechanisms [11,12]; even young children do this [13]. Adults are more likely to search for explanations when evidence contradicts their beliefs or in the face of bad news or when ascribing blame or responsibility [14]. One benefit of such curiosity is that mechanistic understanding reveals the prerequisites for and boundary conditions on outcomes [15].

People also understand the causal logic of intervention. Sloman and Lagnado [10] and Waldmann and Hagmayer [16] provided direct evidence that people follow the logic of intervention when reasoning and making judgments [17]. Several recent studies have demonstrated that intervention facilitates learning of causal structure relative to mere observation in adults [18–20] and children [21]. Even rats are sensitive to the logic of intervention [22].

Finally, people use causal reasoning in planning and decision making. Experts use simplified models of their domain of expertise to evaluate courses of actions by simulating what would happen if particular strategies were pursued [23,24]. Hagmayer and Sloman's studies [25] of whether and how causal assumptions affect choice indicate that people consider not only evidential relations among actions and outcomes; they treat their choices as interventions to infer their causal effects before making a decision (Box 1). People are willing to take actions that affect the probability of a desired outcome only slightly, but refuse to take actions that have no causal impact even if strongly correlated with the outcome. If no information about the underlying causal model was provided, our participants tended to derive a model from their background knowledge. Moreover, when confronted with several causally effective options, people preferred the alternative with the higher causal expected utility. Our studies of Newcomb's paradox have found that once the causal structure underlying the decision is clarified, choices are more likely to be consistent with causal relative to evidential expected utility.

In summary, studies of learning, attribution, explanation, reasoning, judgment and decision making all suggest that people are highly sensitive to causal structure. A growing number of these studies show a grasp of the logic of intervention in the form of the undoing effect.

Choices – interventions or observations?

From the perspective of a deliberating decision maker, a choice is made by the decision maker himself or herself; any factor that influences choice must have its effect through the decision maker. The decision might be informed and even largely determined by facts about the world, but those facts exert their force only by virtue of the decision maker's deliberation. Decision makers must

Box 1. Evidence that causal structure influences choice

Hagmayer and Sloman [25] presented participants with the following scenario:

Recent research has shown that of 100 men who help with the chores, 82 are in good health whereas only 32 of 100 men who do not help with the chores are. Imagine a friend of yours is married and is concerned about his health. He reads about the research and asks for your advice on whether he should start to do chores or not to improve his health. What is your recommendation?

Several possible causal structures could underlie the correlation between chores and health. We presented different groups of participants two of them. Before being asked to give their final answer, one group was informed in a Direct Cause condition that research has discovered that the cause of this finding is that doing the chores is an additional exercise every day.

A second group was informed in a Common Cause condition that research has discovered that the cause of this finding is that men who are concerned about equality issues are also concerned about health issues and, therefore, both help to do the chores and eat healthier food.

If decisions are made on the basis of a mere statistical relation between the action and the outcome, the same recommendations should be made regardless of the causal structure provided. By contrast, if participants took the causal model into account, they should recommend doing the chores only given a direct cause model; only in that would doing the chores trigger the mechanism increasing health.

In a common-cause model, the choice to do the chores would eliminate the causal relation between the degree of concern and chores and therefore render doing the chores independent of health. Thus, the action would have no impact on the outcome.

Participants proved to be highly sensitive to the underlying causal structure. Across several scenarios, only 23% of the participants given a common-cause model recommended action compared with 69% given a direct causal link.

believe that they determine their own choices otherwise there would be no reason to deliberate. As it is up to the decision maker whether and how to take those facts into account to make a decision, the decision itself provides no new information about the facts. Moreover, as far as the decision maker is concerned, the given facts are fixed and therefore choosing cannot reveal anything to the decision maker about the causes of the current choice that the decision maker did not already know. In other words, choice itself cannot be diagnostic of the external factors that influence the decision. Hence, we propose, the decision maker's choice is an intervention because it renders the action chosen independent of all its normal causes in the decision maker's model of the situation in that it provides no new information to the decision maker about the choice's causes. Moreover, we claim that choice is a strong intervention, one that determines the value of the variable (i.e. the action) intervened on [9]. Motivation for this claim comes from research showing that young children assume that actions are caused by desires and intentions and not external factors. Moreover, free agency is considered a central property of animate objects [26].

Note that the action deliberately determined by choice is only a strong intervention for variables that the decision maker can fully control. For example, choosing to go to the movies determines the action of going to the movies but does not by itself determine the enjoyment of the movie (the movie itself partially determines that). The choice is only a weak intervention with respect to the movie outcome.

However, choice might not be an intervention from the perspective of an observer. From an observer's perspective, choices depend on other variables, like environmental affordances and personality traits of the decision maker. For example, while the decision maker's perspective is that she freely chooses to see another Julia Roberts movie, the observer's perspective might be that it is Julia Roberts who causes her to see the movie. Figure 2 sketches the causal models underlying the two perspectives. Whereas choice is an intervention from the perspective of the decision maker's model, i.e. the model used to actually make the choice, choices are merely effects when judged from an observer's perspective (Kant [27] argues for the dual nature of choice). Indeed, in planning and forecasting, observers assume that actors have less control over an outcome than the actors themselves assume [28]. Note that a decision maker could potentially be the observer of his or her own choice, perhaps most easily at some point in time outside the period during which the decision at hand is being made.

A causal model framework for choice

Based on the assumption that choices are interventions, we propose an integration of the novel insights from causal Bayes nets into a descriptive account of decision making. The central idea of this framework is that deliberate planning and decision making is based on simplified causal models of the decision context and imaginary interventions. We suggest that a decision maker goes through three phases to make a decision. The decision-making process is depicted in Figure 3.

Phase 1. Representation of the decision problem: world model construction

In this phase, a causal model of the decision problem is constructed. First, the decision maker's goals are represented as a distribution of preferred causal consequences (e.g. low body weight). Then, factors that are causally relevant to those consequences (e.g. diet, sports and body-mass set point) are identified. In general, only a few factors will be considered depending on a variety of contextual factors resulting in a simplified model [23,29]. Information about these factors might be directly available or inferred from background knowledge. Next, the decision maker constructs a causal model of the decision problem describing how these factors bring about causal

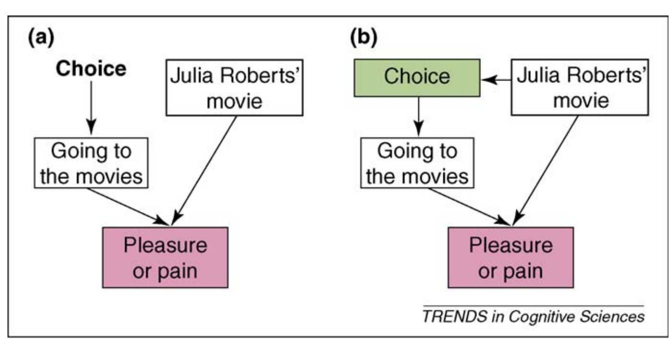

Figure 2. **(a)** Decision maker's perspective: decision maker's model for the sake of choice. Choice is an intervention on a variable within the model. **(b)** Observer's perspective: model of decision maker's choice. Choice is a variable within the model.

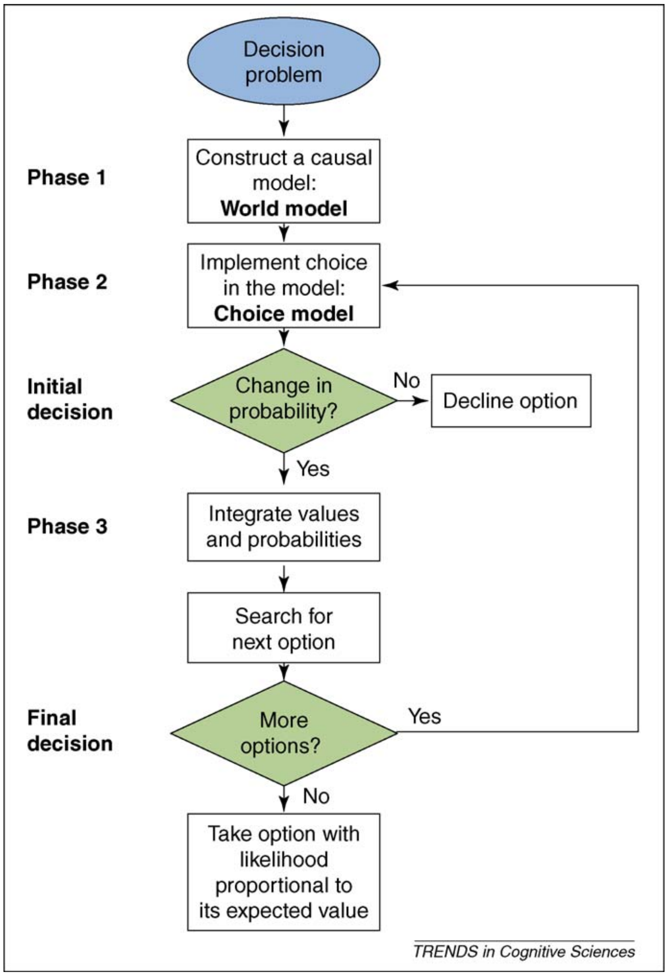

Figure 3. A causal model framework for decision-making, as discussed in this article (see text for details).

consequences. Finally, the available information is used to make inferences about the presence of variables that have not been observed and to infer how likely the desired consequences are to occur under some more or less specified set of conditions. In Bayesian terms: the model is updated conditional on all known values of factors (methods for doing this can be found in, e.g. Halpern [30]).

The construction of the world model can be facilitated if the given information already conforms to a causal model [31]. Goal and attribute framing can affect which causal consequences are considered. Reference points can be represented within a causal model by assigning specific values to the consequences considered. The status quo is captured by assigning actual values to the consequences; specific goals by values to be obtained. We assume that the default is to update the model by assigning current values to the variables, which implies that the status quo would serve as the reference point in most cases [32,33].

Phase 2. Representation of options: choice model construction

To evaluate the consequences of a potential action, a new model must be set up representing the consequences of the intervention resulting from each choice considered deliberately by the decision maker. To do so, the variable being intervened on in the world model has to be disconnected from its normal causes. The choice model differs from the decision maker's world model in four ways:

(i) a choice–intervention node is added to represent the intervention;
(ii) the manipulated variable is given the value assigned by choice;
(iii) undoing is implemented, i.e. the manipulated variable is cut off from its normal causes;
(iv) the probabilities of relevant consequences are updated, i.e. the interventional probabilities are calculated.

An example for a choice model can be seen in Figure 2a. Note that computing the interventional probabilities requires a causal model.

Initial decision

Choices that would result in actions that do not affect desired consequences are discarded right away. In other words, when omission and commission have the same causal consequences, the default choice is not to act because any kind of action incurs some cost in time, effort, the assumption of responsibility or something else.

An option might also be accepted straight away – without comparison to subsequent options – if a choice sufficiently increases the probability of the desired outcome. This would constitute *satisficing* [34]. In extreme cases, people under a great deal of pressure tend to take the first option considered if it is assumed to be causally effective [23,24].

Phase 3. Estimating causally expected values

The *interventional probabilities* resulting from choice and values of desired outcomes are integrated into a causal expected value for a given option. Several theories describe the integration process [35]. For the purposes of this paper, we are agnostic about the details of this process, although we suspect that it involves more qualitative than quantitative reasoning.

Final decision: picking the option yielding the best causal consequences

If more than one option is available to the decision maker, a choice model has to be constructed for each and causal expected values have to be estimated. The final decision is made by choosing options with the most favorable causal consequences: we propose that actions are chosen with a probability proportional to their causal expected utility [36].

Post-decisional processes

After the actual consequences of a decision are obtained, learning can occur. That is, the causal model might be modified to represent the environment and the consequences of choice more veridically. Several algorithms exist to model learning in a causal network [8], some intended to be psychologically plausible [37,38]. The fact that choice is an intervention complicates learning about the causes of a choice, but the data suggest that it eases learning about its consequences [39].

Conclusions

The causal model framework for choice is a theoretical framework designed to account for people's sensitivity to

Box 2. Questions for future research

- Can people reason about arbitrarily complex causal systems or are there principled limitations on what can be computed?
- How is abstract causal knowledge (e.g. principles of human behavior) incorporated into causal reasoning?
- Do people always reason with a single causal model or are we able to integrate uncertainty about causal structure into our reasoning?
- How is causal knowledge integrated with other sorts of knowledge (e.g. spatial)?
- Is causal reasoning more or less accurate if it is conscious and deliberate versus automatic and outside awareness?
- Is the model depicted in Figure 3 generally consistent with how people make decisions?

causal structure and the logic of intervention when making decisions. We all know immediately that (unfortunately) coloring your hair will not reduce your girth and this kind of causal knowledge is used quickly and effectively to make decisions of many kinds.

The construction of causal models turns out to be useful for strategic planning of interventions. In business, strategy maps involving causal relations have emerged as tools for strategic planning [40,41]. Causal models of psychiatric disorders have also been recommended [42] and have proven useful for planning therapeutic interventions [43].

Not all decision making can be described using causal models. People use various different strategies in different contexts [44]. They even resort to non-consequentialist decision making [45]. But our review shows that people frequently calculate the effects of their actions on the causal system they are operating in. This is sensible because, with the right causal model and the right calculations, it can provide the best guess of the outcome of a choice. Nevertheless, people sometimes engage in bad causal reasoning to protect their self-esteem, as in cases of self-handicapping [46] and self-deception [47]. And faulty causal reasoning can hurt statistical judgment [11]. But even bad causal reasoning arises from the human tendency to try to figure out how the world works and to appreciate that choices are interventions that change the state of the world. Future work will help us specify the capacity of human causal reasoning, when and how motivation can lead us astray, and the conditions for acquiring accurate causal models (Box 2).

Acknowledgements
This work was funded by NSF Award 0518147. Clark Glymour, Tania Lombrozo, Benjamin Pitt, Emanuel Robinson and anonymous reviewers provided useful comments on previous drafts.

References

1 Joyce, J.M. (1999) *The Foundations of Causal Decision Theory*. Cambridge University Press
2 Meek, C. and Glymour, C. (1994) Conditioning and intervening. *Br. J. Philos. Sci.* 45, 1001–1021
3 Lewis, D. (1981) Causal decision theory. *Australas. J. Philos.* 59, 5–30
4 Nozick, R. (1995) *The Nature of Rationality*. Princeton University Press
5 Skyrms, B. (1980) *Causal Necessity*. Yale University Pres
6 Glymour, C. (2003) Learning, prediction and causal Bayes nets. *Trends Cogn. Sci.* 7, 43–48
7 Pearl, J. (2000) *Causality*. Cambridge University Press
8 Spirtes, P. *et al.* (1993) *Causation, Prediction, and Search*. Springer
9 Woodward, J. (2003) *Making Things Happen. A Theory of Causal Explanation*. Oxford University Press
10 Sloman, S.A. and Lagnado, D. (2005) Do we 'do'? *Cogn. Sci.* 29, 5–39
11 Dawes, R.M. (1999) A message from psychologists to economists: mere predictability doesn't matter like it should (without a good story appended to it). *J. Econ. Behav. Organ.* 39, 29–40
12 Sloman, S.A. (2005) *Causal Models: How We Think about the World and its Alternatives*. Oxford University Press
13 Wellman, H.M. and Lagattuta, K.H. (2004) Theory of mind for learning and teaching. The nature and role of explanation. *Cogn. Dev.* 19, 479–497
14 Hilton, D.J. *et al.* (2005) The course of events: counterfactuals, causal sequences and explanation. In *The Psychology of Counterfactual Thinking* (Mandel, D. *et al.*, eds), pp. 44–73, Psychology Press
15 Lombrozo, T. and Carey, S. (2006) Functional explanation and the function of explanation. *Cognition* 99, 167–204
16 Waldmann, M.R. and Hagmayer, Y. (2005) Seeing versus doing: two modes of accessing causal knowledge. *J. Exp. Psychol. Learn. Mem. Cogn.* 31, 216–227
17 Hagmayer, Y. *et al.* Causal reasoning through intervention. In *Causal Learning: Psychology, Philosophy, and Computation* (Gopnik, A. and Schulz, L., eds), Oxford University Press. (in press)
18 Lagnado, D.A. and Sloman, S.A. (2004) The advantage of timely intervention. *J. Exp. Psychol. Learn. Mem. Cogn.* 30, 856–876
19 Lagnado, D. and Sloman, S.A. Time as a guide to cause. *J. Exp. Psychol. Learn. Mem. Cogn.* 32, 451–460
20 Steyvers, M. *et al.* (2003) Inferring causal networks from observations and interventions. *Cogn. Sci.* 27, 453–489
21 Kushnir, T. and Gopnik, A. (2005) Young children infer causal strength from probabilities and interventions. *Psychol. Sci.* 16, 678–683
22 Blaisdell, A.P. *et al.* (2006) *Causal Reasoning in Rats*. Science
23 Klein, G. (1998) *Sources of Power*. MIT Press
24 Zsambok, C.E. and Klein, G. (1997) *Naturalistic Decision Making*. Erlbaum
25 Hagmayer, Y. and Sloman, S.A. (2005) Causal models of decision making: choice as intervention. In *Proceedings of the Twenty-Seventh Annual Conference of the Cognitive Science Society*, Stresa, Italy
26 Gopnik, A. and Meltzoff, A.N. (1997) *Words, Thoughts, and Theories*. MIT Press
27 Kant, I. (1785/1974) *Kritik der praktischen Vernunft*. Suhrkamp
28 Buehler, R. *et al.* (1994) Exploring the planning fallacy: why people underestimate their task completion times. *J. Pers. Soc. Psychol.* 67, 366–381
29 Ahn, W-K. *et al.* (1995) The role of covariation versus mechanism information in causal attribution. *Cognition* 54, 299–352
30 Halpern, J. (2003) *Reasoning About Uncertainty*. MIT Press
31 Pennington, N. and Hastie, R. (1993) Reasoning in explanation-based decision making. *Cognition* 49, 123–163
32 Levin, I.P. *et al.* (1998) All frames are not created equal: a typology and critical review of framing effects. *Organ. Behav. Hum. Decis. Process.* 76, 149–188
33 Heath, C. *et al.* (1999) Goals as reference points. *Cogn. Psychol.* 38, 79–109
34 Simon, H. (1957) *Models of Man: Social and Rational*. Wiley
35 Koehler, D.J. and Harvey, N. (2004) *Blackwell Handbook of Judgment and Decision Making*. Blackwell Publishing
36 Hernstein, R.J. (1961) Relative and absolute strength of responses as a function of frequency of reinforcement. *J. Exp. Anal. Behav.* 4, 267–272
37 Gopnik, A. *et al.* (2004) A theory of causal learning in children: causal maps and Bayes nets. *Psychol. Rev.* 111, 3–32
38 Tenenbaum, J.B. and Griffiths, T.L. (2003) Theory-based causal inference. *Adv. Neural Inf. Process. Syst.* 15, 35–42
39 Lagnado, D.A. *et al.* Beyond covariation: cues to causal structure. In *Causal Learning: Psychology, Philosophy, and Computation* (Gopnik, A. and Schulz, L., eds), Oxford University Press. (in press)
40 Hodgkinson, G.P. *et al.* (2004) Causal cognitive mapping in the organizational strategy field: a comparison of alternative elicitation procedures. *Organ. Res. Methods* 7, 3–26
41 Kaplan, R.S. and Morton, D.P. (2004) *Strategy Maps: Converting Intangible Assets into Tangible Outcomes*. Havard Business School Publishing

42 Fiedler, P. (1997). Towards disorder-specific treatments. In *Therapieplaning in der modernen Verhaltenstherapie* [Therapy planning in modern behavior therapy] (Reinecker, H. and Fiedler, P., eds), Pabst
43 Morton, J. (2004) *Understanding Developmental Disorders: A Causal Modeling Approach.* Blackwell
44 Weber, E.U. *et al.* (2004) '*How* do I choose thee? Let me count the ways'. A textual analysis of similarities and differences in modes of decision making in the U. S. A. and China. *Management and Organization Review* 1, 1–32
45 Baron, J. (1994) Nonconsequentialist decisions. *Behav. Brain Sci.* 17, 1–42
46 Covington, M.V. and Roberts, B.W. (1994) Self-worth and college achievement: motivational and personality correlates. In *Student Motivation, Cognition, and Learning: Essays in Honor of Wilbert J. McKeachie* (Pintrich, P.R. *et al.*, eds), pp. 157–187, Erlbaum
47 Quattrone, G. and Tversky, A. (1984) Causal versus diagnostic contingencies: on self-deception and on the voter's illusion. *J. Pers. Soc. Psychol.* 46, 237–248