Springer

Dear Author,

Here are the proofs of your article.

- You can submit your corrections **online**, via **e-mail** or by **fax**.
- For **online** submission please insert your corrections in the online correction form. Always indicate the line number to which the correction refers.
- You can also insert your corrections in the proof PDF and **email** the annotated PDF.
- For fax submission, please ensure that your corrections are clearly legible. Use a fine black pen and write the correction in the margin, not too close to the edge of the page.
- Remember to note the **journal title**, **article number**, and **your name** when sending your response via e-mail or fax.
- **Check** the metadata sheet to make sure that the header information, especially author names and the corresponding affiliations are correctly shown.
- **Check** the questions that may have arisen during copy editing and insert your answers/ corrections.
- **Check** that the text is complete and that all figures, tables and their legends are included. Also check the accuracy of special characters, equations, and electronic supplementary material if applicable. If necessary refer to the *Edited manuscript*.
- The publication of inaccurate data such as dosages and units can have serious consequences. Please take particular care that all such details are correct.
- Please **do not** make changes that involve only matters of style. We have generally introduced forms that follow the journal's style.
  Substantial changes in content, e.g., new results, corrected values, title and authorship are not allowed without the approval of the responsible editor. In such a case, please contact the Editorial Office and return his/her consent together with the proof.
- If we do not receive your corrections **within 48 hours**, we will send you a reminder.
- Your article will be published **Online First** approximately one week after receipt of your corrected proofs. This is the **official first publication** citable with the DOI. **Further changes are, therefore, not possible.**
- The **printed version** will follow in a forthcoming issue.

Please note

After online publication, subscribers (personal/institutional) to this journal will have access to the complete article via the DOI using the URL: http://dx.doi.org/[DOI].
If you would like to know when your article has been published online, take advantage of our free alert service. For registration and further information go to: http://www.springerlink.com.

Due to the electronic nature of the procedure, the manuscript and the original figures will only be returned to you on special request. When you return your corrections, please inform us if you would like to have these documents returned.

Metadata of the article that will be visualized in OnlineFirst

ArticleTitle	From colliding billiard balls to colluding desperate housewives: causal Bayes nets as rational models of everyday causal reasoning	
Article Sub-Title		
Article CopyRight	Springer Science+Business Media B.V. (This will be the copyright line in the final PDF)	
Journal Name	Synthese	

Corresponding Author	Family Name	**Hagmayer**
	Particle	
	Given Name	**York**
	Suffix	
	Division	
	Organization	King’s College London
	Address	Capital House, 42 Weston Street, London, SE1 3QD, UK
	Email	yhagmay@gwdg.de
Author	Family Name	**Osman**
	Particle	
	Given Name	**Magda**
	Suffix	
	Division	School of Biological and Chemical Science
	Organization	Queen Mary College, University of London
	Address	Mile End Road, London, E1 4NS, UK
	Email	m.osman@qmul.ac.uk

	Received	7 December 2011
Schedule	Revised	
	Accepted	17 July 2012

Abstract

Many of our decisions pertain to causal systems. Nevertheless, only recently has it been claimed that people use causal models when making judgments, decisions and predictions, and that causal Bayes nets allow us to formally describe these inferences. Experimental research has been limited to simple, artificial problems, which are unrepresentative of the complex dynamic systems we successfully deal with in everyday life. For instance, in social interactions, we can explain the actions of other’s on the fly and we can generalize from limited observations to predict future actions and their consequences. Our main argument is that none of these inferences (i.e., induction, generalization, explanation, and prediction) can be achieved without causal reasoning. As a case in point we use the popular television series desperate housewives and show how causal Bayes nets are able to explain the inferences made in social contexts. Crucially, causal Bayes nets also allow us to understand why we can infer so much from so little when making sense of a protagonist’s behavior.

Keywords (separated by '-') Causal reasoning - Bayes nets - Rational models - Everyday decision making

Footnote Information

Journal: 11229
Article: 162

Author Query Form

Please ensure you fill out your response to the queries raised below and return this form along with your corrections

Dear Author

During the process of typesetting your article, the following queries have arisen. Please check your typeset proof carefully against the queries listed below and mark the necessary changes either directly on the proof/online grid or in the 'Author's response' area provided below

Section	Details required	Author's response
References	Please check the following reference is in list not in text: Gopnik and Schulz (2007); Osman (2008); Pearl (1988); Sloman and Hagmayer (2006); Waldmann et al. (2006a);. Please cite in text or delete from the list	

Synthese
DOI 10.1007/s11229-012-0162-3

From colliding billiard balls to colluding desperate housewives: causal Bayes nets as rational models of everyday causal reasoning

York Hagmayer · Magda Osman

Received: 7 December 2011 / Accepted: 17 July 2012

Abstract Many of our decisions pertain to causal systems. Nevertheless, only recently has it been claimed that people use causal models when making judgments, decisions and predictions, and that causal Bayes nets allow us to formally describe these inferences. Experimental research has been limited to simple, artificial problems, which are unrepresentative of the complex dynamic systems we successfully deal with in everyday life. For instance, in social interactions, we can explain the actions of other's on the fly and we can generalize from limited observations to predict future actions and their consequences. Our main argument is that none of these inferences (i.e., induction, generalization, explanation, and prediction) can be achieved without causal reasoning. As a case in point we use the popular television series desperate housewives and show how causal Bayes nets are able to explain the inferences made in social contexts. Crucially, causal Bayes nets also allow us to understand why we can infer so much from so little when making sense of a protagonist's behavior.

Keywords Causal reasoning · Bayes nets · Rational models · Everyday decision making

Many of our everyday judgments and decisions are made with respect to causal systems. These systems are often complex (i.e., they encompass many variables and causal mechanisms), dynamic (i.e., they change autonomously and through

Y. Hagmayer (✉)
King's College London, Capital House, 42 Weston Street, London, SE1 3QD, UK
e-mail: yhagmay@gwdg.de

M. Osman
School of Biological and Chemical Science, Queen Mary College, University of London, Mile End Road, London, E1 4NS, UK
e-mail: m.osman@qmul.ac.uk

Journal: 11229-SYNT Article No.: 0162 ☐ TYPESET ☑ DISK ☐ LE ☐ CP Disp.:2012/7/23 Pages: 12 Layout: Small-X

direct intervention) and intransparent (i.e., their underlying structure is not obvious) (Dörner 1989; Osman 2010a,b). Nevertheless, as complex as they are, we are able to explain events, make predictions, and choose effective actions to generate particular outcomes within them. Take for example people living in a suburban neighborhood. The social interactions within the neighborhood constitute a dynamic system which can change gradually over time (e.g., developing close friendships), or suddenly when a single momentous event occurs (e.g., a neighbor gets accused of child harassment). To make sense of the complex and changing interplay between people, verbal and physical behaviors are used to form an impression (i.e., create a model) of others' personalities. Forming a model enables people to understand their own and others' actions, habits and emotional expressions. Crucially, it also enables them to predict behaviors and actions in novel situations by going beyond previous experiences.

The main thesis of this paper is that when it comes to complex dynamic systems, such as the social one just described, everyday inferences are made possible because of abstract theories which allow us to rapidly generate causal models. In turn, these models enable us to make predictions for events never previously experienced. We will show how Bayes nets are a means to formally model knowledge, situation-specific models and inductive inferences. We discuss the ways in which Bayes nets can model people's lay-theories in social domains and can be used to explain the inductive inferences, rapid explanations and predictions people are able to make on the fly. To this end, we use the popular TV series desperate housewives for illustrative purposes. We end by arguing that Bayes nets can also provide a rational model of inductive inferences in complex, dynamic domains, social and otherwise.

1 Causal considerations in dynamic environments

There are at least two domains in psychology in which decision makers' handling of complex dynamic systems in the real world have been investigated. One is the naturalistic decision-making (NDM) research program which examines decision making in vivo (Johnson 2003; Lipshitz et al. 2001), and the second is the dynamic decision making (DDM) program which simulates real world problems in the laboratory (Broadbent and Ashton 1978; Dörner 1989; Osman and Speekenbrink 2011). Both programs are focused on understanding the proficiency of decision-makers in conditions of uncertainty and complexity. That is, the research is designed to understand how people with varying levels of expertise tackle situations in which there are intricate interplays between people (e.g., in management or organization) or between variables within artificial systems (e.g., in industrial or economic systems), (Osman 2010a,b). Findings indicate that through experience people are able to learn to successfully manipulate and manage such systems (see Osman 2010b for a review). Broadly speaking experience enables decision makers to (a) evaluate a course of action by surveying the problem as a whole, (b) generate solutions quickly by pattern matching to previously stored plans of action, and then (c) implement decision rules that stop searches in memory for alternative plans of action (Lipshitz et al. 2001). Some theoretical accounts suggest that people learn to control systems by strengthening the association between perceptual features of problems and successful actions (e.g., Dienes and Fahey 1995).

Journal: 11229-SYNT Article No.: 0162 TYPESET ✓ DISK LE CP Disp.:2012/7/23 Pages: 12 Layout: Small-X

Other accounts assume that previous instances of both failed and successful actions are stored in memory, which become activated by the observed features of a problem via pattern matching processes (Lipshitz et al. 2001). Neither account assumes that inferences are made about the causal structure underlying the system. This means that decision makers don't necessarily need causal knowledge when interacting with a dynamic system (cf. Berry and Broadbent 1995).

This is at odds with recent research on causal decision making in both static environments (Hagmayer and Meder 2012; Hagmayer and Sloman 2009; Sussman and Oppenheimer 2011) and dynamic (Hagmayer et al. 2010), which shows (i) that people take causal knowledge into account when making repeated decisions, and (ii) that they start to induce causal models based on the observed feedback spontaneously. It is also at odds with research on story comprehension in film and television (Wyer and Radvansky 1999), which provides good evidence that people spontaneously generate causal representations to structure their understanding of a developing social context (Brownstein and Read 2007; Radvansky et al. 2005).

2 The importance of causal considerations

There are strong theoretical arguments which explain why causal knowledge and reasoning is important for learning about and interacting within complex systems (cf. Spirtes et al. 2000; Hagmayer and Lagnado 2011) and there is some evidence substantiating these claims.

First, when predicting a target outcome variable from other variables within a system it is important to consider the causal structure connecting these variables to the outcome. This is most obvious when the variables' predictive validities (i.e., the likelihood of the outcome being present if the variable is present) have been learnt separately and need to be integrated to make a prediction based on more than one variable. If the variables are causes of the outcome, the predictive validities will reflect their causal impact on the outcome, hence their predictive validities should be added to derive predictions ($P(\text{Out}|\text{A\&B}) = P(\text{Out}|\text{A}) + P(\text{Out}|\text{B}) - P(\text{Out}|\text{A}) * P(\text{Out}|\text{B})$). However, if the variables are effects of the outcome, predictive validities reflect their diagnosticity. Therefore they should be integrated according to Bayes rule ($P(\text{Out}|\text{A\&B}) = P(\text{A}|\text{Out}) * P(\text{B}|\text{Out}) * P(\text{Out}) / P(\text{A\&B}|\text{Out}) * P(\text{Out}) + P(\text{A\&B}|\sim\text{Out}) * P(\sim\text{Out})$). These different integration rules may result in substantially different predictions (Hagmayer and Lagnado 2011). For example, if both variables have a high predictive validity and one is present while the second is absent, then the outcome is very likely to be present when the variables are causes, but the outcome is only moderately likely if the variables are effects of the outcome. Rehder (2003) provides ample evidence that people are sensitive to causal structure when making predictions in the context of categorization.

Second, when deciding on an intervention it is crucial to consider how the manipulated variable is causally connected to the outcome. Simulation studies by Meder and colleagues (2010) have shown that predictive validity is sufficient to pick a variable for an intervention as long as all variables considered are causally affecting the outcome. This is true even if the causal variables are confounded or interact with each

Journal: 11229-SYNT Article No.: 0162 TYPESET DISK LE CP Disp.:2012/7/23 Pages: 12 Layout: Small-X

other. However, predictive validities do not allow us to make good decisions if some variables are effects of the outcome. Effect variables may be highly predictive of the outcome, but their manipulation won't yield any desired effect (Meder et al. 2010). There is some evidence that people consider causal structure in decision making and rely on their background causal knowledge to decide on action (Hagmayer and Sloman 2009). They did so even when the statistical relation among the action and the outcome was kept constant.

Third, when being confronted with a situation in which the causal system changes or a new option becomes available, it is crucial to consider the underlying causal structure. Hagmayer and Meder (2012) showed that while making repeated decisions with respect to causal system, most participants spontaneously learnt how the variables of a system were causally related among each other and to an outcome. When a variable was unexpectedly removed from the system or a novel option was introduced, participants used their causal knowledge to pick the option that maximized the expected outcome under the new circumstances. Participants not acquiring causal knowledge but resorting to other forms of reinforcement learning were less able to react flexibly and make adaptive choices.

Forth, when learning about a complex causal system, pre-existing abstract causal knowledge is crucial to enable the induction process. Abstract causal knowledge (i.e., framework theories) speeds up learning by constraining the number of hypotheses that may explain the data observed. As Tenenbaum and colleagues (2011) have shown, abstract knowledge is necessary for efficient induction from limited data. Without these top-down constraints the data needed to differentiate between hypotheses quickly surpasses the available observations. Developmental research (e.g., Schulz et al. 2008) provides evidence supporting these claims. Using a general notion of causal laws (i.e., the assumption that there are types of causes generating certain types of effects) children were quickly able learn how different types of objects causally affected each other and they used this knowledge about causal relations on the type level to categorize new objects, make predictions and choose actions.

A number of boundary conditions for causal induction in contexts in which dynamic decision making occurs have to be pointed out (cf. Hagmayer et al. 2010), which may explain why so little causal learning and reasoning has been observed in DDM so far. First, people need feedback that enables the induction of causal models; that is, feedback that is a valid indicator of causal relationships (e.g., temporal contiguity, consequences resulting from interventions in the system). However, in dynamic decision making tasks the outcome feedback provided is often too coarse and the number of observations is too small to facilitate causal induction. Therefore, a second boundary condition is pre-existing theoretical knowledge, which constrains the number of potential hypotheses to enable induction of causal models. The problem is that many dynamic decision problems that have been studied are artificial and preclude the usage of prior knowledge (e.g., Berry and Broadbent (1995) sugar factory). Some may even violate everyday assumptions. The final boundary condition is that people are more likely to refrain from causal induction when they can successfully rely on prior experiences, as has been shown in recognition based decision making research (Brownstein and Read 2007). Only when the current problem went beyond prior experiences, people engaged in mental simulation – a process that depends on causal models

Journal: 11229-SYNT Article No.: 0162 ☐ TYPESET ☑ DISK ☐ LE ☐ CP Disp.:2012/7/23 Pages: 12 Layout: Small-X

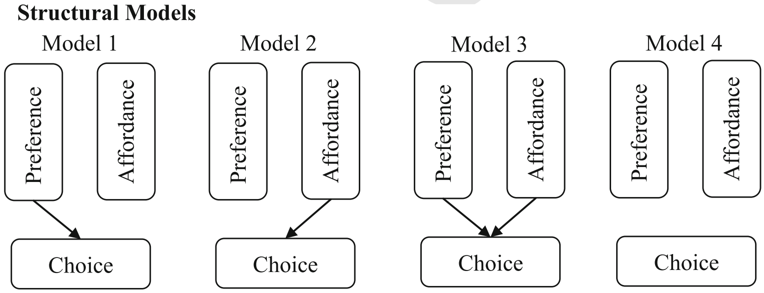
(cf. Klein 1998). Thus, we do not suggest that people always engage in causal considerations when dealing with a complex system or making judgments and decisions.

To sum up, we propose that people use abstract causal knowledge in induction (because it allows for rapid induction from sparse data), and causal model representations in decision making and inference (because it enables predictions and decisions when things change). They may do so even when it comes to complex dynamic systems like social interactions if certain boundary conditions of causal induction are met.

3 Causal Bayes nets

Causal Bayes nets offer a formal framework to model induction, inference, and decision making with respect to causal systems (Pearl 2000; Spirtes et al. 2000). They can be used to model the integration of new observations with previous evidence and preexisting causal knowledge by means of bayesian updating. The main advantage of Bayesian updating is that it follows the laws of probability theory. Importantly, this means that causal Bayes nets can be used to derive a rational model of everyday reasoning in complex domains (see final section). Only a brief introduction is provided at this point (see Pearl 2000, for a comprehensive introduction).

Causal Bayes nets use directed acyclic graphs to represent the structure of a causal system. Figure 1 shows a simple causal Bayes net model of preferential choice. There are at least four possible causal structures that could capture how personal preferences and external affordances may affect choice: choice may be due to (1) preferences, (2) external affordances, (3) both preferences and affordances or (4) neither



Probabilistic Model

P(Choice, Preference, Affordance, Model) =
P(Model 1) * [P(Choice|Preference) * P(Preference) * P(Affordance)]+
P(Model 2) * [P(Choice|Affordance) * P(Preference) * P(Affordance)]+
P(Model 3) * [P(Choice|Preference, Affordance) * P(Preference) * P(Affordance)] +
P(Model 4) * [P(Choice) * P(Preference) * P(Affordance)]

Fig. 1 A potential causal Bayes net model of preferential choice. The four graphs (Model 1–4) represent the four possible causal structures by which preference, affordance and choice could be causally connected. The probabilistic model represents the probabilistic dependencies entailed by the four graphs

Journal: 11229-SYNT Article No.: 0162 TYPESET DISK LE CP Disp.:2012/7/23 Pages: 12 Layout: Small-X

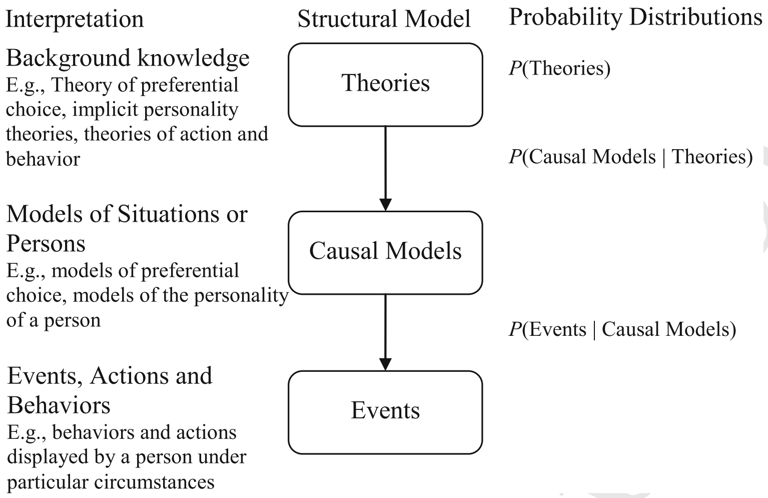


Fig. 2 Hierarchical causal Bayes net representing the relations between theories, causal models and events

(which means that choice is determined by other unknown factors). Each of these models has a certain probability representing its uncertainty (P(Model X)). In addition, each of the structural models is connected to a probabilistic model, which captures the probabilistic dependencies entailed by it. For example, Model 1 entails that choice is conditionally dependent on preference, but not on affordance. Taken together the probabilistic models of the four structural models yield the causal Bayes net of preferential choice.

Often theoretical background knowledge is available that constrains the potential structural models (Tenenbaum et al. 2011). For example, in the causal Bayes net model of preferential choice depicted in Fig. 1, the possibility that choice may affect external affordances was ignored; this is because choice temporally follows the presence of external affordances which entails that it cannot be its cause. The influence of such abstract theoretical knowledge can be modeled by Hierarchical Bayes nets. Figure 2 provides an illustration.

Hierarchical Bayes nets allow us to model various types of inferences, including induction (i.e., an inference about the causal structure underlying a situation and/or the theories constraining causal models), diagnosis and explanation (i.e., an inference about the most likely causes of an event), and prediction (i.e., an inference about similar or novel situations). The most fundamental inference mechanism is Bayesian updating. In Bayesian updating the *posterior probability* of a cause, a model, or theory conditional on an observation, is calculated using Bayes theorem.

3.1 Diagnosis

In the case of diagnosis, the following formal operations are needed to calculate the posterior probability of a potential cause: $P(\text{cause}_\text{j}|\text{event}) = \Sigma \text{i} P(\text{event}|\text{cause}_\text{j}, \text{model}_\text{i}) * P(\text{cause}_\text{j}, \text{model}_\text{i}) / \Sigma \text{i} \Sigma \text{j} P(\text{event}|\text{cause}_\text{j}, \text{model}_\text{i}) * P(\text{cause}_\text{j}, \text{model}_\text{i})$. That

Journal: 11229-SYNT Article No.: 0162 TYPESET DISK LE CP Disp.:2012/7/23 Pages: 12 Layout: Small-X

is the posterior probability of a cause is calculated by taking into account all possible causal models. Let us take for example a social gathering. Even though we know they had eaten an hour before, we observe that a fellow diner accepts a plate of sushi at a dinner party. The causal models depicted in Fig. 1 show the potential explanations for the observed behavior: social affordances and/or a strong liking for the food on offer. Background knowledge about dinner parties may entail that behaviors at dinner parties are strongly affected by norms (i.e., $P(\text{event}|\text{affordance} = \text{high}) >> P(\text{event}|\text{affordance} = \text{low})$ making Models 2 and 3 *apriori* more likely. Therefore Bayes rule entails that the behavior is most likely due to being polite. Interestingly, Model 3 (i.e., both preference and affordance affect choice) also implies that when we learn that the person enjoys sushi (i.e., has a strong preference) the likelihood of the diner just being polite drops. Learning that there is a strong preference *explains away* affordance (cf. Pearl 2000).

3.2 Induction

In the case of induction, Bayesian updating is used to make inferences about the posterior probability of causal models and/or theories conditional on the observed events: $P(\text{model}_{\text{i=x}}|\text{ events}) = P(\text{events}|\text{ model}_{\text{i=x}}) * P(\text{model}_{\text{i=x}})/\Sigma i P(\text{events}|\text{ model}_{\text{i}}) * P(\text{model}_{\text{i}})$. To illustrate, somebody moving to suburbia may not know the intricate social rules that govern dinner parties and therefore may assume that people's behavior is solely guided by preference (i.e., P(Model 1) is very high). However, by observing that people at dinner parties consume food even when it is utterly disgusting, the new neighbor will update his/her knowledge and believe that politeness is an important factor. Bayesian updating would entail that now Model 2 (affordances cause choice) is more likely than Model 1 (preferences cause choice).

Bayesian updating explains why successful induction often requires only a few observations. As different theories and causal models make differential predictions about the events to be expected, very few observations of the respective events are necessary to find out which causal model is most likely. For example, only Model 2 in Fig. 1 implies that behaviors will be displayed by an individual which contradict their preferences because they are conforming to social norms. Hence, few observations of this pattern suffice to favor Model 2 (i.e., to entail a higher posterior probability of Model 2 than the other models). Bottom-up theories like associative learning or other reinforcement learning theories have difficulty in explaining how rapid inductions can be generated from such limited experience; they would predict that inferences require many more learning experiences. Top-down theories, which assume that previous beliefs are important in learning, often do not explain how these beliefs affect induction in new contexts. Causal Bayes nets, by contrast, do provide an explicit account.

4 Bayes net models of lay-theories of action and behavior

To show the conceptual power of Bayes nets, we now turn to theories which describe how lay people make inferences in everyday social interactions or the observation

Journal: 11229-SYNT Article No.: 0162 TYPESET DISK LE CP Disp.:2012/7/23 Pages: 12 Layout: Small-X

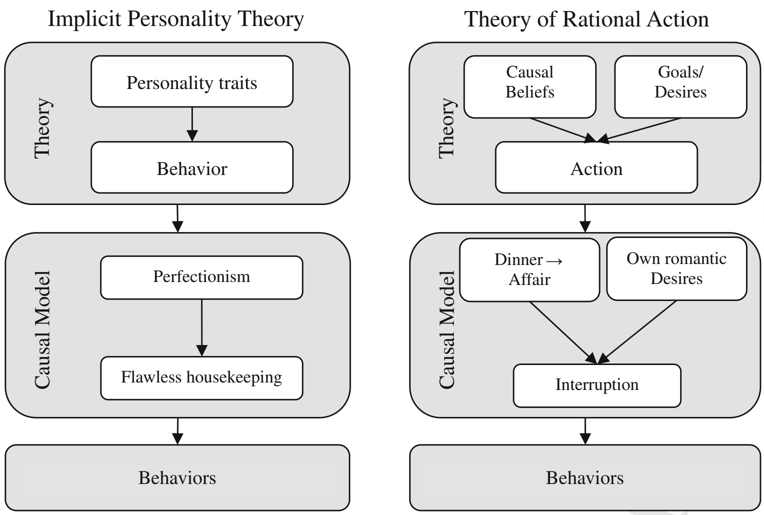


Fig. 3 Possible implementations of implicit personality theory (Asch 1946) and a theory of rational action as hierarchical Bayes nets

thereof (cf. Kunda 1999). We will show how these theories can be modeled using hierarchical Bayes nets. To illustrate we will use the popular TV series desperate housewives. We argue that one reason that viewers are easily able to understand, explain and predict the behaviors and actions of the main characters, is through the construction of causal models, which is guided by their lay-theories of action and behavior. Bayes nets allow us to model viewers' theories and inductive inferences, without assuming that viewers had any previous experiences with the suburban lifestyle in the US.

The first framework supposes that people have implicit theories of personality (Asch 1946) for which there is strong empirical support (cf. Borkenau 1992). This account can be implemented as the Hierarchical Bayes net depicted in Fig. 3. The fundamental assumption at the theory level is that people have multiple personality traits that are interconnected and that cause stable behavioral patterns over time. This implies that: (1) Personality traits are not independent of each other but form clusters (e.g., stereotypes), and (2) behaviors caused by personality traits should be relatively unaffected by external affordances, pressures and rational deliberations. These implications constrain and thereby ease the induction of causal models at the next level down (i.e., the construction personality profiles) from observed behaviors.

For all four main protagonists of desperate housewives (Susan, Bree, Gabriella, Lynette) "typical" behaviors are repeated 3–4 times in the series' pilot. Hence, viewers should have a good idea about their personalities despite the limited time in which they could have familiarized themselves with them (i.e. 45 min). Let us take Bree. There are several instances in which Bree displays perfectionism (e.g., intensive cleaning, perfect self-styling), but we only see two occasions in which she is shown to be vicious (i.e., unjustifiably accusing her son of drug abuse, causing an allergic reaction in her husband). An induction of a personality profile based on these few observations enables the viewer to understand and predict Bree's behavior in subsequent episodes (e.g., Bree's non-emotional handling of her marital problems).

Another lay-theory of behavior assumes that people operate as rational agents, deriving their actions from their beliefs and desires. This framework theory has already been implemented as a Hierarchical Bayes net (see Fig. 3 right hand side) and its predictions have been empirically supported (Baker et al. 2007; Goodman et al. 2009). The fundamental assumption at the theory level is that causal beliefs are used to anticipate the outcomes of available actions and desires to evaluate expected outcomes (by entailing a utility function over outcomes). The action expected to generate the best outcome is assumed to be chosen (Goodman et al. 2009). Again there are distinctive implications: (1) Actions are dependent on the causal beliefs of the decision maker and her/his goals and desires, (2) Beliefs and in consequence actions may change over time. Hence different actions are to be expected at different points in time, and different actions are to be expected when either beliefs or goals differ between people.

In desperate housewives a romantic triangle begins when Susan and Edie (a secondary character) become interested in an attractive bachelor named Mike. Using the rational agent framework, viewers would be able to infer Susan's beliefs and goals based on observations of: (1) her negative reactions to Edie's approaches towards Mike, (2) her own attempts to interact with Mike, and (3) a snoopy neighbor's comment about Edie hosting a private dinner. Considering Susan's belief that the dinner won't be an innocent occasion and her objective to begin a romantic relation with Mike, viewers would not be surprised to see that later in the episode Susan tries to sabotage the evening.

There are a number of other lay-theories of actions and behaviors which may also be modeled using Hierarchical Bayes nets (e.g., Malle 1999). In fact, a Hierarchical Bayes net may encompass several different framework-theories and use available evidence to update respective beliefs. In clear cut cases only one model will prevail (e.g., Bree's perfectionist personality), while in ambiguous cases (e.g., Susan burning down Edie's house) several models may account for the observations and may therefore be used for predictions.

5 Bayes nets of psychological theories versus Bayes nets as rational models

In the previous section we outlined how Bayes nets can be used to implement psychological theories describing how people make sense of interactions in social domains by resorting to lay-theories of behavior. Bayes nets require us to specify the abstract assumptions people may bring to bear, and the implications of these assumptions for the causal models people may derive from their observations. There are no constraints with respect to the abstract assumptions that could be implemented. As pointed out above, even potentially contradictory lay-theories could be integrated in the same hierarchical Bayes net. There is hardly any other formal theory that allows for integrating abstract knowledge, previous experiences, task-specific models and observations in such a systematic way. Constraint satisfaction networks representing the coherence among different kinds of beliefs (including theoretical assumptions and observations) are one exception (Thagard 2000; Wang et al. 2006), but these networks do not provide a rational model (see the following).

Journal: 11229-SYNT Article No.: 0162 TYPESET ☑ DISK LE CP Disp.:2012/7/23 Pages: 12 Layout: Small-X

Bayes nets may also serve as a methodological tool to plan future research. Bayes nets allow researchers to analyze the implications that follow from the data presented to participants given their prior knowledge and their previous learning experiences. They can also be used to test if the theoretical beliefs, people are assumed to have, are sufficient for the inductive inferences participants are supposed to make. Thereby these models allow researchers to test whether the experimental task presented to participants can be solved in the ways envisioned before. To give a practical example, a Bayes net model of a dynamic decision making task allows to test whether any knowledge about the underlying causal mechanisms could be derived from the data that is made available to participants. No other theory allows for such a prediction.

Bayes nets can also be used to provide a rational model of an optimal learner and decision maker (cf. Chater and Oaksford 2008). According to Anderson (1990) a rational analysis needs to include a formal model of the environment to which an agent is adapted. Causal Bayes nets are able to model any kind of causal system (Pearl 2000). Moreover, a hierarchical Bayes net of the cognitions of an agent allows us to compute the optimal inferences that can be derived from the agent's theoretical beliefs, task-specific causal models and available evidence. This is because such a model would specify probability distributions over all potential framework theories, causal models and parameters and would use Bayesian updating to compute the respective posterior probability distributions based on the observations made by the agent. As Bayesian updating is based on probability theory, all inferences are optimal under the given circumstances. Anderson (1990) also requires rational models to be descriptive of the behavior shown by agents. One may suspect that this may not be the case for hierarchical Bayes nets as no person would be able to consider and integrate all the probability distributions mentioned above. However, there is substantial evidence that rational Bayes net models are in fact descriptive of people's inductive inferences in many domains (Griffiths and Tenenbaum 2009; Tenenbaum et al. 2011).

Currently there are no rational Bayes nets that model the sophisticated and intricate exchanges that occur between people and our cognitions about them. However, so-called hidden Markov-models and Dynamic Bayes nets are able to model sequences of events over time (Griffiths et al. 2008) which we would argue are appropriate to represent dynamic interactions between agents, objects and events. While these models have been successfully applied to model sequences of gestures and language parsing (Griffiths et al. 2008) their success in accounting for lay people's ability to predict complex social interactions and their consequences is still an open question worth pursuing.

References

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale: Lawrence Erlbaum Associates.

Asch, S. E. (1946). Forming impressions of personality. *Journal of Abnormal and Social Psychology, 41*, 258–290.

Baker, C., Saxe, R., & Tenenbaum, J. (2007). Action understanding as inverse planning. *Cognition, 113*, 329–349.

Berry, D., & Broadbent, D. E. (1995). Implicit learning in the control of complex systems. In P. A. Frensch & J. Funke (Eds.), *Complex problem solving* (pp. 103–130). Hillsdale: Lawrence Erlbaum Associates.

Journal: 11229-SYNT Article No.: 0162 TYPESET DISK LE CP Disp.:2012/7/23 Pages: 12 Layout: Small-X

Borkenau, P. (1992). Implicit personality theory and the five factor model. *Journal of Personality, 60*, 295–327.

Broadbent, D. E., & Ashton, B. (1978). Human control of a simulated economic system. *Ergonomics, 78*, 1035–1043.

Brownstein, A., & Read, S. (2007). Situation models and memory: The effects of temporal and causal information on recall sequence. *Memory, 15*, 730–745.

Chater, N., & Oaksford, M. (2008). *The probabilistic mind: Prospects for Bayesian cognitive science*. Oxford: Oxford University Press.

Dienes, Z, & Fahey, R. (1995). Role of specific instances in controlling a dynamic system. *Journal of Experimental Psychology: Learning, Memory, Cognition, 21*, 848–862.

Dörner, D. (1989). *The logic of failure*. New York: Henry Holt.

Goodman, N., Baker, C., & Tenenbaum, J.B. (2009). Cause and intent. Social reasoning in causal learning. *Proceedings of the 31st annual conference of the cognitive science society.*

Gopnik, A., & Schulz, L. (2007). *Causal learning: Psychology, philosophy, and computation*. Oxford: Oxford University Press.

Griffiths, T. L., Kemp, C., & Tenenbaum, J. (2008). Bayesian models of cognition. In R. Sun (Ed.), *Cambridge handbook of computational cognitive modeling* (pp. 59–100). Cambridge: Cambridge University Press.

Griffiths, T. L., & Tenenbaum, J. B. (2009). Theory-based causal induction. *Psychological Review, 116*, 661–716.

Hagmayer, Y., & Lagnado, D. (2011). Causal models in judgment and decision making. In M. Dhami, A. Schlottmann, & M. Waldmann (Eds.), *Decision making as a skill* (pp. 351–404). Oxford: Oxford University Press.

Hagmayer, Y., & Meder, B. (2012). Repeated causal decision making. Journal of Experimental Psychology: Learning, Memory & Cognition.

Hagmayer, Y., Meder, B., Osman, M., Mangold, S., & Lagnado, D. (2010). Spontaneous causal learning while controlling a dynamic system. *Open Psychology Journal, 3*, 145–169.

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

Johnson, C. (2003). *Failure in safety-critical systems: A handbook of accident and incident reporting*. Glasgow: University of Glasgow Press.

Klein, G. (1998). *Sources of power: How people make decisions*. Cambridge: MIT Press.

Kunda, Z. (1999). *Social cognition: Making sense of people*. Cambridge: MIT Press.

Lipshitz, R., Klein, G., Orasanu, J., & Salas, E. (2001). Taking stock in naturalistic decision making. *Journal of Behavioral Decision Making, 14*, 331–352.

Malle, B. F. (1999). How people explain behavior: A new theoretical framework. *Personality and Social Psychology Review, 3*, 23–48.

Meder, B., Gerstenberg, T., Hagmayer, Y., & Waldmann, Y. (2010). Observing and intervening: Rational and heuristic models of causal decision making. *Open Psychology Journal, 3*, 119–135.

Osman, M. (2008). Evidence for positive transfer and negative transfer/anti-learning of problem solving skills. *Journal of Experimental Psychology: General, 137*, 97–115.

Osman, M. (2010). Controlling uncertainty: A review of human behavior in complex dynamic environments. *Psychological Bulletin, 136*, 65–86.

Osman, M. (2010). *Controlling uncertainty: Learning and decision making in complex worlds*. Oxford: Wiley.

Osman, M., & Speekenbrink (2011). Information sampling and strategy development in complex dynamic control environments. *Cognitive Systems Research,* 355–364.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. San Mateo: Morgan Kaufmann Publishers.

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge: Cambridge University Press.

Radvansky, G. A., Copeland, D. E., & Zwaan, R. A. (2005). A novel study: The mental organization of events. *Memory, 13*, 796–814.

Rehder, B. (2003). Categorization as causal reasoning. *Cognitive Science, 27*, 709–748.

Schulz, L. E., Goodman, N., Tenenbaum, J., & Jenkins, A. (2008). Going beyond the evidence: Preschoolers' inferences about abstract laws and anomalous data. *Cognition, 109*, 211–223.

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

Journal: 11229-SYNT Article No.: 0162 TYPESET ✓ DISK LE CP Disp.:2012/7/23 Pages: 12 Layout: Small-X

Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, prediction, and search*. Cambridge: MIT Press.
Sussman, A., & Oppenheimer, D. (2011). A causal model theory of judgment. *Proceedings of the 33rd annual conference of the cognitive science conference*.
Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science, 331*, 1279–1285.
Thagard, P. (2000). *Coherence in thought and action*. Cambridge: MIT Press.
Waldmann, M. R., Hagmayer, Y., & Blaisdell, A. P. (2006). Beyond the information given: Causal models in learning and reasoning. *Current Directions in Psychological Science, 15*, 307–311.
Wang, H., Johnson, T., & Zhang, J. (2006). The order effect in human abductive reasoning: An empirical and computational study. *Journal of Experimental & Theoretical Artificial Intelligence, 18*(2), 215–247.
Wyer, R. S., & Radvansky, G. A. (1999). The comprehension and validation of social information. *Psychological Review, 106*, 89–118.