What if? Counterfactual reasoning, pretense, and the role of possible worlds

Daphna Buchsbaum (daphnab@berkeley.edu)
Caren M. Walker (caren.walker@berkeley.edu)
Alison Gopnik (gopnik@berkeley.edu)
Department of Psychology, University of California, Berkeley, Berkeley, CA, USA

Nick Chater (nick.chater@wbs.ac.uk)
Behavioural Science Group, Warwick Business School, University of Warwick, Coventry, UK

David Danks (ddanks@cmu.edu)
Department of Philosophy, Carnegie Mellon University, Pittsburgh, PA, USA

Christopher G. Lucas (cglucas@cmu.edu)
Charles Kemp (ckemp@cmu.edu)
Department of Psychology, Carnegie Mellon University, Pittsburgh, PA, USA

Eva Rafetseder (Eva.Rafetseder@sbg.ac.at)
Josef Perner (josef.perner@sbg.ac.at)
Department of Psychology, University of Salzburg, Salzburg, Austria

Keywords: counterfactuals; pretense; causal reasoning

Symposium Moderator: Daphna Buchsbaum

Motivation

Counterfactual thinking, where one envisions alternative possible events and their outcomes, is hypothesized to be one of the primary ways in which we reason about causal relationships (e.g., Pearl, 2000; Woodward, 2003). Recent computational and experimental work suggests that both adults and children may reason about causality in a manner consistent with probabilistic graphical models – coherent, complex representations of causal structure that allow distinctive kinds of inferences (e.g., Gopnik et al., 2004; Griffiths & Tenenbaum, 2009). In particular, the causal models approach supports and distinguishes two types of inferences, predictions, on the one hand, and interventions, including counterfactual interventions, on the other. In predictions, we take what we think is true now as a premise and then use the model to calculate what else will be true. In counterfactuals, we take some value of the model that we currently think is not true as a premise, and calculate what would follow if it were.

Intuitively, childhood pretense bears a striking resemblance to counterfactual inference, but this relationship has not been widely explored. In general, pretend play seems paradoxical. Why should children spend so much time thinking about unreal worlds? Moreover, why would counterfactual inference itself be useful, since it is also about things that are not real? In this symposium we will explore the ways in which pretense and counterfactual thinking might be related (Buchsbaum, Walker & Gopnik; Rafetseder & Perner), the types of computations that might underly both kinds of thought (Lucas & Kemp; Chater) and the ways in which both might contribute to our causal understanding of the world, even without exposing us to new data (Chater; Danks).

Children's complex causal reasoning in pretend play

Authors: Daphna Buchsbaum, Caren M. Walker & Alison Gopnik In causal counterfactuals and in causal interventions we take some value of a causal model that we currently think is not true as a premise, and calculate what would follow if it were. We propose that these crucially important abilities – creating possible causal interventions and testing alternative causal hypotheses – depend on the same cognitive machinery that children use when they pretend: adopting a premise that is currently not true, creating an event sequence that follows from that premise, and quarantining the result of this process from reality.

Empirical results with preschool children support these ideas. Buchsbaum, Bridgers, Weisberg, and Gopnik (2012) found a significant and specific relationship between counterfactual inference and pretense in a causal task. In a new study, we gave children a complex causal structure involving four different variables (e.g. the sun comes up, which makes the rooster crow and the birds chirp, and the rooster crowing wakes up the farmer). Children's counterfactual inferences about this complex structure paralleled their inferences about pretense and both were significantly accurate. Interestingly, children were more likely to make "backtracking" counterfactual inferences when explicitly asked to reason counterfactually. In contrast, they were significantly more likely to treat the "fixed" variable as an intervention ("non-backtracking") when asked to pretend its value.

Representations, counterfactuals, and pretense

Author: Nick Chater Cognitive science views thought as computation. Computations are often conceived of as functions between input and output. But it may be more produc-

tive to explore the standard mantra from computer science, that a computer program consists of an algorithm operating over a data structure (see Chater & Oaksford, in press; and related work by Pearl, 2000). The behaviour of the algorithm is well-defined even if there are interventions into the contents of the data structure, during the computation (if, for example, the contents of the Turing machine tape, or a register in a pocket calculator, are modified, the algorithm will react in a well-defined way). Thus, a computer program can be viewed as defining a rich set of counterfactuals over possible modifications to the data structure, as the computation unfolds.

It turns out that this point of view provides a natural analysis of what it means for information to be represented: that which is represented can be modified by an intervention on the data structure. From this perspective, human counterfactual thinking and pretend play may have a common basis: they may be different sources of evidence concerning the flexibility with which the cognitive system is able to modify its own data structures, to reason about how the world might have been (e.g., modifying the representation of the past and tracing the consequences), or how it might differently be conceived (in children's play acting, modifying a representation of a banana to be a representation of phone).

Counterfactuals, causal learning, simulation, and pretense

Author: David Danks Causal structures provide information not just about what actually did occur, but about what would have occurred in various alternative scenarios. Counterfactuals are thus a key — in fact, necessary — guide for learning causal structures. Any method for learning about causal structures in the world must employ counterfactuals, whether explicitly or implicitly. The standard ways to judge counterfactuals for causal learning are through the use of interventions, or by focusing on "similar" (in relevant ways) cases. In many situations, however, these methods are too risky, too expensive, or infeasible for any number of other reasons. We must instead find other ways to judge counterfactuals.

Simulations based on one's present, uncertain causal beliefs provide a natural method for discovering surprising implications and incoherence in one's causal beliefs. We can use what we currently think about some causal structure to consider alternative possibilities, and thereby learn about our own (implicit) expectations and beliefs. Entirely mental simulation of causal relations is a challenging task, however, even for adults who have received training in it. One way to simplify the task is to ground the simulations in external, physical events that are analogous (in appropriate ways) to the underlying causal structure. That is, pretense and pretend play can help us learn about causal structures in the real world.

A unified theory of counterfactual reasoning

Authors: Christopher G. Lucas & Charles Kemp Bayesian networks have been used to account for many aspects of causal reasoning, including inferences about counterfactual scenarios. We present a Bayes net model of counterfactual reasoning that generalizes and extends the work of Pearl (2000). The model distinguishes between counterfactual observations and counterfactual interventions, and can reason about both backtracking and non-backtracking counterfactuals. Several experiments demonstrate that our model accounts better for human inferences than Pearl's original proposal and a more recent Bayes net account developed by Rips (2009).

Counterfactual reasoning vs reasoning counter-to-fact

Authors: Eva Rafetseder & Josef Perner Pretense has some affinity with counterfactual reasoning. It typically contains a counterfactual supposition that something (a prop) is something other than what it really is, and like reasoning, it proceeds to further suppositions in a constrained, non-arbitrary way. Rafetseder, Cristi-Vargas, and Perner (2010) have distinguished counterfactual reasoning from hypothetical reasoning counter to fact. Reasoning counter to fact makes suppositions that may contradict known facts and then uses known regularities to draw further inferences. Counterfactual reasoning is more constrained; it has to adhere to the nearest possible world constraint, i.e., the reasoning from the counterfactual assumption has to stay as close as possible to what actually happened. Conformity to this constraint develops rather late around 6 to 12 years. In hypothetical reasoning typical regularities (e.g., If [whenever] somebody takes shoes off floors tend to be clean) are applied to counterfactual questions (e.g., If Carol had taken her shoes off, would the floor be dirty or clean?) without regard to what actually happened (e.g., that Max had also been walking across the floor with dirty shoes). The importance of this distinction for pretense is that the affinity of pretense to counterfactuality is limited to reasoning with premises counter to fact.

References

Buchsbaum, D., Bridgers, S., Weisberg, D. S., & Gopnik, A. (2012). The power of possibility: Causal learning, counterfactual reasoning, and pretend play. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*, 2202-2212.

Chater, N., & Oaksford, M. (in press). Programs as causal models: Speculations on mental programs and mental representation. *Cognition*.

Danks, D. (2009). The psychology of causal perception and reasoning. In H. Beebee, C. Hitchcock, & P. Menzies (Eds.), *Oxford handbook of causation* (p. 447-470). Oxford University Press.

Gopnik, A., Glymour, C., Sobel, D., Schulz, L., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review*, *111*, 1-31.

Griffiths, T., & Tenenbaum, J. (2009). Theory-based causal induction. *Psychological review*, *116*(4), 661–716.

Pearl, J. (2000). *Causality.* Cambridge University Press.

Rafetseder, E., Cristi-Vargas, R., & Perner, J. (2010). Counterfactual reasoning: Developing a sense of "nearest possible world". *Child Development*, *81*(1), 376-289.

Rips, L. J. (2009). Two causal theories of counterfactual conditionals. *Cognitive Science*, 1-47.

Woodward, J. (2003). *Making things happen: A theory of causal explanation.* Oxford University Press.