Causal vs. Evidential Decision Making in Newcomb's Paradox

York Hagmayer (york.hagmayer@bio.uni-goettingen.de)
Department of Psychology, University of Goettingen
Gosslerstrasse 14, 37073 Goettingen, Germany

Steven A. Sloman (Steven_Sloman@brown.edu)
Cognitive & Linguistic Sciences, Brown University,
Box 1978, Providence, RI, USA

Newcomb's Paradox

Newcomb's Paradox (Nozick, 1969) has been regarded as a critical test for causal decision making theories. Decision makers are confronted with the following problem (abbreviated by the authors):

"You have great confidence in a particular demon's ability to predict your choices. This demon is going to predict your choice in the following situation. There are two boxes, a transparent one that contains $1000 and an opaque one that contains either $1,000,000 or nothing. You can either choose to take what is in both boxes or to take only what is in the opaque box. You know (and the demon knows you know, and you know the demon knows, etc.) that if the demon predicts you will take what is in both boxes, he will put nothing in the opaque box. But if the demon predicts that you will choose only the opaque box, he puts the $1,000,000 in it. First, the demon makes his prediction, then he puts the money in the opaque box or not based on his prediction (you don't get to see whether he does or not), then you make your choice. So far the demon correctly predicted participants' choices most of the time. Imagine you are this situation, what do you choose?"

Based on the evidence that the demon was able to correctly predict participants' choices most of the time, it seems rational to open only the opaque box, because it has the higher evidential expected utility. However, based on the information that the decision is made after the demon has made his prediction and cannot change the amount of money he allocated in the opaque box, it seems rational to prefer to open both boxes, because the choice cannot affect the demon's prediction anymore and it gives an additional $1000 (higher causal expected utility).

Causal Analysis

Although the temporal sequence indicates that no causal relation exists among the decision maker's choice and the demon's prediction, a closer analysis reveals that the paradox is ambiguous from a causal point of view. First, the high accuracy of the demon's predictions indicate a causal relation among the deliberately made decisions and the demon's estimates (see Lagnado et al., in press, for evidence that correlations among actions and outcomes are considered as valid indicators of causality). Second, as a causal relation is possible, the exact timing becomes crucial. In the question asked it is not specified, whether the demon already made his prediction or is about to make his prediction. If an unknown causal relation is possible and the money is not yet allocated, it is perfectly rational from a causal point of view to prefer the one-box option. In contrast, if the demon already made his prediction, the possibly existing causal relation is blocked, and it is better to open both boxes.

In sum, Newcomb's Paradox presents conflicting cues about the causal relatedness of the choice and the demon's prediction. Moreover, important information about the timing of the decision is missing, which would disambiguate the given information. The causal model theory of choice (Hagmayer & Sloman, 2005) predicts that once these information are provided participants should show a clear preference for the option with the higher causal expected utility (i.e. the two-boxes option).

Empirical Evidence

We confronted participants in two studies with a neuroscientific version of Newcomb's Paradox in which the demon was replaced by an algorithm analyzing brain scans. Given the ambiguous original problem, a large number of participants chose the one-box option. Choices were justified by various reasons, including causal structure, evidential relations, and the principle of dominance.

In the second experiment the underlying causal structure and the timing of the decision was specified and manipulated. Participants now preferred the option with the higher causal expected utility over the option with the higher evidential expected utility. In addition, a majority of participants now justified their decision by causal considerations. These results support the causal model theory of choice.

References

Hagmayer Y., & Sloman, S. A. (2005). Causal models of decision making: Choice as intervention. *Proceedings of the Twenty-Seventh Annual Conference of the Cognitive Science Society*, Stresa, Italy.

Lagnado, D. A., Waldmann, M. R., Hagmayer Y., & Sloman, S. A. (in press). Beyond covariation: Cues to causal structure. In Gopnik, A. & Schulz, L. (Eds.), *Causal learning: Psychology, philosophy, and computation*. Oxford: Oxford University Press.

Nozick, R. (1969). Newcomb's problem and two principles of choice. In N. Rescher (ed.), *Essays in honor of Carl G. Hempel* (pp. 107-133). Dordrecht: Reidel.