Causal Reasoning from Forces

Aron Barbey (abarbey@emory.edu) and **Phillip Wolff (pwolff@emory.edu)**
Department of Psychology, Emory University
532 N. Kilgo Circle, Atlanta, GA 30322 USA

Everyday language suggests we use forces in causal reasoning. We say, for example, *The force of his argument cannot be denied*, or *Your argument doesn't go through*, or *The social progressive argument has tremendous moral force*. Recent work in force dynamics suggests how these intuitions might be fleshed out computationally.

According to force dynamics, people represent causal relationships in terms of configurations of force (Talmy, 1988; Wolff & Zettergren, 2002). One force, $\mathbf{F_A}$, is associated with an affector, that is, the entity that acts on another entity. Another force, $\mathbf{F_B}$, is associated with a *patient*, the entity that is acted on by the affector. A third force, $\mathbf{F_{BA}}$, is the resultant produced from the addition of these forces. Various causal concepts entail different configurations of force. For example, as shown in Figure 1, in a CAUSE configuration $\mathbf{F_A}$ and $\mathbf{F_B}$ are in opposition and the resultant, $\mathbf{F_{BA}}$, is towards the endstate, E.

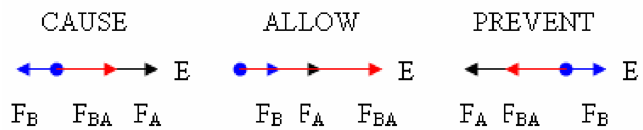

Fig. 1. Causal concepts and their configurations

The *transitive dynamics model* specifies how these configurations of force are combined to generate conclusions. Consider the inference problems in Figure 2.

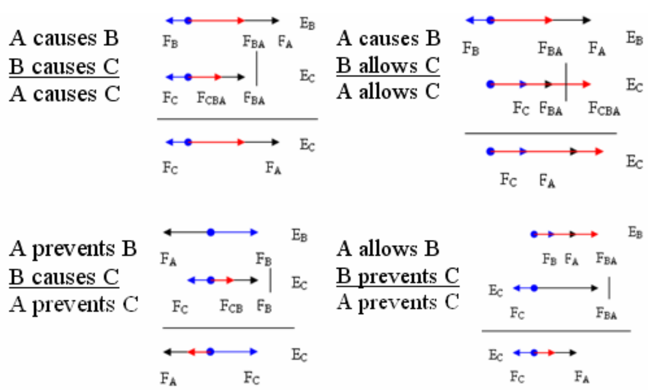

Fig. 2. Transitive arguments and configurations of force

The model holds that people use the resultant of the 1st premise as the affector vector in the 2nd premise in the case of CAUSE and ENABLE relations, but not PREVENT relations. In the case of PREVENT relations, the magnitude of the resultant may equal 0, so the model holds that in the case of PREVENT relations, people use the force of patient in the 1st premise, $\mathbf{F_B}$, as the affector vector in the 2nd premise. These common forces (either $\mathbf{F_{BA}}$ or $\mathbf{F_B}$) connect the premises and allow the forces between them to be oriented with respect to one another.

The model holds that the affector in the conclusion is based on the affector vector in the 1st premises, $\mathbf{F_A}$, and that the endstate in the conclusion, $\mathbf{E_C}$, is based on the endstate from the 2nd premise. The patient vector in the conclusion is based on the patient vectors in the 1st or 2nd premises. Roughly stated, the force of the patient is towards the endstate, $\mathbf{E_C}$, if either $\mathbf{F_C}$ or $\mathbf{F_B}$ point toward that endstate, implying that the patient would make progress towards the endstate in the absence of $\mathbf{F_A}$.

Table 1. Conclusions drawn by participants (N = 19)

Second premise	**First premise**	
	A causes B	**A allows B**
B causes C	A causes C (17)	A allows C (13)
B allows C	A allows C (18)	A allows C (19)
B prevents C	A prevents C (18)	A prevents C (15)
¬B causes C	A prevents C (17)	A prevents C (14)
	A prevents B	**¬A causes B**
B causes C	A prevents C (15)	¬A causes C (18)
B allows C	A prevents C (16)	¬A allows C (17)
B prevents C	A causes/allows C (6)	¬A prevents C (18)
¬B causes C	A causes C (11)	¬A prevents C (10)

The predictions of the model were tested in a replication of Goldvarg and Johnson-Laird's (2001) Experiment 4. Participants reviewed sixteen syllogisms that involved psychological terms (e.g., *obedience causes motivation, motivation causes eccentricity*). Participants indicated what, if anything, followed. Table 1 shows the conclusion predicted by the transitive dynamics model along with the number of participants choosing that conclusion. As can be seen, the model predicted the modal response for all but one of the arguments, prevent-prevent. Here we believe participants' modal response, "prevents," was influenced by atmosphere of the argument. The results also replicated the findings of Goldvarg and Johnson-Laird (2001), except for three arguments, but it is for these three arguments that the predictions of our model differ from those of Goldvarg and Johnson-Laird's. In sum, the results demonstrate how reasoning might be accomplished through chains of forces.

References

Goldvarg, E., & Johnson-Laird, P. (2001). Naive causality: A mental model theory of causal meaning and reasoning. *Cognitive Science, 25*, 565-610.

Talmy, L. (1988). Force dynamics in language and cognition. *Cognitive Science, 12*, 49-100.

Wolff, P., & Zettergren, M. (2002). A vector model of causal meaning. *Proceedings of the twenty-fourth annual conference of the Cognitive Science Society* (pp. 944-949). Mahwah, NJ: Lawrence Erlbaum Associates.