Assessing Psychological Theories of Causal Meaning and Inference

Sergio E. Chaigneau (sergio.chaigneau@uai.cl)
Escuela de Psicología, Universidad Adolfo Ibáñez, Av. Diagonal Las Torres 2640
Peñalolén, Santiago-Chile

Aron K. Barbey (barbeya@ninds.nih.gov)
Cognitive Neuroscience Section, National Institute of Neurological Disorders and Stroke, Building 10, Room 7D49,
MSC 1440, Bethesda, MD 20892-1440

Abstract

We focused on three theories of causal meaning: mental model, force dynamics, and causal model theory. These theories differ in the ascribed meaning of causal verbs like *cause*, *enable*, *allow*, and *prevent*, and also differ in the mechanisms they propose for causal inference. In Experiment 1, we tested their mechanism for causal inference. As predicted by causal model theory, given problems of the form A prevents B / B prevents C, participants concluded that A causes or allows C. In Experiment 2, we tested these theories' proposed meanings for *allow* and *enable*. As predicted by causal model theory, participants labeled conjunctive causes as enablers.

Keywords: Causal meaning, causal inference, causal reasoning, enabling relations.

Causal knowledge provides the basis for higher-level thought, supporting explanatory and predictive inferences that are essential for learning and controlling the environment to achieve goals. The cognitive foundations of causal meaning and inference, however, remain largely unknown. How do people represent the meaning of *cause*, *allow*, and *prevent*, and how is this knowledge applied to support further inferences? In this work, we investigated these questions with a focus on two central issues. First, we focused on how causal relations are combined to derive inferences. Second, we examined the cognitive representation of *allow*. We framed these issues in the context of existing theories of causal meaning and inference, assessing the predictions of the mental model theory (Goldvarg & Johnson-Laird, 2001), force dynamics theory (Wolff, 2007; Barbey & Wolff, 2006, 2007), and causal model theory (Sloman, Barbey & Hotalling, 2008).

Mental Model Theory Mental model theory proposes that causal relations are represented by a distinct set of mental models or possible state of affairs (Goldvarg & Johnson-Laird, 2001). Within this framework events are represented by capitals (e.g., A), their presence in lowercase (a) and their absence in lowercase with a tilde (~a). A causes B represents three possibilities: the occurrence of A and B (a, b), the absence of A in the presence of B (~a,b), and a null event in which neither A nor B occur (~a, ~b). A allows or allows B represents the occurrence of A and B (a, b), the presence of A in the absence of B (a, ~b), and a null event in which neither occur (~a, ~b). Finally, A prevents B represents the presence of A in the absence of B (a, ~b), the absence of A in the presence of B (~a, b), and the null event (~a, ~b).

Mental model theory holds that people typically represent only one model for reasoning, namely the first of each set above (Johnson-Laird & Byrne, 2002). This is a factual or explicit model. The second and third models are counterfactual models that often remain implicit. Using only factual models to reason, sometimes produces invalid conclusions (as shown in Table 1).

When multiple causal statements are involved, Goldvarg and Johnson-Laird propose models are combined following a set of rules. For statements of the general form "A *relation* B" and "B *relation* C", consistent models are combined, resulting inconsistent and redundant models are eliminated, and the resulting possibilities holding between A and C are inspected to find which relation, if any, corresponds. Because people generally use only factual models to reason, conclusions are sometimes invalid or illusory. In contrast, if people are able to use factual and counterfactual models, conclusions are always valid (see Table 1 for examples of mental model derivations).

Table 1: Example derivations in mental model theory.

Problem	Derivation	Conclusion
A causes B	a, b	
B prevents C	b, ~c	
	a, b, ~c	Premises integrated
	a, ~c	A prevents C (valid)
A prevents B	a, ~b	
B prevents C	b, ~c	
	…	Premises not integrated
	a, ~c	A prevents C (invalid)

In Experiment 1, we were interested in evaluating predictions for double prevent problems. As Table 1 shows, because A prevents B cannot be integrated with B prevents C (it would produce an inconsistent model that includes both b and ~b), mental model theory predicts people will consider only the first event of the first premise and the second of the last, erroneously concluding that A prevents C.

Force Dynamics Theory Force dynamics theory proposes that mental representations of causal relations reflect one of the properties of causes in the physical world: the interaction of forces (Talmy, 1988; Wolff, 2007; Barbey & Wolff, 2006; Barbey & Wolff, 2007).

In force dynamics, causal relations represent the interaction of two main entities: an affector and a patient (the entity acted upon by the affector). In Wolff's (2007) formulation, these entities are analyzed in terms of three dimensions: (1) the tendency of the patient for the endstate; (2) the presence or absence of concordance between the affector and the patient; (3) progress toward the endstate (i.e., whether or not the endstate occurs). Table 2 summarizes how these dimensions represent the concepts *cause*, *allow*, and *prevent*. For example, the sentence "The explosion caused the bridge to collapse," represents a state of affairs in which the patient (the bridge) did not have a tendency to collapse, the affector (the explosion) acted against the patient, and the result (the collapse of the bridge) occurred.

Table 2. Force dynamic representations of several causal concepts.

	Patient tendency for the endstate	Affector-patient concordance	Endstate approached
Cause	No	No	Yes
Allow	Yes	Yes	Yes
Prevent	Yes	No	No

Force dynamic dimensions are formally represented in the language of vectors. As Figure 1 illustrates, the patient, B, has a tendency for the endstate, E, when the vector associated with the patient points in the same direction as the vector that specifies the endstate. Thus, the patient vector points in the same direction as the endstate vector for *allow* and *prevent*, but not in the case of *cause*. Concordance occurs when the vectors associated with the patient and affector point in the same direction. As illustrated in Figure 1, the patient and affector are concordant for *allow*, but not in the cases of *cause* and *prevent*. Finally, the result is expected to occur when the resultant vector points in the same direction as the endstate vector, a property represented by *cause* and *allow*, but not *prevent*.

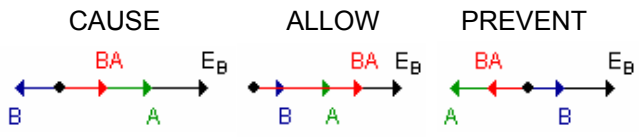
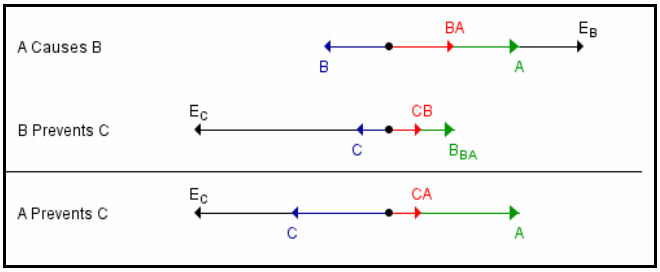


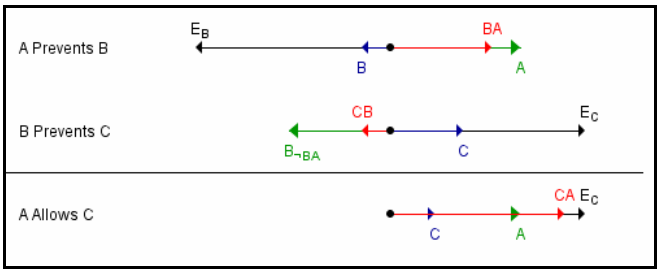
Figure 1. Configurations of force associated with cause, allow, and prevent. A = the affector force, B = the patient force, BA = the resultant of A and B, E = endstate.

The force dynamics theory has been extended to inferences drawn from multiple causal relations (Barbey & Wolff, 2006; 2007). In the context of transitive inference, this is accomplished by representing the configuration of forces that underlie A's relationship to B, and B's relationship to C, and then linking these premises to draw a transitive inference about A's relation to C. As Figure 2 illustrates, the *transitive dynamics model* proposes that the premises are connected by using the resultant vector in the first premise (BA) as the affector vector in the second (B_{BA}). The resultant vector points in the same direction as the affector in the second premise (see Figure 2, Panel A) unless the B terms in the two premises conflict (i.e., if one is negated; see Figure 2, Panels B and C).

Panel A



Panel B



Panel C

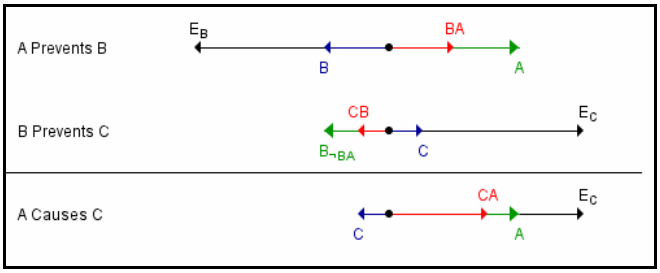


Figure 2. Examples of derivations in force dynamics theory. For prevent-prevent problems, conclusions can be *allow* (panel B) or *cause* (panel C), depending on the particular configuration of forces.

A conclusion is drawn by forming a new configuration of forces based on the two premises. Specifically, the affector in the conclusion is the affector from the first premise; the endstate vector in the conclusion is the endstate vector from the last premise; and the patient in the conclusion is the

resultant of the patient vectors in the premises. The resulting configuration of vectors can then be interpreted according to the semantics for individual causal relations (see Figure 1).

Causal Model Theory Causal model theory (Sloman et al., 2008) utilizes the graphical formalism of causal Bayes nets (Pearl, 2000; Spirtes, et al., 1993; Sloman, 2005). This framework offers a way to represent and reason from causal systems using nodes and links in the form of acyclic causal graphs (a graphical formalism linked to Bayes nets; Pearl, 2000; Spirtes, Glymour, & Scheines, 1993). In a causal graph, events are represented as nodes and mechanisms as links (e.g., *ultraviolet radiation* → *skin damage*). A link between A and B represents a causal mechanism that has A as one of its inputs and B as the output (e.g., *ultraviolet radiation* [A] alters DNA in skin cells, which results in *skin damage* [B]; for discussion, see Sloman & Hagmayer, 2006).

Causal graphs are isomorphic to structural equations (Sloman et al., 2008). In a structural equation, effects are represented as functions of other events (e.g., B := A), with the symbol ":=" meaning that the function is asymmetric (i.e., an effect is a function of its cause, but a cause is not a function of its effect). By modifying the qualitative structure of causal models, different causal relations can be represented.

The verb *cause* represents B := A (i.e., the presence of A entails B, whereas B is uncertain in the absence of A). The verbs *enable* and *allow* represent B := A & C (i.e., A and C are conjunctively necessary causes of B, but none is a sufficient cause). The verb *prevent* represents either B := ~A (i.e., B occurs whenever A is not present, but if A is present, B will not occur) or B := ~A & C (i.e., A needs to be absent and C present for B to occur).

In the face of multiple causal relations (e.g., A causes B, and B prevents C), conclusions are drawn by substituting the repeated term in the second premise with its equivalent from the first premise (see Table 3 for examples of derivations).

Table 3. Example derivations in causal model theory. Two meanings of *prevent* are used.

Problem	Derivations	Conclusion
A causes B	B:=A	
B prevents C	C:= ~B	
	C:= ~A	A prevents C
A prevents B	B:= ~A	
B prevents C	C:= ~B	
	C:= ~(~A)	A causes C
A prevents B	B:= ~A	
B prevents C	C:= ~B & D	
	C:= ~(~A) & D	A allows C

The Experiments

Our experiments investigated the process of causal inference (Experiment 1) and the causal meanings (Experiment 2) assumed in the mental model, force dynamics, and causal model frameworks. Participants in our experiments were all native Spanish-speakers.

Experiment 1: Reasoning from *Prevention*

In causal inference, two or more causal events must be integrated to support a conclusion. We focused specifically on inferences drawn from a chain of prevent relations (A prevents B, B prevents C, what is the relation between A and C?). Double prevent problems are interesting, because they have produced inconsistent results (Goldvarg & Johnson-Laird, 2001; Barbey & Wolff, 2007), and because they have a diagnostic value in distinguishing between our focal theories. Whereas transitive dynamics and causal model theories predict participants will infer that A *causes* or *allows* C (see Figure 2 and Table 3), mental model theory predicts participants will conclude that A *prevents* C (Goldvarg & Johnson-Laird, 2001; see Table 1).

Experiment 1 further examined the degree to which conclusions are drawn from dual processing systems that utilize prior knowledge and deliberative reasoning (Barbey & Sloman, 2007; Kahneman & Frederick, 2005). We think that inconsistencies found in the literature, may be due to people finding it difficult to reason about double prevent problems. These problems involve negations, and people have difficulties in representing things that are absent (Hasson & Glucksberg, 2006). This difficulty may serve as an incentive to use prior knowledge, instead of reasoning, to solve double prevent.

Method

In this experiment, prior knowledge was manipulated by presenting conclusions that by themselves suggested either *cause* or *allow*, or suggested *prevent*. An example of problems with conclusions that suggested *cause* or *allow* is:

Vaccines prevent infections
Infections prevent good health
Vaccines _____ good health

An example of problems with conclusions that suggested *prevent* is:

Financial planning prevents compulsive saving
Compulsive saving prevents bankruptcy
Financial planning _____ bankruptcy

A processing theory we envision about the interaction of prior knowledge and causal reasoning, is the following. Both processes occur in parallel, with prior knowledge facilitating or interfering with reasoning, depending on whether prior knowledge is somehow consistent or inconsistent with it. When processing demands are too high (i.e., reasoning is difficult), people rely solely on prior knowledge (i.e., they fall back on a heuristic).

This simple theory, gave us leverage to test causal model and force dynamics against mental model's predictions. Recall that, for double prevent problems, causal model and force dynamics predict that reasoning will lead participants to conclude *cause* or *allow*. In the condition where prior knowledge suggests the conclusion should be *cause* or *allow*, either theory equipped with our processing assumptions predicts that responses coming from prior knowledge and responses coming from reasoning, should contribute together to produce a higher frequency of *cause* or *allow* responses than *prevent* responses. This, because consistent prior knowledge will facilitate reasoning. In contrast, in the condition where prior knowledge suggests the conclusion should be *prevent*, both theories predict about equal frequencies of *prevent* and of *cause* or *allow* choices. This same prediction obtains if prior knowledge interferes with reasoning (i.e., the tendency to respond *prevent* interferes with responding *cause* or *allow*), or if participants that find problems difficult fall back on a heuristic and arrive at the *prevent* response, while those who reason causally arrive at *cause* or *allow* responses.

Interestingly, mental model theory makes exactly opposite predictions. Recall that mental model predicts reasoning will lead participants to the *prevent* conclusion. In the condition where prior knowledge also suggests *prevent*, coupling mental model with our processing theory predicts that these two sources of responses will contribute together to produce a higher frequency of *prevent* choices than *cause* or *allow* choices. Again, this is because consistent prior knowledge should facilitate reasoning. In contrast, in the condition where prior knowledge suggests the conclusion should be *cause* or *allow*, mental model predicts about equal frequencies of *prevent* and of *cause* or *allow* choices. This same prediction obtains if prior knowledge interferes with reasoning (i.e., the tendency to respond *cause* or *allow* interferes with responding *prevent*), or if participants that find problems difficult fall back on a heuristic and arrive at *cause* or *allow* responses, while those who reason causally arrive at the *prevent* response.

Materials and Participants Participant's responded on a booklet that contained one problem per page. In total, they received 8 two-term problems, 4 from each condition of the prior knowledge factor. After reading each problem, participants had to choose either *prevent*, *cause*, *allow* or *nothing follows* as possible responses. Within each condition, two problems came from the psychological/social domain, and two came from the physical/biological domain. Participants were 48 University Adolfo Ibáñez's undergraduates, who participated for course credit.

Design and Procedure Experiment 1 employed a repeated measures design with two within participants factors: prior knowledge (suggests *prevent*, suggests *cause* or *allow*) and response (*cause* or *allow* responses, *prevent* responses). The order of the 2 levels of the prior knowledge factor, the order of domains (physical/biological first, psychological/social first), and the order of problems within each domain (direct, inverse), were completely crossed.

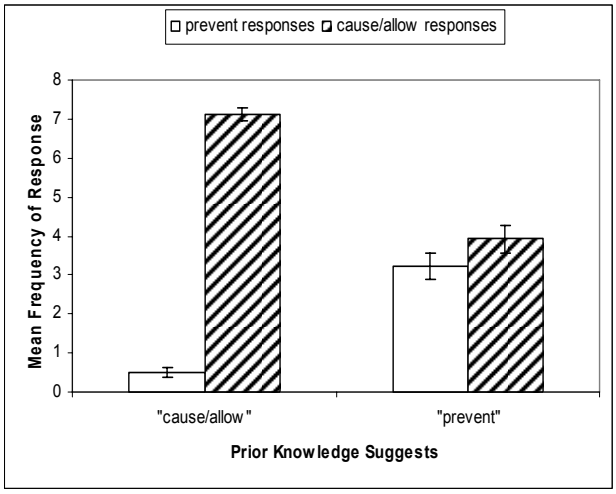


Figure 3. In Exp 1, mean frequencies of *prevent* responses, and of *cause* or *allow* responses, for double prevent problems where prior knowledge suggests *prevent* as response, versus problems where prior knowledge suggests *cause* or *allow* as responses. Error bars are standard errors.

Results

Data were submitted to a 2 x 2 repeated measures ANOVA. As Figure 3 shows, the pattern of results closely matches the predictions of causal model and force dynamics theories. When found a main effect of prior knowledge. When prior knowledge suggested c*ause* or *allow*, there were fewer *nothing follows* conclusions than when prior knowledge suggested *prevent* ($F(1, 47) = 11.87$, $MSe = .21$, $p < .001$, $R^2 = .20$, power = .92). We also found a main effect of response type. *Cause* or *allow* responses were significantly more frequent than *prevent* responses ($F(1, 47) = 87.56$, $MSe = 7.29$, $p < .001$, $R^2 = .65$, power = 1). Most importantly, the interaction between the two factors was also significant ($F(1, 47) = 87.98$, $MSe = 4.78$, $p < .001$, $R^2 = .65$, power = 1). As predicted by causal model and force dynamics, when prior knowledge suggested *cause* or *allow*; *cause* and *allow* responses were significantly more frequent than *prevent* responses. When prior knowledge suggested *prevent*; *prevent* responses were about as frequent as *cause* and *allow* responses. We found a similar pattern when means were computed separately for each domain (i.e., physical/biological or psychological/social domains).

Discussion

These findings support the predictions of the causal model and force dynamics theories, providing evidence that the *cause* or *allow* conclusion follows from a chain of prevents (as observed in Barbey & Wolff, 2007). Furthermore, our results suggest that causal reasoning is supported by dual processing systems that incorporate prior knowledge and

deliberative reasoning. If participants guided their responses only by prior knowledge, they would have been insensitive to causal structure. If participants guided their responses only by causal reasoning, they would have been insensitive to prior knowledge. In contrast, the interaction shown in Figure 3, supports a dual process account. In this same vein, an interesting finding is that we found fewer *nothing follows* responses when prior knowledge suggested *cause* or *allow* than when it suggested *prevent*. This is consistent with facilitation and interference effects of prior knowledge on reasoning.

Experiment 2: Representation of *Allow*

Experiment 2 investigated the cognitive representation of *allow*, examining a central prediction of the causal model theory. Recall that in this theory, the meaning of *cause* is being a sufficient antecedent. Therefore, if the antecedents of an event are not individually sufficient, but rather jointly necessary for the effect to occur, people should label it as an allow event. Other causal models with two antecedents, should not be perceived as allow events. A causal model with two antecedents that independently produce the effect (A → C ← B) does not represent *allow*, nor does a model with two antecedents that form a chain of causes (A → B → C). The mental model and force dynamics theories do not make this distinction and therefore would not predict that *allow* represents antecedents that are jointly necessary for the effect.

To investigate this issue, Experiment 2 measured the proportion of *allow* labels applied to written and graphical descriptions of events representing each of the models described above.

Method

Materials and Participants We created four sets of events, 2 sets from the *psychological* domain (*psych-1* and *psych-2*) and 2 from the *physical* domain (*phys-1* and *phys-2*), all involving two antecedents and one consequent. By manipulating the verbal and graphical description of relations between events in a set, the same antecedents and consequents could be presented in an independent causes model or in a conjunctive causes model. The following are examples of verbal descriptions for the same set of events presented as different models (see the corresponding graphical depictions in Figure 4).

Chain model:

> If A wins a prize,
> B will become sad.
> If B becomes sad,
> C will be mad.

Independent causes model:

> A does not affect B, nor does B affect A.
> If A wins a prize,
> C will be mad.
> If B becomes sad,
> C will be mad.

Conjunctive causes model:

> A does not affect B, nor does B affect A.
> If A does not win a prize, and B becomes sad,
> C will not be mad.
> If A wins a prize, and B does not become sad,
> C will not be mad.
> Only if A wins a prize and B becomes sad,
> C will be mad.

After receiving a vignette and its accompanying graph, participants had to choose between three alternatives to describe the relations that held between A and C, and between B and C (*cause*, *allow*, *unrelated*). Only the first two responses were of interest to us. Consequently, participants could show 3 different patterns of responding: *A causes C* and *B causes C* (the *cause-cause* pattern); *A causes C* and *B allows C*, or *A allows C* and *B causes C* (the *cause-allow* pattern); *A allows C* and *B allows C* (the *allow-allow* pattern). Problems were presented in a booklet, with one problem (i.e., a vignette and its corresponding graph) per page. Participants were 36 University Adolfo Ibáñez's undergraduates, who participated for course credit.

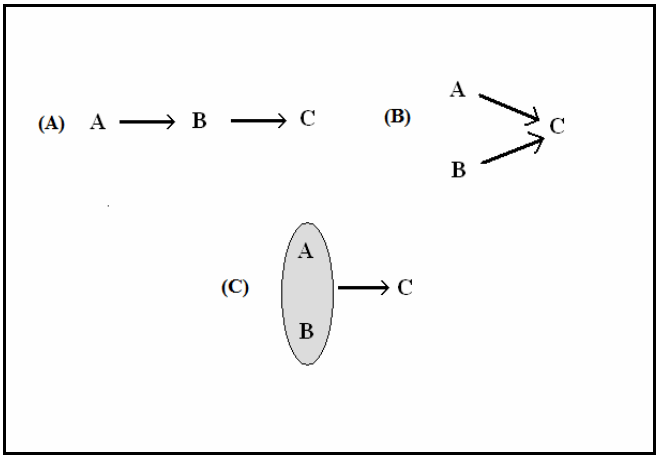


Figure 4: Graphical depictions of models used in Exp 2. Chain (A), Independent (B) and Conjunctive causes (C).

Design and Procedure Four basic versions of the materials were created. Each version contained 3 problems different from each other, 2 from one domain and 1 from the other. For each participant, one problem was a chain model, another a conjunctive causes model, and yet another an independent causes model. Each participant contributed one judgment for each type of model, and for each model we collected 9 judgments from each event set (*psych-1*, *psych-2*, *phys-1*, *phys-2*) for a total of 36 judgments. The order of problems in each basic version was counterbalanced, producing a total of 12 versions (3 orders for each basic version).

Results

For each model, we counted the number of cause-cause, cause-allow or allow-cause responses, and allow-allow responses across participants. Participants never chose to describe the events as *unrelated*. As Table 4 illustrates, our predictions were upheld by the data. For the chain models, the Chi-Square test showed that the three patterns of responding occurred with a similar frequency (χ^2 (2, N=36) = 4.0, p = 0.14). For the independent causes models, the Chi-Square test also showed that the three patterns of responding occurred with about the same frequency (χ^2 (2, N=36) = 4.17, p = 0.12). But for the conjunctive causes model, the Chi-Square test showed that frequencies were far from randomly distributed (χ^2 (2, N=36) = 21.33, p < 0.001), with the allow-allow pattern being much more frequent than the other patterns. Although our sample size did not allow us to make separate tests by domain, the pattern of frequencies was stable. In particular, in both physical and psychological problems, the allow-allow pattern of response was the most frequent.

Table 4. In Exp 2, frequencies for each response pattern within each causal model (expected frequencies were cause-cause = 9, cause-allow = 18, allow-allow = 9).

	Model		
	Chain	Independent	Conjunctive
cause-cause	14	14	5
cause-allow	16	13	10
allow-allow	6	9	21

Discussion

The results support the causal model theory's formulation of *allow*, demonstrating that an allow event represents two antecedents that are jointly necessary (rather than individually sufficient) for the effect. The mental model and force dynamics frameworks are unable to account for the observed findings.

General Discussion

Our experiments support causal model's representations of *cause*, *enable*, *allow* and *prevent*. Results from Experiment 1 support causal model's representation of *prevent*, and its mechanisms for causal inference. Results are also consistent with force dynamics theory. In contrast, mental model theory wrongly predicts that for double-prevent problems people will conclude *prevent*. Participants in Experiment 2, chose the *allow* label for antecedents that were part of a conjunctive cause, and not for antecedents that were by themselves sufficient causes. Neither force dynamics nor mental model theory make this prediction. Only causal model theory was able to account for results in both experiments.

Aside from the empirical advantage causal model theory shows, the relative ease with which predictions are derived, suggests theoretical advantages. Causal model theory can incorporate multiple causes, enablers, or preventions, which is something that other theories do not naturally do. In mental model theory, the cost of incorporating a new variable is an exponential increase in the number of models. In force dynamics, adding a variable requires computing the new configuration of forces (Sloman et al., 2008).

Finally, Experiment 2 shows that causal reasoning depends on prior knowledge and deliberative reasoning. The exploration of how these systems interact to support causal meaning and inference, remains for the future.

References

Barbey, A.K., & Sloman, S.A. (2007). Base-rate respect: From ecological rationality to dual processes. *The Behavioral and Brain Sciences, 30*(3), 241-254.

Barbey, A.K., & Wolff, P. (2006). Causal reasoning from forces. In *Proceedings of the 28th Annual Conference of the Cog. Science Society*. Mahwah, NJ: Erlbaum.

Barbey, A.K., & Wolff, P. (2007). Learning causal structure from reasoning. In *Proceedings of the 29th Annual Conference of the Cog. Science Society*. Hillsdale, NJ: Erlbaum.

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

Hasson, U., & Glucksberg, S. (2006). Does understanding negation entail affirmation? An examination of negated metaphors. *Journal of Pragmatics, 38*(7), 1015-1032.

Johnson-Laird, P.N. & Byrne, R.M.J. (2002). Conditionals: a theory of meaning, pragmatics, and inference. *Psychological Review, 109*(4), 646-678.

Kahneman, D. & Frederick, S. (2005). A model of heuristic judgment. In K.J. Holyoak & R.G. Morrison (Eds.) *The Cambridge Handbook of Thinking and Reasoning*. Cambridge University Press. 267-293.

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

Sloman, S., Barbey, A., & Hotalling, J. (2008). A causal model theory of the meaning of cause, enable, and prevent. *Cognitive Science*.

Sloman, S.A., & Hagmayer, Y. (2006). The causal psycho-logic of choice. *Trends in Cognitive Sciences, 10*(9), 407-412.

Spirtes, P., Glymour, C., & Scheines, R. (1993). *Causation, prediction, and search*. New York: Springer-Verlag.

Talmy, L. (1988). Force dynamics in language and cognition. *Cognitive Science,12*(1), 49-100.

Wolff, P. (2007). Representing Causation. *Journal of Experimental Psychology: General, 136*, 82-111.

Wolff, P., Song, G., & Driscoll, D. (2002). Models of causation and causal verbs. In M. Andronis, C. Ball, H. Elston and S. Neuval (Eds.), *Papers from the 37th Meeting of the Chicago Linguistics Society, Main Session, Vol. 1*. (pp. 607-622) Chicago: Chicago Linguistics Society.