correct account of actual causation; rather, we argue that standard methods will not lead to such an account. A different approach is required.

Keywords Actual causation · Bayesian networks · Combinatorics · Intervention · Intuitions

> Once upon a time a hungry wanderer came into a village. He filled an iron cauldron with water, built a fire under it, and dropped a stone into the water. 'I do like a tasty stone soup' he announced. Soon a villager added a cabbage to the pot, another added some salt and others added potatoes, onions, carrots, mushrooms, and so on, until there was a meal for all.

1 The theses

One philosophical goal is analysis: the provision of necessary and sufficient conditions for a concept, or for the possession or application of a concept. The Western historical source of the goal is Plato's discussion of the concept of 'virtue' in the *Meno*, but the *Meno* is also the source of a method: conjecture an analysis, seek intuitive counterexamples, reformulate the conjecture to cover the intuitive examples of the concept and to exclude the intuitive non-examples; repeat if necessary. Much of contemporary philosophy attempts the same strategy for many concepts: knowledge, belief, reference, causation, and so on. Addressing analyses of 'reference,' Mallon et al. (in press) argue that psychological investigation suggests that intuitions about reference are so varied that no uniform analysis can capture the discrepancies.

Our concern is about analyses of a scientifically and morally important notion, 'actual causation'—about proposed necessary and sufficient conditions for one event to cause another. For an inference to a general analysis from intuitions about cases to be credible, more than psychological consensus is required. The intuitive cases used to justify an analysis must somehow be representative of the possible cases of actual causation or its absence. What is particularly interesting about 'actual causation' is that the possible cases can in some sense be enumerated, and the enumeration can be used to show that consideration of intuitive examples is not representative, and apparently cannot be. Our argument first provides principles for enumerating the number of possible, structurally isomorphic examples of actual causal relations, without regard to the content of the related events. We show that even with very strong equivalence relations, and even considering only the number of events typical of examples in the philosophical literature, the number of possible cases is quite large. Second, we note that the number of equivalence classes grows exponentially as more events are considered. And, third, we show by example that as more events are added, novel kinds of ambiguous cases, or counterexamples to proposed analyses, emerge.

The question of when one event or circumstance causes another has been the subject of two recent collections of philosophical essays (Dowe and Noordhof 2004; Collins et al. 2004), of a lengthy chapter in a prize-winning book (Woodward 2003), of a

connected pair of articles amounting to a short book (Halpern and Pearl 2005a), as well as of several other recent articles (Giles 2005; Spohn 2005; Hiddleston 2005

Analyses of actual causation for deterministic cases have assumed that the relation obtains between *values*

graph, there are $2^4 \times 2^2 = 2^6$ possible truth functions. If we again treat the no-edge graph as just one case, then there are 601 causal models over the three potential causes. Since any causal model among the potential causes can be paired with any structure for the effect, there are 190,517 possible causal models altogether. And the number of cases (not structures) is much larger: each possible structure corresponds to 2^{C} cases, where C is number of exogenous (i.e., no parent) variables in that structure. (Until

Table 1 Numbers of truth functions

Number of parents	Number of truth functions	Number of truth functions with test pairs
1	4	2
2	16	10
3	256	218
4	65,536	64,594
5	> 4 " 10	

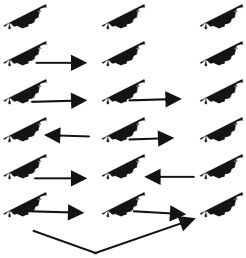
Table 2 Counting graphs with test pairs

Number of graphs of form…	× Number of test pair truth functions per graph	= Number of structures
1 disconnected graph	1	1
6 graphs of the form ->	2	12
6 graphs of the form -> ->	$2 \times 2 = 4$	24
3 graphs of the form <- ->	$2 \times 2 = 4$	12
3 graphs of the form -> <-	10	30
6 graphs of the form 	$2 \times 10 = 20$	120

cases (i.e., 6 possible structures); 10 permissible truth functions for the three two-edge cases (i.e., 30 possible structures); and 218 test pair truth functions for the single three-edge case. There are thus 255 possible structures (assuming the test pair condition) over the three potential causes and effect. Since every structure among the causes is consistent with every structure between the potential causes and the effect, we have $255 \times 199 = 50,745$ structures on three potential binary causes and one binary effect. The test pair restriction eliminates nearly 75% of the possible causal models, but that is not nearly reduction enough for intuition to survey the cases. Moreover, the combinatorics rapidly get much worse as the number of potential causes increases. The "simple" situation of five causes (i.e., all have $C \rightarrow E$) with *no* causal connections among them, and where we impose the test pair condition, corresponds to more than 4 billion possible structures.

2.2 Unlabeled graphs and other restrictions

We can additionally consider restrictions on the space of possible graphs. The idea with graphical models is that structure alone is considered, not the names given to variables or the substantive content of the events. In the absence of specific information about the meaning of variables, $X \rightarrow Y$ is structurally identical to $X \leftarrow Y$. If we group together directed acyclic graphs that are identical except for the variable names, then there are only six possible structures over three potential causes:



The middle column of Table 2 shows the number of test pair truth functions for each of these six graphs. The counts of structures involving the effect are more complicated if variable names do not matter. For example, if $X \rightarrow Y \leftarrow Z$ among the potential causes, then X, Y as the causes of E is equivalent to Z, Y being the causes of E; notice

Table 3 Counting unlabeled graphs with test pairs

	0 causes	1 cause	2 cause	3 cause	= Number of test pair truth functions for row structure
* * *	1	2	10	218	231
* → * *	1	3 × 2	3 × 10	218	255
* → * → *	1	3 × 2	3 × 10	218	255
* ← * → *	1	2 × 2	2 × 10	218	243
* → * → *	1	2 × 2	2 × 10	218	243
Three-edge	1	3 × 2	3 × 10	218	255

that X, Z being causes is not equivalent to the other two. Table 3 shows the number of test pair truth functions for E for all combinations of potential cause structure (rows) and number of causes of the effect (column). For cells with two numbers, the first number indicates the number of distinct graphical structures involving the effect when the potential causes are distinguishable only by their structural role (relative to the other potential causes).

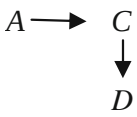
We can compute the total number of causal models by multiplying the right-most column of Table 3 by the relevant number of test pair truth functions over the potential causes, and then summing together. Ignoring variable names, combined with the test pair condition, results in 10,263 possible causal structures. Smaller, but still a busy time for intuitions.

We can impose further plausible restrictions on the space of possible graphs, though they could conflict with some theories of actual causation.[7] All of the various accounts of actual causation agree that $\mathsf{C} = \mathsf{c}$ cannot be an actual cause of $\mathsf{E} = \mathsf{e}$ if there is no directed path from C to E. Moreover, if there is a directed path from C to E, and there is no directed path from B to E, then whether or not $\mathsf{C} = \mathsf{c}$ is an actual cause of $\mathsf{E} = \mathsf{e}$ cannot depend on whether or not $\mathsf{B} = \mathsf{b}$. Various models are thus dispensable or equivalent with respect to testing an account of actual causation. For example, suppose E is a function of a single variable and * → * → * holds among the potential causal variables. The only distinct structure is the one in which E depends on the terminal star. If E depends on the middle variable, then it is equivalent to $* \rightarrow * \cdots *$ over the potential causes, since the last variable cannot be an actual cause, and cannot affect whether the other two variables are actual causes (by the above principles involving directed paths). If E depends on the first variable, then it replicates a case counted among those with $* \cdots * \cdots *$ as the relevant substructure on the causal variables. This restriction results in 20 distinct graphical structures over the three potential causes and E, distributed as shown in the central cells of Table 4. The relevant number of test pairs for the structures among the potential causes (rows) and involving the effect (columns) are also shown in Table 4.

[7] For example, that

Table 6 Example of truth

(3) A and B each fire a bullet that would have missed the target, except that the bullets collide (C = 1) and A's bullet ricochets through the bullseye. What caused the bullseye to be hit (D = 1)?

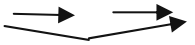


$C = A \bullet B; \quad D = C; \quad A = B = C = D = 1.$

W, HP2005: The actual causes of D = 1 are A = 1, B = 1, and C = 1.

(4) A, a perfect marksman, is about to fire at the bullseye; B is about to jostle A to prevent A from hitting the bullseye; C shoves B out of the way.A fires and hits the bullseye (D). What caused the bullseye to be hit?

$C \rightarrow B \rightarrow A \rightarrow D; \quad D = A; \quad A = (1 - B); \quad B = (12.2881\ 553.715\ \mathrm{l}\ 147.668\ 55$



$P = (1 \% B); \quad T = B + P.B = T = 1, \quad P = 0$

W, HP2005: The actual cause of $T = 1$ is $B = 1$

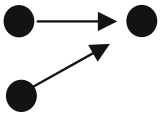
(14) A and B

others will vote: his priors for every vote but his are 50/50 for round-up. No matter how W votes, cases in which C and R

description of circumstances. One would like to know whether judgments of actual causation depend only on the Þnal state or on the transitions that lead to it. One would like to know in what respects systems are sometimes too complex for people to give more than random judgments, or none at all. And many other questions remain unanswered.

There is an enormous psychological literature on human judgment about causation when the joint occurrences of features are repeated (i.e., about type-level causation), and about token causation for extremely simple ÒmechanicalÓ cases (e.g., collisions of objects, inspired by Michotte 1954), but relatively little about actual causation in other contexts. A study by Sloman and Lagnado (2002

indicators for discovering actual causal relations. Those indicators might provide a definition of actual causation, but they need not. The justification of a Euclidean



where a dark node indicates that a local change of state occurred. But the change (or happening) graph representation has no clear functional dependencies that are independent of the actual beginning and end states—no laws—and fails to mark the difference between a change in a node from empty to dark, and a change of that same node from dark to empty; each kind of change becomes a dark node. The same "change graph" would also represent this transition:

7 Conclusion

Causal Bayes nets developed as a formalism for representing causal relations among variables and for studying inferences to such relations and their use in predicting the effects of interventions. That framework is now used more or less without comment in several areas of science. It was natural enough then to take Bayes nets as a framework for actual causation, but it is a mistake to take actual causation generally to be isomorphic to a relation among values of nodes in such a structure, just as it is a mistake to induce vast generalizations about conditions for causal attribution from a baker's dozen of examples.

Our argument is not for an abandonment of formal representations of actual causation, or for promulgating more examples without formal control. We are not arguing for abandoning neuron diagrams or Bayes nets or graphical causal models in philosophical investigations of causal relations. We are not arguing against the possibility of a correct theory of actual causation. It is instead an argument (i) against the adequacy of the unsystematic Socratic strategy that has dominated philosophical discussion of actual causation; (ii) against the sufficiency of Bayes net representations for actual causation without consideration of state transitions; and (iii) against the presumption that, in judging cases, philosophers know best.

References

Ahn, W., & Kalish, C. W. (2000). The role of mechanism beliefs in causal reasoning. In F. C. Keil & R. A. Wilson (Eds.), *Explanation and cognition* (pp. 199–225). Cambridge, MA: The MIT Press.

Ahn, W.-K., Kalish, C. W., Medin, D. L., & Gelman, S. A. (1995). The role of covariation versus mechanism information in causal attribution. *Cognition, 54*, 299–352. doi:10.1016/0010-0277(94)00640-7.

Cheng, P. (1997). From covariation to causation: A causal power theory. *Psychological Review, 104*, 367–405.

Choi, I., Nisbett, R. E., & Norenzayan, A. (1999). Causal attribution across cultures: Variation and universality. *Psychological Review, 125*, 47–63.

Collins, J., Hall, N., & Paul, L. (Eds.). (2004). *Causation and counterfactuals.*

Hiddleston, E. (2005). Causal powers. *The British Journal for the Philosophy of Science, 56*, 27–59. doi:10.1093/phisci/axi102.
Hitchcock, C. (2001). The intransitivity of causation revealed in equations and graphs. *The Journal of Philosophy, 98*, 273–299. doi:10.2307/2678432.
Kvart, I. (2004a). Probabilistic cause, edge conditions, late preemption and discrete cases. In P. Dowe & P. Noordhof (Eds.), *Cause and chance: Causation in an indeterministic world*. New York: Routledge.
Kvart, I. (2004b). Causation: Probabilistic and counterfactual analyses. In J. Collins, N. Hall & L. Paul (Eds.), *Causation and counterfactuals*. Cambridge, MA: MIT Press.
Lewis, D. (1986). Causation. In *Philosophical Papers*, Vol. II, New York: Oxford University Press.
Mallon, R., Machery, E., Nichols, S., & Stitch, S. (in press). Against arguments from reference. *Philosophy and Phenomenological Research*.
McKenzie, C. R. M., & Nelson, J. D. (2003). What a speaker's choice of frame reveals: Reference points, frame selection, and framing effects. *Psychonomic Bulletin & Review, 10*, 596–602.
Menzies, P. (2004). Difference making in context. In J. Collins, N. Hall & L. Paul (Eds.), *Causation and counterfactuals*. Cambridge, MA: MIT Press.
Michotte, A. (1954). *La perception de la causalité*. Louvain: Publications Universitaires de Louvain.
Noordhof, P. (2004). Prospects for a counterfactual theory of causation. In P. Dowe & P. Noordhof (Eds.), *Cause and chance: Causation in an indeterministic world*. New York: Routledge.
Novick, L. R., & Cheng, P. W. (2004). Assessing interactive causal influence. *Psychological Review, 111*, 455–485. doi:10.1037/0033-295X.111.2.455.
Nute, D. (1976). David Lewis and the analysis of counterfactuals. *Nous (Detroit, Mich.), 10*, 455–461. doi:10.2307/2214616.
Pearl, J. (2000). *Causality*. New York: Oxford University Press.
Ramachandran, M. (2004a). Indeterministic causation and varieties of chance raising. In P. Dowe & P. Noordhof (Eds.), *Cause and chance: Causation in an indeterministic world*. New York: Routledge.
Ramachandran, M. (2004b). A counterfactual analysis of indeterministic causation. In J. Collins, N. Hall & L. Paul (Eds.), *Causation and counterfactuals*. Cambridge, MA: MIT Press.
Sher, S., & McKenzie, C. R. M. (2006). Information leakage from logically equivalent frames. *Cognition, 101*, 467–494. doi:10.1016/j.cognition.2005.11.001.
Sloman, S. A., & Lagnado, D. (2002). Counterfactual undoing in deterministic causal reasoning. Proceedings of the twenty-fourth annual conference of the cognitive science society, Maryland.
Spirtes, P., Glymour, C., & Scheines, R. (1993). *Causation, prediction and search*. New York: Springer.
Spohn, W. (2005). Causation: An alternative. *The British Journal for the Philosophy of Science, 57*, 93–119. doi:10.1093/bjps/axi151.
Walsh, C. R., & Sloman, S. A. (2005). The meaning of cause and prevent: The role of causal mechanism. In B. G. Bara, L. Barsalou & M. Bucciarelli (Eds.), *Proceedings of the 27th annual conference of the cognitive science society* (pp. 2331–2336). Mahwah, NJ: Lawrence Erlbaum Associates.
Wolff, P., & Song, G. (2003). Models of causation and the semantics of causal verbs. *Cognitive Psychology, 47*, 276–332. doi:10.1016/S0010-0285(03)00036-7.
Woodward, J. (2003). *Making things happen: A theory of causal explanation.* Oxford: Oxford University Press.