Simulating Explanatory Coexistence: Integrated, Synthetic, and Target-Dependent Reasoning

Scott E. Friedman (friedman@sift.net)
SIFT, 319 N 1st Ave.
Minneapolis, MN 55401 USA

Micah B. Goldwater (micah.goldwater@sydney.edu.au)
The University of Sydney, School of Psychology
Brennan MacCallum Building (A18)
NSW 2006 Australia

Abstract

Understanding the cognitive structure of explanations— and the cognitive processes that assemble them— is a milestone for understanding how people learn and communicate. Recent research on *explanatory coexistence* suggests that people's causal beliefs are less globally coherent than previously thought: people use seemingly-competing supernatural and biological causes to explain different aspects of the same phenomenon, or they assemble supernatural and biological causes into single, coherent explanations (Legare & Gelman, 2008; Legare & Shtulman, 2018; Shtulman & Lombrozo, 2016). This coexistence— and unexpected coherence— of diverse causal mechanisms poses interesting questions about the role of coherence and fragmentation in people's mental models and explanations. This paper presents a computational model of explanatory coherence in the well-characterized domain of disease transmission, extending a previous cognitive model of explanation-based conceptual change (Friedman, Forbus, & Sherin, 2018). Our approach (1) retrieves diverse causal model fragments based on the phenomenon to explain, (2) assembles coherent causal models using relevance-directed abductive reasoning, and (3) selects explanatory paths that support within-explanation and within-scenario coherence. Our model simulates the three different types of explanatory coexistence detailed in the literature.

Keywords: cognitive modeling; explanatory coexistence; AI; abductive reasoning; explanation

Introduction

The cognitive process of explanation has been a central focus of cognitive science since its inception, and it has broad implications for communication, instruction, and conceptual change (Chi, De Leeuw, Chiu, & LaVancher, 1994; Vosniadou, 1994; diSessa & Sherin, 1998; Shtulman & Lombrozo, 2016; Friedman et al., 2018). The more recent focus on *explanatory coexistence*, whereby people utilize diverse— and seemingly incompatible— causal mechanisms in their explanations (Legare & Shtulman, 2018), poses additional questions about how people construct and consider explanations, how explanations are structured, and how explanations cohere with other beliefs.

This paper presents a computational cognitive model of explanation, building on previous cognitive models of conceptual change (Friedman et al., 2018). We apply our cognitive model to simulate human subjects' explanatory coexistence in the domain of disease, as characterized by Legare and Gelman (2008) and later by Legare and Shtulman (2018).

Our cognitive model assembles situation-specific causal models from smaller, generic *model fragments* (i.e., causal knowledge units). Given a new situation to explain, the model explains the situation by:

1. Retrieving causal model fragments based on the situation.
2. Traversing backwards recursively, instantiating model fragments within the situation in an relevance-directed beam search, assuming entities and relations as necessary.
3. Identifying the causal path(s) that maximize an objective coherence function with respect to global assumptions, coverage over the situation, and *presupposition* beliefs.

This model assumes that intuitive and culturally-acquired knowledge coexists, and that the process of assembling explanations is biased principally by coherence. This means that scientific and supernatural causal mechanisms can coexist in the same explanation, e.g., so that supernatural events might cause a biological event that leads to a viral infection, assuming the causal knowledge is primed and applicable.

Our simulation results demonstrate that that our model (1) simulates the three categories of explanatory coexistence in the literature and (2) varies its choice of explanation according to priming in a manner similar to human subjects.

We continue with an overview of explanatory coexistence and computational methods used in our cognitive model. We then describe our approach, present our simulation results, and conclude with a discussion of our results, key psychological assumptions, and directions for future work.

Background

We describe psychology research on explanatory coexistence, and then we review computational modeling techniques relevant to our simulation.

Explanatory Coexistence

There are scientific and religious or supernatural explanations for the same natural phenomena (e.g., creation of the universe, death, disease transmission). It is intuitive that learning scientific explanations for natural phenomena would replace previously learned supernatural explanations; however, evidence over the past decade suggests the opposite: scientific explanations replacing supernatural explanations is the

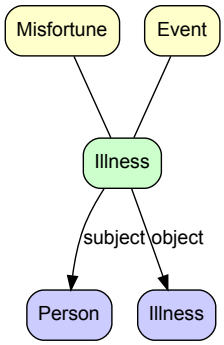

Figure 1: Model fragment for *Illness*.

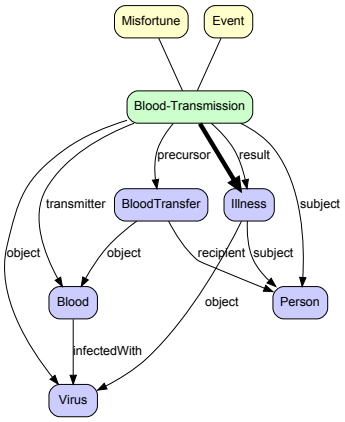

Figure 2: Model fragment for *Blood-Transmission* of disease via *BloodTransfer*.

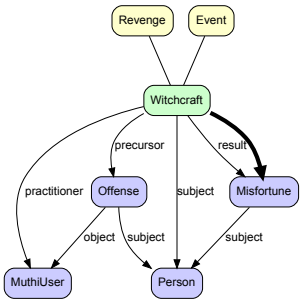

Figure 3: Model fragment for *Witchcraft* causing misfortune after an offense.

exceptional case. More frequently, people utilize both explanations (Shtulman & Lombrozo, 2016).

Legare and Gelman (2008) examined the specific case of explaining HIV transmission in South Africa. Before educational interventions focusing on the biological transmission of HIV, AIDS symptoms were explained as the result of witchcraft. Legare and Gelman (2008) showed that the educational interventions did not replace the bewitchment explanations; instead, both biological and bewitchment explanations coexist. For example, a man may have contracted HIV from sexual intercourse, but was attracted to a woman with HIV because of witchcraft.

Legare and Shtulman (2018) acknowledge the following categories of explanatory coexistence, all of which we simulate in this work:

1. **Integrated** reasoning combines seemingly-incompatible causal mechanisms into a coherent causal structure. For instance, bewitchment could cause somebody to choose a sexual partner who has AIDS, and intercourse with that partner causes disease transmission.
2. **Synthetic** reasoning invokes multiple causal mechanisms without articulating hierarchical or temporal precedence to any, possibly due to competing explanations.
3. **Target-dependent** reasoning applies different mechanisms to distinct aspects of a situation, in a highly-contextualized fashion. The various mechanisms do not participate in the *same* explanation.

Compositional Modeling

Simulating people's causal mental models requires expressive knowledge representation and reasoning (KR&R). An approach using only atomic logical propositions is not expressive enough to suit the mental model literature (Vosniadou, 1994; Chi et al., 1994; diSessa & Sherin, 1998; Gentner & Stevens, 1983) or the analogy literature (Friedman, Barbella, & Forbus, 2012), and an approach using only neural networks does not support sufficient interpretability.

Previous KR&R research on *compositional modeling* (Falkenhainer & Forbus, 1991) provides (1) representations for modeling the structure and continuous processes of dynamic systems, and (2) algorithms for composing these models on-the-fly for novel situations. Structure-behavior-function models (Goel, Rugaber, & Vattam, 2009) expand on this formalism to capture teleology, and have been used to simulate people's mental models.

Following recent cognitive modeling work (see the **Assembled Coherence** subsection), we simulate people's mental models using compositional modeling semantics extended with more expressive event structure (Pustejovsky, 2013). We represent each causal mechanism as a generic *model fragment* that can compose with others into large situation-specific explanations. Each model fragment describes:

- **Categories** that it instantiates, from general (e.g., *Misfortune*) to specific (e.g., *Sexual Transmission* [of a virus]).
- **Participants** are the entities or events that interact within the described mechanism. Each participant has one or more categories of its own. Model fragments with the same binding of participants are semantically equivalent.
- **Constraints** are *existence* conditions specified over the participants. If the constraints hold over participants in a situation, the model fragment may be *instantiated*.
- **Consequences** are functional or behavioral representations specified over the participants. They are asserted into the situation when the model fragment is instantiated.[1]

We include diagrams of three of the ten model fragments used in this simulation domain: Figure 1 shows a simple fragment describing an *Illness* state: a *subject* participant of type *Person*; a *object* participant of type *Disease*; and super-categories of *Event* and *Misfortune*.

Figure 2 shows the more complex fragment *Blood Transmission* [of disease], including sub-events of *Blood Transfer*

[1] A consequence may have *conditions* that must hold in the situation for them to be asserted, but in this paper each conclusion is a causal relationship without conditions.

and *Illness*, as well as a conclusion stating that, if instantiated, the *Blood Transmission* is a cause of the *Illness*. Importantly, this model fragment constrains its own participants: its subject (the *Person* at lower-right) is the subject of the *Illness* and the *Blood Transfer*. These constraints are required for the fragment— and its causal structure— to be realized.

Finally, Figure 3 illustrates a simple *Witchcraft* fragment, with a conclusion stating that an instantiated *Witchcraft* can cause a *Misfortune* of its subject. Per Figures 1 and 2, both *Illness* and *Blood Transmission* are types of *Misfortune*, so they can be directly caused by *Witchcraft*. This allows assembly of larger causal models, provided the situation (or explicit assumptions) satisfies the fragments' constraints. We describe assumptions below, and their affect on explanation quality.

Scalability of compositional modeling is a key consideration as the number of fragments grows, since the deductive closure of possible models can grow geometrically. We later describe a relevance-based heuristic in our approach that jointly (1) reduces the compositional modeling search space drastically and (2) helps model priming effects.

Abductive Reasoning

Abductive reasoning generates multiple explanations for observations— potentially generating assumptions along the way— and then selects the *"best"* explanation and its constituent assumptions as inferences or rationale for the observations. Previous computational approaches have modeled explanation quality as numerical cost (Charniak & Shimony, 1994) or as likelihood maximization with Bayesian approaches (Raghavan & Mooney, 2010). Our approach uses cost-based abductive reasoning to select explanations built from model fragments, and could be extended to use Bayesian approaches if we had estimates of subjects' beliefs of prior probability distributions. To improve scalability over previous abduction approaches— since the search space can grow geometrically (Poole, 1993)— our approach uses a relevance-based heuristic to guide its search for explanations.

Assembled Coherence

This paper extends recent work on the *assembled coherence (AC) theory* (Friedman et al., 2018) of mental models and conceptual change.

AC theory proposes that fragmented knowledge is assembled into larger, coherent mental models through the process of *abductive reasoning* (i.e., reasoning to the best explanation). Once assembled, these mental models are evaluated against a network of *presupposition* beliefs and then reused in novel situations by partial reformulation or by analogy (Friedman et al., 2012). This incorporates ideas from both the knowledge-in-pieces (diSessa & Sherin, 1998) and framework theory (Vosniadou, 1994) perspectives of mental models, and postulates that the two perspectives are compatible and complementary.

AC theory has been implemented in computational cognitive models to simulate explanation-based conceptual change in the domains of force dynamics (Friedman & Forbus, 2010), the day-night cycle (Friedman et al., 2012), the human circulatory system (Friedman & Forbus, 2011), and seasonal change (Friedman et al., 2018).

Approach

Our computational model generates a causal explanation by (1) retrieving model fragments based on the scenario to explain, (2) instantiating causal model fragments in an effect-to-cause beam search prioritized by relevance, (3) scoring coherent explanatory paths for coherence, and (4) selecting the most optimal explanatory path. We describe each of these processes below.

Retrieving causal knowledge. Given a new situation to explain, the system retrieves its model fragments (i.e., causal mechanisms) based on the categorical and relational overlap of the situation with those of its model fragments.

Specifically, given a situation s and a model fragment m, we compute relevance $Rel(m,s)$ with respect to the model fragment's participant categories C_m and constraint relations R_m and the situation's categories C_s and relations R_s. We use a simple Jaccard distance as a relevance function:

$$Rel(m,s) = \frac{|C_m \cap C_s| + |R_m \cap R_s|}{|C_m \cup C_s| + |R_m \cup R_s|} \quad (1)$$

This relevance function is a very coarse estimate of a model fragment's applicability to a situation, and we use it for simplicity: a model fragment's relevance strictly increases with situation-shared categories (e.g., *Person*, *Blood*, *SexualIntercourse*) and relations (e.g., *infectedWith*, *knows*, *motherOf*), and its relevance decreases monotonically relative to its total number of categories and relations. This approach is similar to performing spreading activation (Crestani, 1997) from categories and relations to relevant model fragments but allows indexing for scalability.

We discuss other plausible retrieval and salience factors in the conclusion of this paper.

Relevance-directed beam search. Given its relevance over causal model components, the system performs an incremental backward search through the space of possible causal models. This process is given an *explanandum* (i.e., event or assertion to explain), such as the illness of an individual, and then performs the following recursive operations for its explanation queue.

For each item x in its queue, it finds *applicable* model fragments that have x's type as a habitat consequence, e.g., if x is an *Illness*, then *BloodTransmission* (Figure 2) and *Witchcraft* (Figure 3) both apply. It selects applicable model fragments within the top 10% relevance window and attempts to compose the retrieved model fragment(s), constraining them by binding x as the consequent participant. The composition algorithm may assume any participants necessary to compose at least one instance, provided it obeys the input binding(s). The system then adds these new instances (e.g., *BloodTransmission* or *Witchcraft*) to the queue and will focus on those next, repeating.

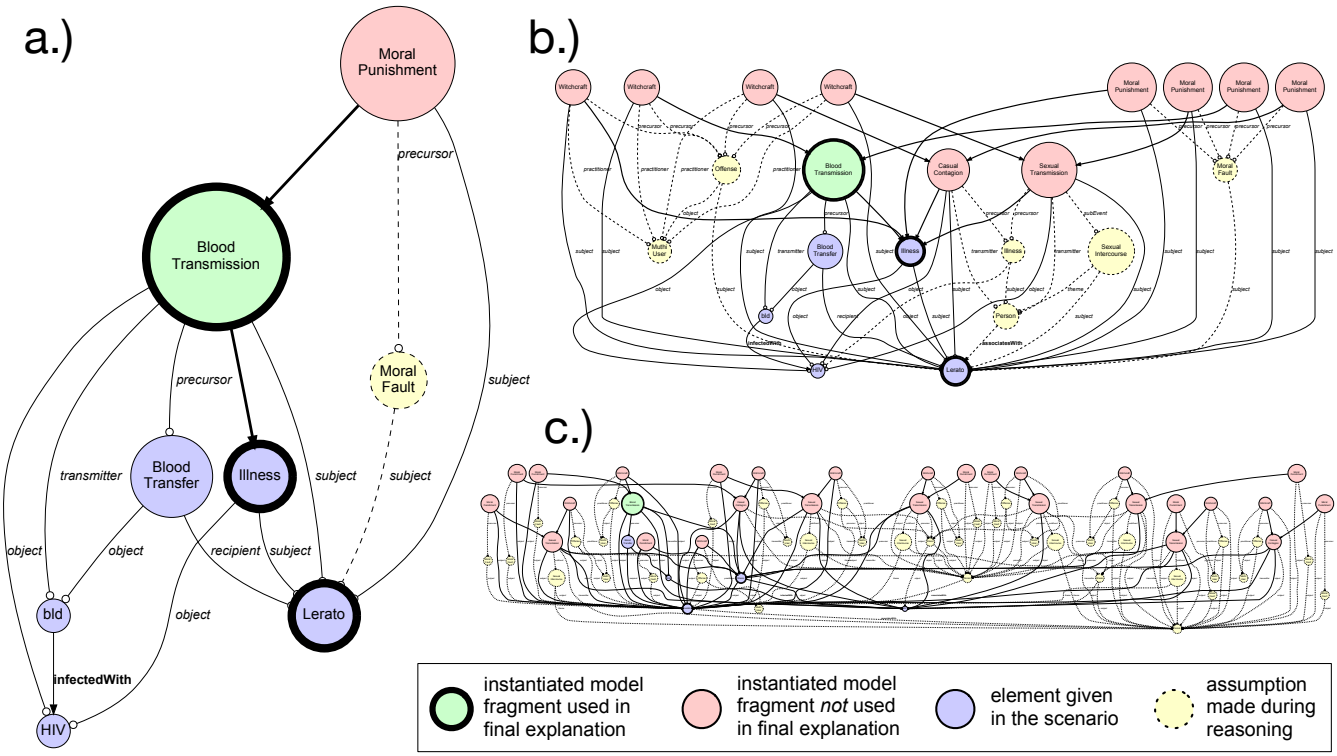

Figure 4: Explanations generated and selected for the same prompt of Lerato's AIDS after a blood transfer: (a) with relevance-directed causal search; (b) with undirected causal search; and (c) with exhaustive forward-chaining. All approaches utilize the same causal knowledge and result in the same final explanation, but with orders of magnitude difference in computation.

Explanation structure. Relevance-directed beam search produces a network of model fragment instances, as illustrated in Figure 4(a). In Figure 4(a), we provided the system the situation plotted in blue: Lerato has HIV and was the recipient of a blood transfer. Lerato and her *Illness* instance are outlined in bold for clarity. As with the human subjects of Legare and Gelman (2008), Lerato's illness is the explanandum in every simulation in this paper, but we vary the details of the situation to model priming effects.

The green (e.g., *Blood Transmission*) and red (e.g., *Moral Punishment*) elements in Figure 4(a) are instances of model fragments that were retrieved and assembled by this algorithm. The difference is that the green instances were chosen as part of the *best explanation* (described below), and the red instances were assembled and considered by the system, but were not ultimately included in the best explanation.

The yellow elements (e.g., *Moral Fault*) were *assumed* during the course of instantiation in order to satisfy model fragment participants and constraints.

In summary, Figure 4(a) shows that the system assembled a *Blood Transmission* event as a cause of Lerato's HIV, given that Lerato was the recipient of a blood transfer that was infected with HIV. It explained the *Blood Transmission* with a possible *Moral Punishment*, and assumed that Lerato committed some *Moral Fault* in the course of instantiating the *Moral Punishment*. The *Moral Punishment* (in red) was not included in the best explanation due to the additional assumption, since this reduces the coherence score (described below). All of our simulation results use this color-coding.

For reference, we contrast the Figure 4(a) explanation structure resulting from relevance-directed beam search with two other (less efficient) explanation-assembly algorithms to characterize the strength of our approach:

- Figure 4(b) illustrates the same situation and explanandum (i.e., blue nodes) using a backward search *without* relevance as a heuristic: it regresses from effects to causes, but tries *all* causes rather than those primed by the situation.
- Figure 4(c) illustrates the same situation and explanandum using exhaustive forward search. This instantiates all applicable events and then repeats.

Neither of these graphs' structure are legible, but we include them to visualize the difference in computation across approaches. Both of these alternative approaches select the exact same final explanation (in green) as the more efficient relevance-directed beam search in Figure 4(a). This suggests that relevance from the situation is a useful heuristic for approximating coherence while assembling explanations in a large space of possible explanatory paths.

These plots also demonstrate that our qualitative models are capable of expressing a wide range of explanations, many of which are incoherent and not employed by people.

This means we have not trivially *"baked in"* the explanation within the knowledge representation; rather, it is the product of assembly and assumption (described above) and coherence assessment, which we describe next.

Scoring explanations for coherence. After assembling explanation structure from model fragments— and potentially making assumptions in the process— the system traverses the explanation structure to select a *best explanation*. This is the culmination of *abductive reasoning* (i.e., inference to the best explanation), which has been formulated as likelihood maximization (Raghavan & Mooney, 2010), simplicity, and other measures of explanation quality (Lombrozo, 2007).

Our system scores explanations by (1) identifying connected causal subgraphs of at least one cause (i.e., of Lerato's *Illness*), (2) scoring those subgraphs for coherence, where larger scores indicate greater coherence, and (3) selecting the highest-scoring subgraph as the best explanation.

The coherence score is the sum of epistemic *features* of a causal graph, where features positively or negatively contribute to coherence. Each feature is scored once for each causal graph, so many model fragment instances can rely on one assumption and incur the cost once. We employ a simple order-of-magnitude scoring technique over these features:

- *Model Fragments* (-1) penalize for increasing complexity.
- *Assumptions* (-10) penalize for increasing complexity.
- *Situation premises* (10) are situation events and entities that participate in model fragments, increasing explanatory inclusion (i.e., coherence) over the stated situation.
- *Causal associations* (100) are presuppositions that associate categories of causes and effects, e.g., *witchcraft* causally contributes to *illness*.
- *Causal dissociations* (-100) are presuppositions that dissociate categories of causes and effects, e.g., *witchcraft* does not cause *physical effects*.

These features coarsely quantify coherence: within-explanation coherence, explanation-to-situation coherence, and explanation-to-presupposition coherence. Following Vosniadou (1994), we model presuppositions as overarching belief-level constraints on people's explanations acquired culturally or via observation. We do not believe our list of is complete, since factors like analogical structure, narrative structure, likelihood, and other factors all contribute to people's explanatory preferences (Lombrozo, 2007).

Simulation

Our simulation setup is a variation of a human experiment by Legare and Gelman (2008): as exemplified in the previous section, we prompt the system to explain how Lerato contracted HIV. We use priming conditions from their study— *biological* priming, *bewitchment* priming, *neither* priming, and *both* types of priming— by varying the information we provide about Lerato. We provide two alternative types of biological priming: sexual intercourse and blood transfer. We also provide a "moral" priming condition, since other results from Legare and Gelman (2008) suggests that some subjects believe immoral behavior can cause illness.

Legare and Gelman (2008) report that 60% to 70% of their subjects exhibited some case of explanatory coexistence, where both supernatural and biological mechanisms (a) explained aspects of the scenario (*target-dependent*); (b) were juxtaposed (*synthetic*); or (c) coexisted in a causal chain (*integrated*). Subjects were sensitive to priming effects: biological and bewitchment priming was associated with more of those mechanisms appearing in explanations. We next review our simulation results for seven priming conditions, shown in the Figure 5 explanation graphs.

Target-dependent reasoning. Graphs (a-e) are all evidence of target-dependent reasoning. Graph (a) is *no priming*, where the system assumes immoral behavior as a simple cause for the disease. Graph (b) is immoral priming, which removes the need for the assumption of immorality. Graph (c) is bewitchment priming, mentioning a practitioner who knows Lerato, which results in assuming an offense, and also considering *Moral Punishment*, but ultimately choosing *Witchcraft* as an explanation. Graph (d) is biological priming with mention of receiving infected blood, resulting in a *Blood Transmission* explanation. Graph (e) is biological priming with mention of sexual intercourse with an HIV-infected partner, resulting in a *Sexual Transmission* explanation, but considering that the illness or the sexual transmission might have been caused by immoral behavior.

Synthetic reasoning. Graph (g) demonstrates one possible example of synthetic reasoning, where a presupposition causally associates *Witchcraft* with *Illness*, and we prime both biological and witchcraft causes. In this case, the *Sexual Transmission* fragment coheres with the situation (i.e., it requires no assumptions), and the *Witchcraft* fragment coheres with the presupposition (rendered in black), so the union of those causes of the illness is higher-scoring than either alone. Our system has no hard constraint to select single causes at causal junctions; however, selecting two causes— when either alone is sufficient— is counter-intuitive. The "synthetic reasoning" category of explanatory coexistence is not as well-specified as the other two, and could plausibly represent multiple sub-strategies, e.g., where subjects integrate causes in parallel, mention multiple salient or competing causes, or are vaguely verbalizing a more integrated causal chain (as below). This suggests further research with human subjects.

Integrated reasoning. Graphs (f) and (h) are evidence of integrated reasoning. Graph (f) is priming of both bewitchment and biology, resulting in an integrated explanation: witchcraft caused the sexual transmission of HIV during the sexual encounter. Graph (h) is priming of moral and biology, resulting in another integrated explanation: immoral behavior caused transmission of HIV during a transfer of blood. Legare and Gelman (2008) did not explicitly attempt the priming condition in Graph (h), but our model predicts that an integrated explanation is plausible for these mechanisms.

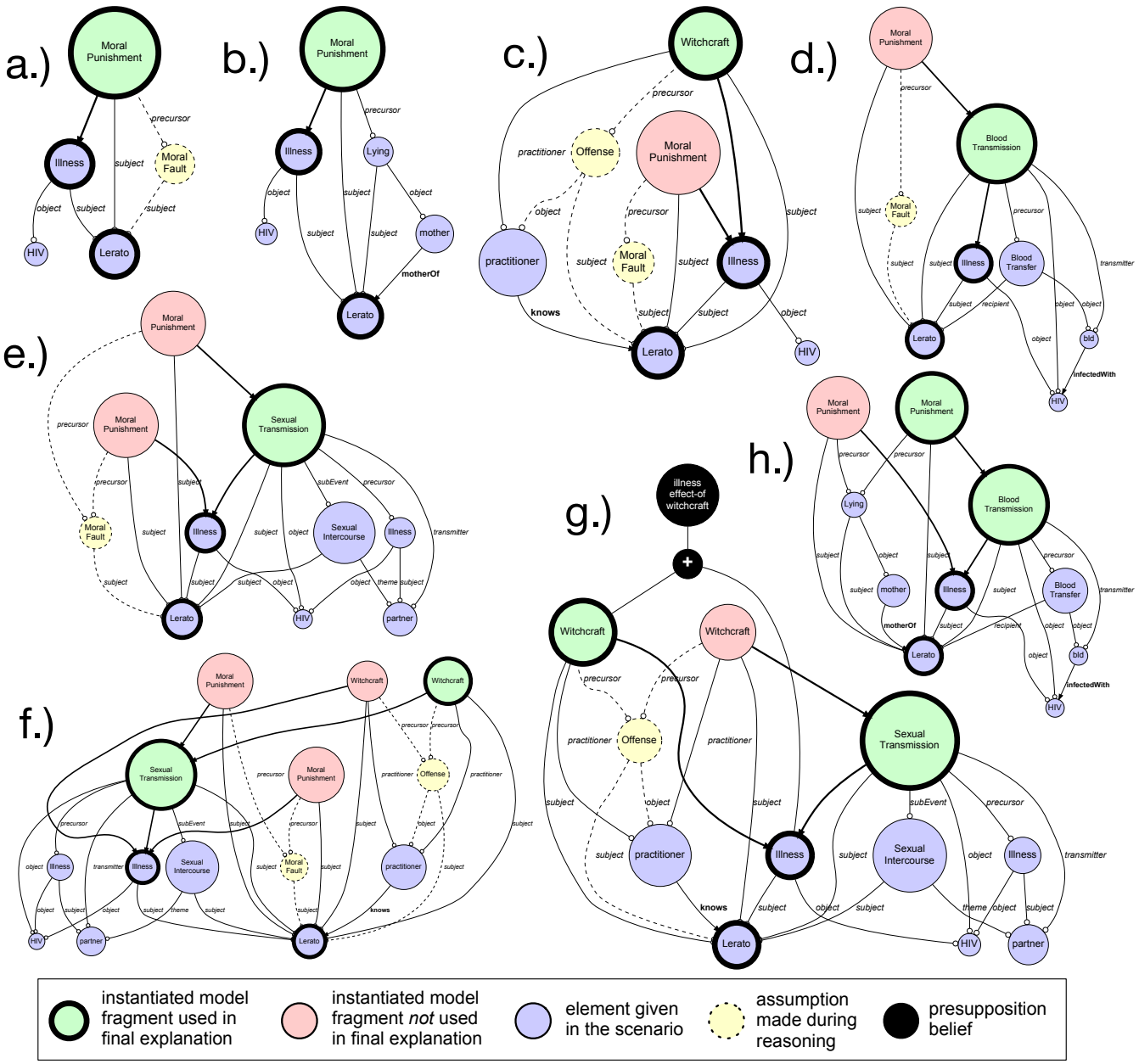

Figure 5: Simulation results with identical causal knowledge, by varying priming: (a) no priming; (b) priming immoral behavior; (c) priming with witchcraft practitioner; (d) priming with blood transfer; (e) priming with sexual encounters; (f) priming with *both* sexual encounters and witchcraft; (g) same priming but including a presupposition that illness is caused by witchcraft; and (h) priming with *both* immorality and blood transfer.

Conclusion

This paper presents a computational cognitive model that simulates all three categories of explanatory coexistence (Legare & Shtulman, 2018; Legare & Gelman, 2008), using the same psychological assumptions as previous models of conceptual change and self-explanation (Friedman et al., 2018). Our computational model retrieves diverse causal model fragments based on relevance to the scenario, and then assembles and evaluates explanations that may integrate both biological and supernatural causes.

We simulated different explanatory coexistence outcomes by varying high-level presuppositions and priming effects; we did *not* vary any causal models, likelihood values, or retrieval parameters across trials. The simulations demonstrate that the model's explanation-assembly is sensitive to priming effects, similar to people (Legare & Gelman, 2008). We showed that salient high-level beliefs— which have been termed *presuppositions* (Vosniadou, 1994)— bias the system to prefer explanations that cohere with their constraints.

Psychological claims and assumptions. This model and simulation support the claims that (1) explanatory coexistence may be the rule rather than the exception (Legare & Shtulman, 2018) and (2) explanatory coherence is a secondary property of assembling and assessing fragmentary, reusable causal knowledge (Friedman et al., 2018). These claims must be framed within the assumptions and limitations of our computational cognitive model.

Our model does not explicitly represent prior probabilities or joint probabilities across causal mechanisms. On the one hand, this allows it to flexibly assemble human-like causal explanations with diverse, seemingly-conflicting mechanisms; however, it could produce uncharacteristic explanations in other domains. This is an empirical question we will investigate in future work, described below. Although we did not encode likelihoods in this work, our model is compatible with Bayesian and statistical relational learning: its situation-specific explanation structure supports statistical inference (Raghavan & Mooney, 2010), and its coherence score could inform likelihood judgments in absence of prior probabilities.

Our model simplifies the psychological processes of knowledge activation and explanation assessment. Some activation and assessment factors *not* modeled here include structural similarity to previous situations, prior likelihood estimates for any given causal mechanism, and probability distributions over causal mechanisms conditionalized on the situation. Implementing these factors would increase the power of our model, but at the expense of interpretability: these factors make additional assumptions about the belief state of each simulated subject, such as their episodic knowledge and the likelihood they ascribe to each causal mechanism.

Future work. In addition to the domain of disease, people's explanatory coexistence has been characterized in the domains of death and human origins (Legare & Shtulman, 2018). Simulating these domains will provide additional empirical evidence of our model's generality.

In addition to other domains, running this model of explanation on other explanation tasks will help qualify its broader psychological plausibility. Also, applying this model of explanation within larger models of explanation-based learning and conceptual change will help us refine the model's parameters knowledge representations.

Finally, this paper's simulations utilized a purely qualitative comparison between human and machine explanations as a proof of concept, but we plan to model quantitative properties, such as subjects' reaction time, in future work.

Acknowledgments

We acknowledge CogSci reviewers for their insightful reviews, and we thank Cristine Legare and Andrew Shtulman for helpful discussions on the direction of this work.

References

Charniak, E., & Shimony, S. (1994). Cost-based abduction and MAP explanation. *Artificial Intelligence*, *66*, 345–374.

Chi, M. T., De Leeuw, N., Chiu, M.-H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive science*, *18*(3), 439–477.

Crestani, F. (1997). Application of spreading activation techniques in information retrieval. *Artificial Intelligence Review*, *11*(6), 453–482.

diSessa, A. A., & Sherin, B. L. (1998). What changes in conceptual change? *International journal of science education*, *20*(10), 1155–1191.

Falkenhainer, B., & Forbus, K. D. (1991). Compositional modeling: finding the right model for the job. *Artificial Intelligence*, *51*(1-3), 95–143.

Friedman, S. E., Barbella, D., & Forbus, K. (2012). Revising domain knowledge with cross-domain analogy. *Advances in Cognitive Systems*, *2*, 13–24.

Friedman, S. E., Forbus, K., & Sherin, B. (2018). Representing, running, and revising mental models: A computational model. *Cognitive Science*, *42*(4), 1110–1145.

Friedman, S. E., & Forbus, K. D. (2010). An integrated systems approach to explanation-based conceptual change. In *Proceedings of AAAI 2010* (pp. 1523–1529).

Friedman, S. E., & Forbus, K. D. (2011). Repairing incorrect knowledge with model formulation and metareasoning. In *Proceedings of ijcai 2011* (Vol. 22, pp. 887–893).

Gentner, D., & Stevens, A. L. (1983). *Mental models*. Psychology Press.

Goel, A., Rugaber, S., & Vattam, S. (2009). Structure, behavior, & function of complex systems: The structure, behavior, & function modeling language. *AI EDAM*, *23*, 23–35.

Legare, C. H., & Gelman, S. A. (2008). Bewitchment, biology, or both: The co-existence of natural and supernatural explanatory frameworks across development. *Cognitive Science*, *32*(4), 607–642.

Legare, C. H., & Shtulman, A. (2018). Explanatory pluralism across cultures and development. *Metacognitive Diversity: An Interdisciplinary Approach*, 415.

Lombrozo, T. (2007). Simplicity and probability in causal explanation. *Cognitive psychology*, *55*(3), 232–257.

Poole, D. (1993). Probabilistic horn abduction and bayesian networks. *Artificial intelligence*, *64*(1), 81–129.

Pustejovsky, J. (2013). Dynamic event structure and habitat theory. In *Proceedings of the 6th international conference on generative approaches to the lexicon* (pp. 1–10).

Raghavan, S., & Mooney, R. J. (2010). Bayesian abductive logic programs. In *Statistical relational artificial intelligence* (pp. 82–87).

Shtulman, A., & Lombrozo, T. (2016). Bundles of contradiction: A coexistence view of conceptual change. *Core knowledge and conceptual change*, 49–67.

Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. *Learning and instruction*, *4*(1), 45–69.