Title
Mapping a Plurality of Explanations with NLP: A Case Study of Mothers and Health Workers in India

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 45(45)

Authors
Friedman, Scott E
Schmer-Galunder, Sonja
Sarathy, Vasanth
et al.

Publication Date
2023

Peer reviewed

Mapping a Plurality of Explanations with NLP: A Case Study of Mothers and Health Workers in India

Scott E. Friedman, Sonja Schmer-Galunder, Vasanth Sarathy, Ruta Wheelock, Matthew McLure, Drisana M. Mosaphir, Robert P. Goldman, Noam Benkler, Pavan Kantharaju
{friedman, sgalunder, vsarathy, rwheelock, mmclure, dmosaphir, rpgoldman, nbenkler, pkantharaju}@sift.net
SIFT, Minneapolis, MN, USA

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

Cristine H. Legare (legare@austin.utexas.edu)
Department of Psychology, The University of Texas at Austin, Austin, TX, USA

Abstract

Understanding the values, norms, behaviors, and causal beliefs of communities is a central goal of cognitive science, with practical benefits of grasping and improving community factors such as healthcare delivery. These cultural causal beliefs are evident, in part, within narratives, interview transcripts, ethnography, and other textual sources, but analyzing these texts presently involves tedious expert hand-coding or relatively shallow qualitative text analysis or classification. We present a novel approach for extracting graphical causal models from text via NLP, including qualitative causality, intentions, teleology, sentiment, welfare, social influence, and other rationale. The factors (i.e., nodes) of these causal models are tagged with ethnographic attributes and word-senses, allowing aggregation of causal models over thousands of passages to identify correlations and recurring themes. We apply this approach to a corpus of narrative interviews about maternal and child health and healthcare delivery in Bihar, India, corroborating the hand-coded results of human experts and also identifying novel insights about explanatory structure.

Keywords: NLP; causal reasoning; computational social science; cultural anthropology; cognitive science

Introduction

Eliciting explanations and narratives—and explanations *of* narratives—is a widespread, effective method of gathering evidence of a community's diverse causal beliefs, values, and mental models. Unfortunately, extracting and summarizing these mental models is presently a time-intensive, manual task, and the manual outcome may be difficult to formalize or compare across domains. Fortunately, advances in *natural language processing* (NLP) have made it easier to extract entities and relations from text (Devlin et al., 2019; Eberts & Ulges, 2020; Friedman et al., 2022) and theoretical advances in cognitive science inform how we might encode the diverse plurality of explanatory structures (Legare & Shtulman, 2018; Shtulman & Lombrozo, 2016; Lombrozo, 2010).

This paper presents a domain-general computational approach that (1) extracts semantic graphs from passages of text using NLP, (2) uses these semantic graphs to approximate causal mental models, combining causal relations from the AI and cognitive psychology literature, (3) summarizes causal factors using context-sensitive tagging of word-senses (Loureiro & Jorge, 2019; Fellbaum, 2010) and (4) summarizes these causal models over multiple (potentially thousands) of passages to characterize beliefs and rationale within and across narratives and populations.

We apply our approach on a pre-existing dataset of interviews of mothers and health workers in India. The dataset includes over 10,000 responses explaining the choices and behaviors of characters within narrative vignettes focused on maternal and child health and healthcare delivery.

The empirical results we present in this work provide evidence for our primary research claims:

- Diverse causal relations from across cognitive science—including qualitative increase/decrease, intentions and goals, teleology, welfare, and social influences—can be extracted from explanations using transformer-based NLP.
- These diverse causal relations help characterize the causal structure of textual explanations and they allow the summary and comparison of explanations across populations and across narratives.
- Summarizing causal factors with word-senses (such that "pneumonia" and "fever" are both tagged as **ill_health.n**) helps unify causal factors across individuals and narratives, despite differences in syntax and word choice.

We continue with a review of relevant work in NLP and causal mental models, and then we describe the technical approach of the present work. We then review the dataset used in this work and present the results of our analysis, noting where our approach corroborates human annotators and where it extends the previous analysis. We conclude with a discussion of present limitations and near-term future work.

Related Work: NLP for Causal Language

Recent approaches in NLP extract causal relations from scientific texts (Mueller & Abdullaev, 2019; Eberts & Ulges, 2020; Magnusson & Friedman, 2021), focusing primarily on directed qualitative relations such as *increase* and *decrease*. The present work on causal language analysis is informed by these efforts, as we likewise rely on large language models

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society.*

(Beltagy et al., 2019; Devlin et al., 2019) and we also utilize qualitative causal relations. Unlike these scientific causal relation extractors, we incorporate causal relations and attributes to represent socially-informed, belief-informed, and goal-driven behavior.

Other NLP research is aimed at capturing social norms (Forbes et al., 2020) or cultural values (Benkler et al., 2022) using large language models, but these do not extract or map out explicit causal relationships between factors, which is a core contribution of this work, as we describe next.

A Plurality of Causal Structures

Our approach assumes *explanatory pluralism* (Legare & Shtulman, 2018; Shtulman & Lombrozo, 2016) in that we do not classify texts or explanations with any *single* category of causality; rather, one explanation may coherently include diverse causal factors—such as biological and supernatural—in the same causal chain (e.g., Legare et al., 2012). We therefore characterize explanations by aggregating the multiple factors and relations within their rationale.

In this explanatory pluralism setting, we represent explanations as graphs. For concrete examples, we refer to Figures 1-5, demonstrating the output of our computational approach for the sentence "Anita fed colostrum to her child so that it does not have any diseases such as pneumonia, fever, turning body color blue" from our dataset of interview responses. Each graph contains *factors* (nodes or multi-node subgraphs) and *relations* (directed, labeled edges between factors). Each node of a factor may have multiple *attributes* that indicate multi-class labels and *word senses* estimating the WordNet synonym sets (Fellbaum, 2010) for that node, within the sentence context.

Figure 1 shows the semantic nodes and relations extracted directly from text via NLP, where nodes are colored by attributes and edge color indicates semantic roles (gray) or causal status (blue). Figure 2 shows the attributes, and word-senses for the factor "diseases" (possessed by "her" "child"). Figure 3 shows the causal model the system refines from Figure 1, by applying the relations and attributes of the Figure 1 semantic parse globally. The rest of this section describes some relevant causal relations represented in this work, including some not shown in Figures 1-5.

Qualitative change. Qualitative causality is prevalent in language, e.g., "prices dropped" or "Having more children will increase expenses," absent numbers or quantitative functions. Mention of qualitative changes may indicate hedging, knowledge of a longer causal chain, or knowledge of an underlying causal mechanism (Ahn & Kalish, 2000); despite this ambiguity, this language conveys a meaningful, directed influence. The qualitative reasoning and simulation literature has formalized *qualitative proportionalities* when one factor increases or decreases another, and *direct influences* when one factor increases by the rate of another (Forbus, 2019, 1984; Kuipers, 1986), and qualitative probabilistic networks also provide semantics for one variable qualitatively increas-

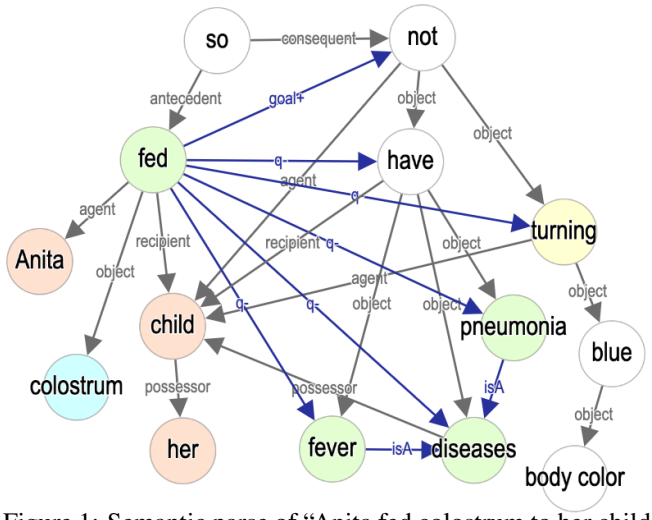

Figure 1: Semantic parse of "Anita fed colostrum to her child so that it does not have any diseases such as pneumonia, fever, turning body color blue."

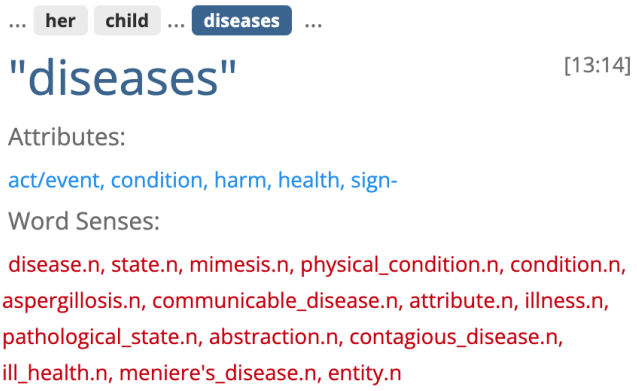

Figure 2: Attributes and word senses inferred for the span "diseases" within the context of the parse shown in Figure 1.

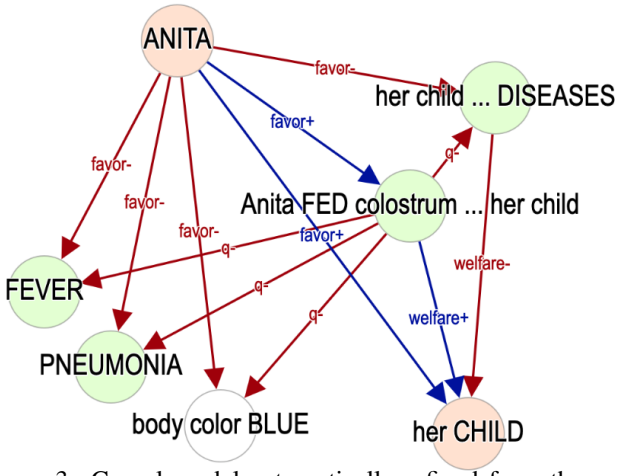

Figure 3: Causal model automatically refined from the semantic parse shown in Figure 1.

ing or decreasing another's likelihood (Wellman, 1990). Informed by this formalism, we capture **q+** (increase) and **q-** (decrease) and **q*** (unspecified change) edges in semantic parses (e.g., Figure 1) and causal graphs (e.g., Figure 3).

Condition and welfare. In domains of social interaction, commonsense reasoning, and healthcare, the welfare and well-being are central considerations for agent decisions. Our schema represents impacts on well-being and welfare at the node-level with **health** and **condition** attributes. This is shown in Figure 2 for the node "diseases," which also accompanies the **sign-** attribute, indicating negative valence for welfare influence. As shown in Figure 3, our system propagates local welfare influences like "disease" throughout the global graph to build a causal model, such that the feeding of colostrum is **welfare+** for the child since it *prevents* (**q-**) the "diseases" which is **welfare-** for the child.

Intention, function, and favor. Intentional, goal-directed behaviors (opposed to accidental, mechanistic behaviors) support purposeful, design-based explanations (Dennett, 1989) and carry a distinctive causal status (Lombrozo, 2010; Lagnado & Channon, 2008; McClure et al., 2007). Consequently, we represent intentional indicators (e.g., "tried to" or "plans to") with **intent** attributes, and we represent factor-to-factor intention as **goal+** relations, e.g., from "fed (colostrum)" to "not (have diseases)" in Figure 1, expressing that the goal of feeding colostrum is to not have diseases. Intention also manifests as teleological explanations of object function, design, and affordance (Pustejovsky, 1991; Lombrozo & Carey, 2006), which we capture as **function** links (not shown in Figures 1-5).

Actors' intentions and preferences are evidence for what they want to achieve, maximize, or minimize, and what/who they want to benefit. We encode this as **favor+** (likewise, **favor-**) for an actor's intent to maximize (likewise, minimize) the likelihood, amount, or benefit of the target. In Figure 3, the system infers that Anita is **favor+** to achieving the feeding of colostrum and the benefit of her child, and she is **favor-** to all of the elements that feeding colostrum prevents, using the causal ascription we describe below. These relations capture actors' various perspectives in the narrative, which may capture conflicting objectives across actors in the narrative.

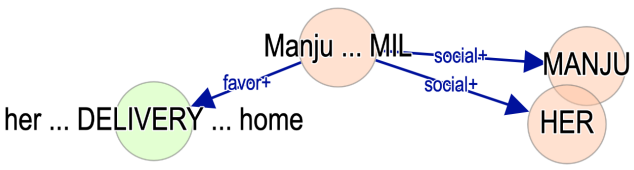

Figure 4: Causal model produced for "Manju's MIL must have told her to have delivery at home" from the dataset.

Social influence. People's intentions—or even their inferred intentions—influence other people (Malle, 2006). Social influence may be direct ("she told me to give birth at home") or indirect ("I do it because the other kids do"). In the domains of traditional and biomedical medicine, social influence impacts the rituals of both patients and practitioners, and the influencer might be the broader community, e.g., where fear of ostracism influences an individual's decision-making (Legare et al., 2020). We represent direct social influence as **social+/-** links, e.g., where one individual directly influences another (e.g., intending, asking, permitting, forbidding), as shown in the Figure 4 causal model for "Manju's MIL must have told her to have delivery at home." This expresses the MIL (mother-in-law) of Manju socially influencing Manju, including the MIL **favor+** for home delivery. Our approach does not presently capture indirect social influence such as fear of ostracism, influence of popularity, or admiration.

Explicit rules and rationale. Norms and causal rationale may manifest with concise but explicit syntax such as "if," "because," "unless," and (in Figure 1) "so." Our approach encodes **rule** and **rationale** attributes on these nodes and attaches **antecedent** and **consequent** links where appropriate, e.g., in Figure 1 where the rationale "so" connects the "fed colostrum" antecedent subgraph to the "not have any disease" consequent subgraph. These are used to infer **rationale+/-** edges in the causal model.

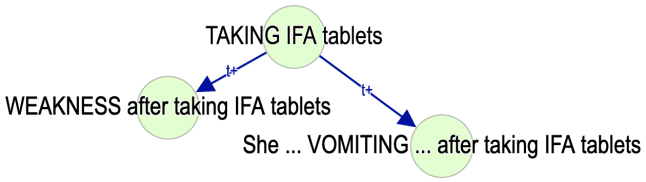

Figure 5: Causal model produced for "She feels like vomiting and weakness after taking IFA tablets" from the dataset.

Temporal precedence. Temporal precedence does not imply causality, but if we assume that an explanation contains mostly relevant information (Grice, 1975), mentioning a sequence of events could be evidence for goal-directed behavior, script-like norms, or mechanism-based causes. The causal model produced for the text "She feels like vomiting and weakness after taking IFA tablets" is shown in Figure 5, containing temporal **t+** relations from the preceding event to the subsequent events. Despite the causal ambiguity, these temporal relations were relevant to the respondent, and these may corroborate other categories of causal relations listed above (e.g., qualitative causality, rationale, intentions, etc.).

Approach

Our technical approach for building a causal model from an unstructured text involves (1) using NLP to parse text into a semantic graph as shown in Figure 1, (2) using NLP to infer word senses of spans as shown in Figure 2, and (3) using the semantic parse and word senses to refine a causal model like those shown in Figures 3-5.

NLP for semantic parsing. Given a passage of text, the system parses unstructured text into relational graphs using the SpEAR NLP architecture (Friedman et al., 2022) which extends the SpERT architecture (Eberts & Ulges, 2020) with attribute prediction (see attributes in Figure 2) and attention-based neural components. SpEAR utilizes a transformer-based encoder (e.g., Devlin et al., 2019; Beltagy et al., 2019) to encode the text into vector representations and then it extracts relevant nodes, relations, and attributes. The output is a graph representation containing (1) nodes, where each corresponds to a continuous sequence of one or more words, (3) a set of directed, labeled relations over the inferred nodes, and (3) a set of attributes on the each inferred nodes.

SpEAR can extract graphs from texts using many possible graph schemas of nodes, relations, and attributes. For the present work, the graph schema includes the causal relations and attributes described above, additional relations (i.e., all gray relations in Figure 1) informed by work in semantic role labeling (Palmer et al., 2010; Bonial et al., 2014), and attributes representing different social institutions based on anthropology theory (Weber, 2017). SpEAR has been used in previous work to characterize moral disengagement in language (Friedman et al., 2021) and extract the causal structure of scientific claims (Magnusson & Friedman, 2021).

NLP for word-sense tagging. Given the graph representation from the previous step, a LMMS language model (Loureiro & Jorge, 2019) encodes the text into vectors and matches each extracted node to zero or more word-senses from WordNet (Fellbaum, 2010) using a cosine similarity metric. After identifying proximal word senses, the system traverses upward through the WordNet hierarchy to assign more abstract word senses such as **physical_condition.n** for "disease" in Figure 2. This word-sense inference may contain mismatches and over-general senses (such as **entity.n** and **attribute.n** as senses for "disease" in Figure 2, which are too general to be actionable in this analysis) but other tags provide practical semantic tags for indexing causal factors, as we demonstrate in our results.

Causal model refinement. The system then approximates a causal model from the semantic graph, attributes, and word senses, which we describe using the outputs shown in Figures 1-3. It begins by adding causal links directly from the semantic parse: the various **q-** links in the Figure 1 parse and their incident nodes are added to the Figure 3 causal model; the **goal+** intentional relation in the parse supports a **favor+** link in the causal model; the "diseases" node with its **harm** attribute in Figure 2 supports a **welfare-** from "diseases" to the possessor "her child;" and so forth. Next, the system extends the causal model using domain-general patterns. For example, Anita's **favor+** for feeding colostrum, combined with the **q-** of feeding colostrum to the various ailments, supports inferences that Anita is **favor-** to those ailments. Further, the feeding preventing (**q-**) the diseases, combined with the diseases' **welfare-** to the child, is evidence that the feeding is **welfare+** to the child. These and other causal patterns iteratively extend the causal graph until no more causal relations can be inferred.

Finally, the system filters nodes from its causal model that are causally redundant with other nodes. For example, the nodes for "not" and "have" in Figure 1 are pruned from the causal model in Figure 3 because the downstream node "diseases" captures the same causal structure.

Dataset

The dataset used in this work includes responses from interviews with mothers and Accredited Social Health Activists (ASHAs) in Bihar, India (Legare et al., 2020). The interviews involve narrative vignettes where a young mother is considering doing something the ASHA recommends, and the respondent is prompted to explain the characters' rationale among other factors. The interviews include eight topics of ASHA advice, each of which is plotted in Figure 6(a-d) and Figure 7(e-h): feeding the baby colostrum; exclusive breastfeeding; taking IFA tablets; hospital or home delivery; family planning without children; family planning as a parent; vaccinations during pregnancy; and vaccinations during infancy.

Across the eight vignettes, respondents explained why the mother chose to follow the ASHA's advice (blue bars in Figures 6 and 7) or chose not to (orange bars in Figures 6 and 7), comprising 1,570 explanations across participants. The respondents also answered questions about decision-making roles, resolving disagreements, and on how the timing and rationale for ASHA recommendations, for a total of 10,320 responses. The present work focuses on the 1,570 responses providing rationale for the mother's decision-making, which we describe next.

Results

Figures 6(a-d) and 7(e-h) plot six categories of causal relations (labeled "Relations" at bottom) and nine categories of causal factors (labeled "Node/Subgraph Factors" at bottom), over eight different vignette topics. We review each of these topics, highlighting key differences within and across prompts. In all plots, the y-axis plots the number of instances of each causal relation or factor per participant explanation, log-scaled with averages labeled above each series with 95% confidence intervals. Significance is computed with paired t-tests and indicated under relations and factors as $\sim$ ($p \leq 0.1$), * ($p \leq 0.05$), ** ($p \leq 0.01$), *** ($p \leq 0.001$), **** ($p \leq 0.0001$).

One broad theme across prompts is a significantly increased incidence of mentioning actual diseases, strength/health, and well-being for characters as rationale for ASHA-consistent decisions (blue) versus ASHA-inconsistent decisions (orange). The **welfare+** causal relation occurs significantly more frequently in ASHA-consistent rationale, with an order of magnitude higher frequency for IFA and infant vaccination, as evidence that ASHA-consistent advice is believed to benefit characters. The **Strong/Healthy** word-sense factor was at significantly higher in ASHA-consistent

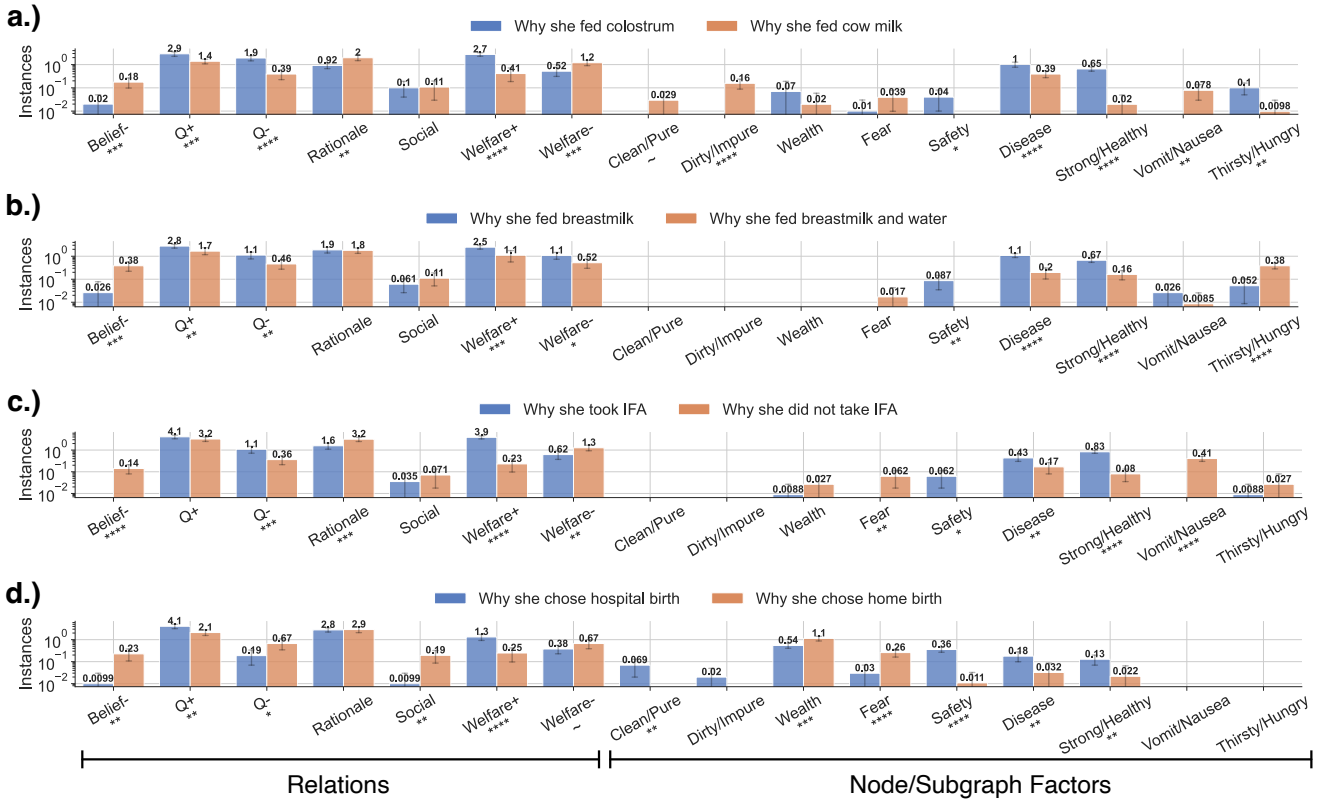

Figure 6: Results of parsing respondents' explanations for characters' choices in four healthcare narratives: (a) a mother chooses to feed her child colostrum *vs.* cow milk; (b) a mother chooses to feed her child breastmilk *vs.* breastmilk with water; (c) a pregnant woman chooses to take IFA tablets *vs.* not taking IFA; (d) a woman chooses to give birth at the hospital *vs.* at home.

rationale for most prompts, by an order of magnitude in most. Likewise, the **Disease** and **Safety** word-sense factors appeared significantly more frequently in all ASHA-consistent rationale except family planning, presumably because family planning is not a primary disease prevention or health security strategy.

A second broad theme is that characters' beliefs, fears, and social influence only appeared significantly as rationale for ASHA-*inconsistent* decisions. For instance, **Fear** is mentioned as a significant rationale for avoiding IFA (c), avoiding hospital births (d), and avoiding both family planning (e-f) and vaccines (g-h). Likewise, characters' false beliefs or lack of knowledge (**Belief-**) was cited significantly higher for all ASHA-inconsistent behaviors except family planning. Finally, direct social influence **Social** only contributed significantly as an influence to deliver at home. Combined with the above theme, this suggests a tendency toward explaining ASHA-consistent behaviors with objective mechansisms and health benefits, and ASHA-inconsistent behaviors with characters' subjective fears and beliefs.

Some relations and factors were more sensitive to the topic of the prompt. **Dirty/Impure** was significant only for feeding cow milk (a) as colostrum is believed by some to be dirty milk, possibly due to its color. **Clean/Pure** was significant for hospital births, countered by **Wealth** for delivering at home due to hospital bills (d). **Wealth** was also a significant rationale for parents to engage in family planning to save costs (f). **Vomit/Nausea** was a significant factor in avoiding IFA tablets (c), and feeding babies cow milk (a), where these factors were not mentioned in the ASHA-recommended rationale.

Figures 6 and 7 plot causal factors but not the characters' sentiment toward them. To this end, we can assess the most- and least-favored concepts across all prompts by tracing the **favor+/-** links inferred by the system. Figure 8 shows the ten least favored (above the line) and ten most favored (below the line) elements, indicating what female characters in the vignettes sought to prevent or diminish (negative favor) or achieve or benefit (positive favor). The least-favored elements comprised primarily illness and disease, with some sub-categories of food or medicine included. The most-favored elements comprised children, feeding, and delivery. As shown in Figure 8, most **favor** links in this domain were positive, about achieving goals and benefiting others.

Conclusion

We described our approach for encoding cognitive science theories about causal reasoning and explanatory pluralism into a NLP system for analyzing explanations. We applied our system on a previous dataset about healthcare delivery in India, using a mix of domain-general causal indicators and

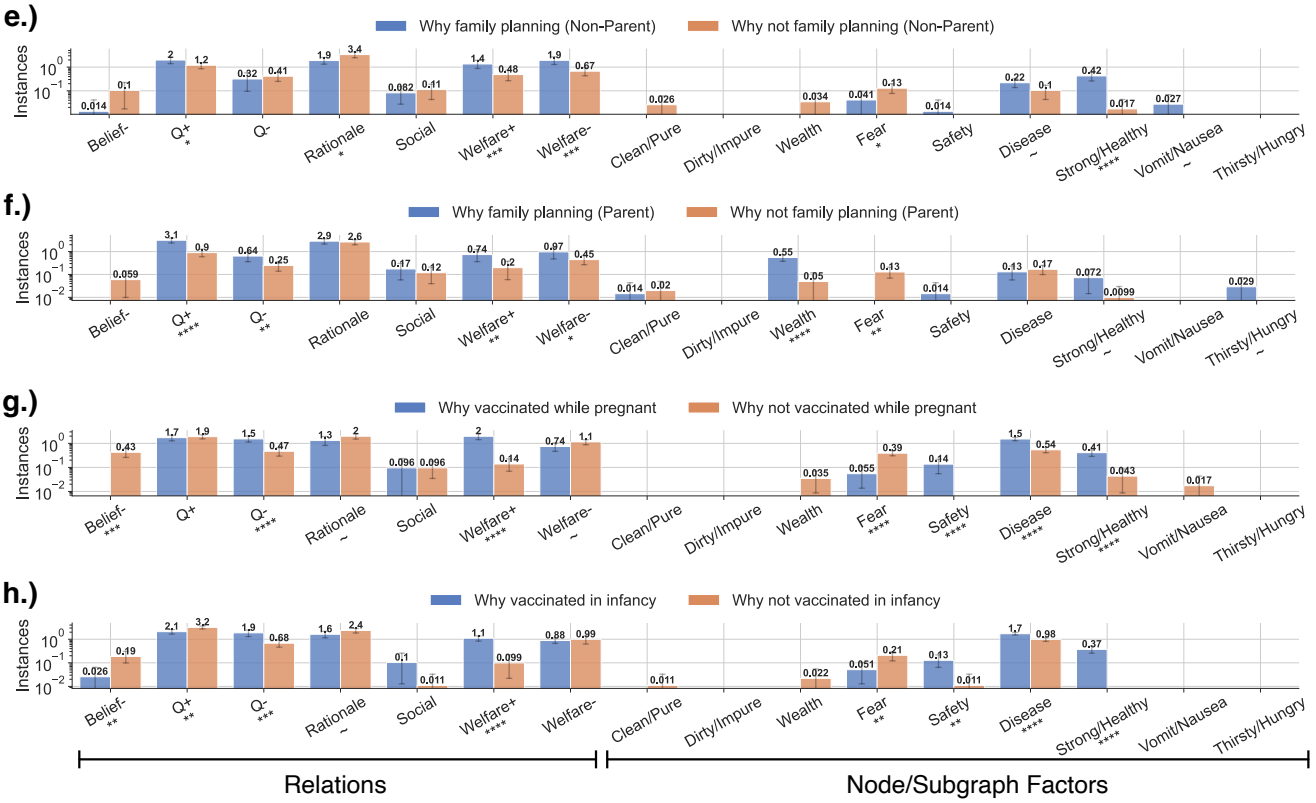

Figure 7: Results of parsing respondents' explanations for characters' choices in four healthcare narratives: (e) a woman without children utilizes family planning *vs.* not; (f) a mother utilizes family planing *vs.* not; (g) a woman is vaccinated during pregnancy *vs.* not; (h) a woman vaccinates her infant *vs.* not.

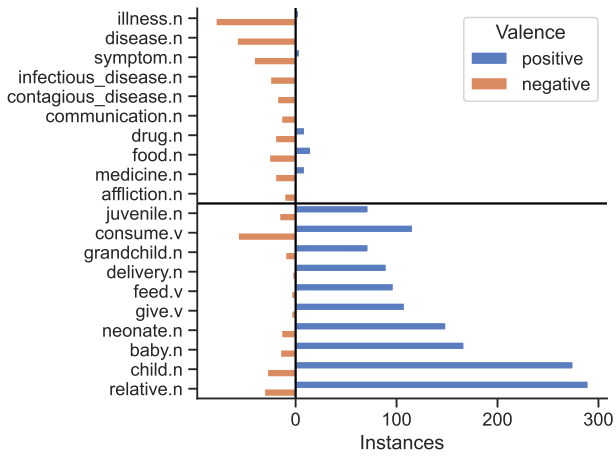

Figure 8: Ten most negative (top) and ten most positive (bottom) favored word senses by female characters.

domain-specific causal cultural and biological factors such as the **Dirty/Impure** factor to corroborate findings about attitudes toward colostrum (Bandyopadhyay, 2009; Khan, 1990). Our results support our claims that diverse causal relations and factors can be extracted from explanations and used to interpret community beliefs and concerns, whether differentiating between conflicting narratives (Figures 6 and 7) or combining them to assess character favor toward or against different factors (Figure 8).

Limitations. The limitations of our approach are important for interpreting the results. Relying on large language models for NLP may suboptimally influence the NLP in ways consistent with sexism and racism, based on their training data (Bolukbasi et al., 2016), and our WordNet resource is known to be over-developed in some areas and incomplete in other areas, including basic categories of gender identity (Hicks et al., 2016). Consequently, utilizing any such technical approach should include assessment of conceptual coverage and the potential for algorithmic inaccuracy or bias.

Future Work. We plan to apply this approach to compare explanations across roles (e.g., mothers versus ASHAs) rather than comparing across prompts, and to apply it to new domains to further demonstrate generality. Finally, we envision using the semantic parses (Figure 1) and causal models (Figure 3) to build *runnable* models and fuzzy cognitive maps (Gray et al., 2014; Jetter & Kok, 2014), ultimately helping summarize collective concern and collective intelligence into a heterogeneous model.

Acknowledgements

The research was supported by funding from the Defense Advanced Research Projects Agency (DARPA HABITUS W911NF-21-C-0007-04). The views, opinions and/or findings expressed are those of the authors and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government. The authors wish to thank reviewers for their helpful feedback.

References

Ahn, W.-k., & Kalish, C. W. (2000). The role of mechanism beliefs in causal reasoning. *Explanation and cognition*, 199–225.

Bandyopadhyay, M. (2009). Impact of ritual pollution on lactation and breastfeeding practices in rural west bengal, india. *International Breastfeeding Journal*, *4*(1), 1–8.

Beltagy, I., Lo, K., & Cohan, A. (2019). SciBERT: A pretrained language model for scientific text. *arXiv preprint arXiv:1903.10676*.

Benkler, N., Friedman, S., Schmer-Galunder, S., Mosaphir, D., Sarathy, V., Kantharaju, P., ... Goldman, R. P. (2022). Cultural value resonance in folktales: A transformer-based analysis with the world value corpus. In *SBP-BRiMS 2022* (pp. 209–218).

Bolukbasi, T., Chang, K.-W., Zou, J. Y., Saligrama, V., & Kalai, A. T. (2016). Man is to computer programmer as woman is to homemaker? debiasing word embeddings. *Advances in neural information processing systems*, *29*.

Bonial, C., Bonn, J., Conger, K., Hwang, J. D., & Palmer, M. (2014). Propbank: Semantics of new predicate types. In *Lrec* (pp. 3013–3019).

Dennett, D. C. (1989). *The intentional stance*. MIT press.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019, June). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT 2019* (pp. 4171–4186). Minneapolis, Minnesota: Association for Computational Linguistics.

Eberts, M., & Ulges, A. (2020). Span-based joint entity and relation extraction with transformer pre-training. *24th European Conference on Artificial Intelligence*.

Fellbaum, C. (2010). Wordnet. In *Theory and applications of ontology: computer applications* (pp. 231–243). Springer.

Forbes, M., Hwang, J. D., Shwartz, V., Sap, M., & Choi, Y. (2020). Social chemistry 101: Learning to reason about social and moral norms. *arXiv preprint arXiv:2011.00620*.

Forbus, K. D. (1984). Qualitative process theory. *Artificial Intelligence*, *24*(1-3), 85–168.

Forbus, K. D. (2019). *Qualitative representations: How people reason and learn about the continuous world*. MIT Press.

Friedman, S., Magnusson, I., Sarathy, V., & Schmer-Galunder, S. (2022). From unstructured text to causal knowledge graphs: A transformer-based approach. *arXiv preprint arXiv:2202.11768*.

Friedman, S., Magnusson, I., Schmer-Galunder, S., Wheelock, R., Gottlieb, J., Miller, C., et al. (2021). Toward transformer-based nlp for extracting psychosocial indicators of moral disengagement. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 43).

Gray, S. A., Zanre, E., & Gray, S. R. (2014). Fuzzy cognitive maps as representations of mental models and group beliefs. *Fuzzy cognitive maps for applied sciences and engineering: From fundamentals to extensions and learning algorithms*, 29–48.

Grice, H. P. (1975). Logic and conversation. In *Speech acts* (pp. 41–58). Brill.

Hicks, A., Rutherford, M., Fellbaum, C., & Bian, J. (2016). An analysis of wordnet's coverage of gender identity using twitter and the national transgender discrimination survey. In *Proceedings of the 8th global wordnet conference (gwc)* (pp. 123–130).

Jetter, A. J., & Kok, K. (2014). Fuzzy cognitive maps for futures studies—a methodological assessment of concepts and methods. *Futures*, *61*, 45–57.

Khan, M. (1990). Breast-feeding and weaning practices in india. *Asia Pac Popul J*, *5*(1), 71–88.

Kuipers, B. (1986). Qualitative simulation. *Artificial Intelligence*, *29*(3), 289–338.

Lagnado, D. A., & Channon, S. (2008). Judgments of cause and blame: The effects of intentionality and foreseeability. *Cognition*, *108*(3), 754–770.

Legare, C. H., Akhauri, S., Chaudhuri, I., Hashmi, F. A., Johnson, T., Little, E. E., ... others (2020). Perinatal risk and the cultural ecology of health in bihar, india. *Philosophical Transactions of the Royal Society B*, *375*(1805), 20190433.

Legare, C. H., Evans, E. M., Rosengren, K. S., & Harris, P. L. (2012). The coexistence of natural and supernatural explanations across cultures and development. *Child development*, *83*(3), 779–793.

Legare, C. H., & Shtulman, A. (2018). Explanatory pluralism across cultures and development. *Metacognitive diversity: An interdisciplinary approach*, *415*.

Lombrozo, T. (2010). Causal–explanatory pluralism: How intentions, functions, and mechanisms influence causal ascriptions. *Cognitive psychology*, *61*(4), 303–332.

Lombrozo, T., & Carey, S. (2006). Functional explanation and the function of explanation. *Cognition*, *99*(2), 167–204.

Loureiro, D., & Jorge, A. (2019, July). Language modelling makes sense: Propagating representations through WordNet for full-coverage word sense disambiguation. In *Proceedings of the 57th annual meeting of the association for computational linguistics* (pp. 5682–5691). Florence, Italy: Association for Computational Linguistics.

Magnusson, I. H., & Friedman, S. E. (2021). Extracting fine-grained knowledge graphs of scientific claims: Dataset and

transformer-based results. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Malle, B. F. (2006). *How the mind explains behavior: Folk explanations, meaning, and social interaction*. MIT press.

McClure, J., Hilton, D. J., & Sutton, R. M. (2007). Judgments of voluntary and physical causes in causal chains: Probabilistic and social functionalist criteria for attributions. *European journal of social psychology*, *37*(5), 879–901.

Mueller, R., & Abdullaev, S. (2019). Deepcause: Hypothesis extraction from information systems papers with deep learning for theory ontology learning. In *Proceedings of the 52nd hawaii international conference on system sciences*.

Palmer, M., Gildea, D., & Xue, N. (2010). Semantic role labeling. *Synthesis Lectures on Human Language Technologies*, *3*(1), 1–103.

Pustejovsky, J. (1991). The syntax of event structure. *Cognition*, *41*(1-3), 47–81.

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

Weber, M. (2017). *Methodology of social sciences*. Routledge.

Wellman, M. P. (1990). Fundamental concepts of qualitative probabilistic networks. *Artificial intelligence*, *44*(3), 257–303.