Tell Me Your (Cognitive) Budget, and I'll Tell You What You Value: Evidential Relationships Between Values, Data, and Generic Causal Claims about the Social World

David Kinney (david.kinney@princeton.edu)
Department of Psychology, Princeton University
Princeton, NJ 08540 USA

Tania Lombrozo (lombrozo@princeton.edu)
Department of Psychology, Princeton University
Princeton, NJ 08540 USA

Abstract

Consider the following two (hypothetical) generic causal claims: "Attending an all-girls school improves girls' math scores" and "attending an affluent all-girls school improves girls' math scores." These claims not only differ in what they suggest about how test scores are distributed across different types of schools (i.e., "the data"), but also have the potential to communicate something about the speakers' values: namely, the prominence they accord to affluence in representing and making decisions about the social world. Here, we examine the relationship between the level of granularity with which a cause is described in a generic causal claim (e.g., all-girls school vs. affluent all-girls school) and the value of the information contained in the causal model that generates that claim. We argue that listeners who know any two of the following can make reliable inferences about the third: 1) the level of granularity at which a speaker makes a generic causal claim, 2) the speaker's decision-theoretic values, and 3) the data available to the speaker. We present results of three experiments in the domain of social categories (N=853) that provide evidence in keeping with these predictions.

Keywords: causation; granularity; generics; social categories.

Introduction

At an event introducing the United States' 2023 federal budget, President Joe Biden began his remarks by saying: "Don't tell me what you value. Show me your budget, and I'll tell you what you value." An agent's choices about how to spend their resources often provide better evidence of that agent's values than their explicit claims. Just as governments have fiscal budgets, agents have *cognitive* budgets in their representations of the environment. And just as fiscal choices can tell us about a government's values, so too can representational choices tell us about an agent's values.

In this paper we present an analysis that illuminates the relationships of mutual constraint that hold between an agent's representational "budget" (i.e., the granularity with which they represent the causal structure of the world), the agent's values (i.e., the decision-theoretic thresholds that determine which information from the environment is valuable to preserve), and the data they are representing. We report three experiments that support these relationships of mutual constraint, such that any one of an agent's representation of the environment, their values, or their data about the environment can be inferred given the other two. We then argue that our results support a central role for decision-theoretic factors in explaining why particular social categories feature in a given person's model of the causal structure of their social environment, and that this has implications for the value judgments we draw about individuals who use particular causal generics.

The representations that we consider are generic causal claims about the social world. Consider the following two claims, which vary in their level of granularity:

- **COMPLEX:** Attending a majority-white school with a small recent immigrant population improves math performance.
- **SIMPLE:** Attending a school with a small recent immigrant population improves math performance.

These are both generic causal claims of the form '*c* causes *e*.' As indicated by their labels, the first claim is strictly more complex than the second in the sense that the cause is described in strictly more detail: the generic cause cited in the first claim contains details about a school's composition in terms of both race and immigration status, while the second claim cites a cause that only contains details about a school's immigration status. Thus, if we compare two speakers who we take to be describing the same causal process based on the same data, one of whom states COMPLEX and the other of whom states SIMPLE, then we begin to get some insight into their respective cognitive budgets. Correspondingly, we get some sense for how important it is to each agent to be precise about a specific causal relationship in their social world.

How exactly does the level of granularity at which a speaker makes a causal claim (and the inferences it licenses about that speaker's cognitive budget) tell us about that speaker's values? Following Tessler and Goodman (2019), we hold that listeners accept a causal claim as true (and that speakers expect listeners to accept a given causal claim as true) when the probability of the effect conditional on the

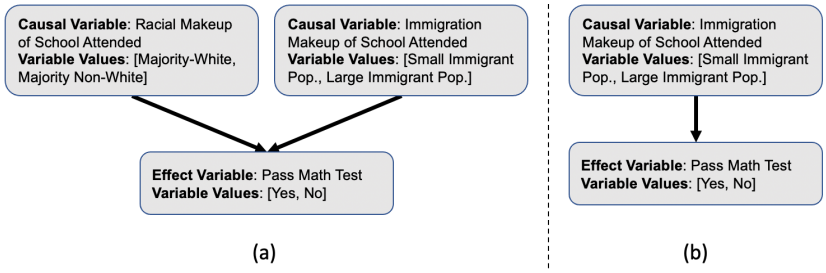

Figure 1: Possible causal models representing the data-generating social system, where (a) corresponds to COMPLEX and (b) to SIMPLE.

cause is above a certain threshold. More precisely, letting e and c be values of variables C and E in a causal model of the relevant system, listeners accept a generic causal claim of the form 'c causes e' as true just in case $p(e|do(c)) > \theta$, where $\theta \in [0,1]$. We follow Pearl (2000) in interpreting the operator $do(c)$ as an exogenous intervention on the system such that $C=c$ (Tessler and Goodman adopt a slightly different causal formalism). In the case of the two causal claims above, suppose that the speaker believes that Table 1 contains accurate data regarding the racial composition, immigration-status composition, and math performance on a math test of different schools. Assume further that these frequencies can be interpreted as interventional conditional probabilities.

Table 1: Hypothetical data about math performance in a large school district.

	% of Students who Pass
Majority-White Schools with Small Recent Immigrant Population	60
Majority-Non-White Schools with Small Recent Immigrant Population	45
Majority-White Schools with Large Recent Immigrant Population	25
Majority-Non-White Schools with Large Recent Immigrant Population	10

Now suppose that the speaker endorses *one* of the following claims:

- **HIGH THRESHOLD.** A school is providing adequate math education if at least 50% of its students pass the math test in question.
- **LOW THRESHOLD.** A school is providing adequate math education if at least 30% of its students pass the math test in question.

A speaker who endorses HIGH THRESHOLD, we argue, is more likely to generate the causal claim COMPLEX than the causal claim SIMPLE. If we take attending a given school to improve math performance just in case enough students pass the test in question for that school to be providing adequate math education, then endorsing HIGH THRESHOLD amounts to adopting a threshold of $\theta = .5$ for accepting as true any claim of the form 'c improves performance on math tests' where c is any value of a causal variable. Between the generic causal claims COMPLEX and SIMPLE, only COMPLEX necessarily satisfies this truth condition. By contrast, a speaker who endorses LOW THRESHOLD adopts a threshold of $\theta = .3$ for accepting as true any claim of the form 'c improves performance on math tests.' Though both COMPLEX and SIMPLE satisfy this truth condition, SIMPLE is less complex, and so a speaker who endorses LOW THRESHOLD while including the claim SIMPLE and not the claim COMPLEX in their mental causal model of their social world is able to expend less of their overall cognitive complexity budget without making sacrifices as to the truth of their claims.

On the basis of the reasoning given above, we claim that in this case and in similarly-structured cases, a speaker who endorses the lower threshold will be judged likely to utter the less complex causal claim, while a speaker who endorses the higher threshold will be judged more likely to utter the more complex causal claim (Prediction 1). We also hypothesize that listeners are capable of making the reverse judgment, and inferring the threshold that a speaker endorses on the basis of the speaker's choice of causal claim (Prediction 2); this is the analog of Biden's claim about budgets.

On our analysis, there is a tight conceptual connection between the acceptance thresholds that speakers endorse when making generic causal claims, and the *decision-theoretic values* of that speaker (i.e., what that agent cares about). This connection is due to the *value of the information* contained in the implicit causal model that a speaker uses to generate a causal claim.

To illustrate, consider a speaker who utters the claim COMPLEX. We take such an utterance to imply that the speaker implicitly represents their social world as having the causal structure given in Figure 1a. Suppose that such a speaker also endorses a 50% threshold for the percentage of students who must pass the test for a school to qualify as providing adequate math performance. This threshold can be understood in decision-theoretic terms: faced with a decision about whether to classify a school as providing adequate math education, this threshold will determine their response. Finally, suppose that Table 1 represents the data regarding the relevant schools. For any given school, if they were to query both of the causal variables in the causal model shown in Figure 1a, then they would have all of the information that they needed to make a decision about whether to deem that school as delivering adequate math education, since it is only by learning both its racial makeup and its immigrant status that they can learn whether, according to their standards, it

provides adequate math education. By contrast, suppose that they could only query the causal variables in the model shown in Figure 1b. They would only learn the composition of the school according to immigration status, and so they would lack information that would be valuable to them in making their decision, since according to their data, even if a school has a small immigrant population, depending on its racial composition it may still not have enough students who pass the math test for it to be considered adequate. This leads them to make a causal claim that implies a mental model that is more complex, but contains more information that they deem decision-relevant.

By contrast, suppose that a speaker endorses a 30% threshold for the percentage of students that must pass the test for a school to be providing adequate math education. For such a speaker, querying the causal variables in the model in Figure 1b would give them all of the information that they need to make a decision about whether to classify a school as providing adequate math education. As such, they can utter the simpler causal claim 'attending a school with a small recent immigrant population improves performance on math tests' without sacrificing any information that is valuable (to them). In their cognitive budgeting process, they are able to purchase simplicity for free, since the information lost in formulating a more coarse-grained causal model of the world is not valuable to them. In this way, speakers' values (i.e., their preferences over states of affairs) both provide evidence for and are evidenced by the level of detail with which they model the causal structure of their social world. For this reason, although Tessler and Goodman regard probabilistic thresholds for accepting a generic causal claim as semantic or truth-conditional thresholds, we regard them primarily as *decision-theoretic* thresholds. That is, they are thresholds that structure the conative attitudes, and subsequent decision norms, of agents. In so doing, we suggest, these thresholds can also influence listeners' expectations as to the generic causal claims that speakers are more likely to utter, given the data available to them. In this way, a speaker's choice of granularity with respect to the causal generics that they use is indicative of their decision-theoretic values, which may or may not be connected to other values such as moral values. Indeed, we take our investigation here to primarily be about decision-theoretic cognition, rather than moral cognition.

We present findings from three experiments that are consistent with our predictions that listers are able to: 1) use information about a speaker's decision-theoretic thresholds and the data available to the speaker to make inferences about the likely level of granularity with which a speaker makes generic causal claims, 2) use information about the level of granularity with which a speaker makes generic causal claims and the data available to the speaker to make inferences about the speaker's decision-theoretic thresholds, and 3) use information about the level of granularity with which a speaker makes generic causal claims and the speaker's decision-theoretic thresholds to make inferences about the data available to the speaker. We conclude by briefly discussing the potential relevance of our results for understanding moral judgments regarding the speakers of generic causal claims involving social categories.

Previous Work

In formal epistemology and philosophy of science, the question of how to make choices about which variable set to use when representing some system has been termed the "variable choice problem" (Woodward, 2016). Kinney (2019) and Kinney and Watson (2020) provide a framework for showing how this problem can be resolved, at least in part, by supposing that agents seek causal models that maximize simplicity while retaining all of the information that is decision-theoretically valuable to them (see Kinney & Lombrozo, 2022, for evidence in favor of a *ceteris paribus* preference for more compressed causal models). The framework presented here extends this idea into the social domain, arguing that agents' use of specific social categories and conjunctions of social categories in their generic causal descriptions of the world can be viewed, at least in part, as a consequence of their desire to represent the world as simply as possible while still accounting for all of the information that the agent needs. In other words, and as argued above, the level of granularity of our generic causal claims about the social world reflects our cognitive budgets. This is an idea with roots in the literature on the psychology of categorization (e.g., Rosch, 1978).

The role of causal generics in thinking about the reality of social groupings comes to the fore in work on psychological essentialism (Gelman, 2003; Haslam et al., 2000) and the "inherence heuristic" (Bigler and Clark, 2014; Cimpian, 2015; Cimpian and Salomon, 2014a, 2014b; Gelman and Roberts, 2017; Hussak and Cimpian, 2018; Salomon and Cimpian, 2014). According to this literature, the use of generics, including causal generics, may communicate that the group is a natural kind, such that individuals are members of the group in virtue of their possessing inherent, essential properties that are stable across time and space (Benitez et al., 2022; Cimpian and Markman, 2011; Foster-Hanson et al., 2022; Gelman, 2013; Leslie, 2014; Ritchie, 2021; Wodak et al., 2015), although the association between generic claims and an essentialist construal has also been called into question (Noyes and Keil, 2019; Prasada et al., 2013; Vasilyeva and Lombrozo, 2020). Indeed, Foster-Hanson and Rhodes note that "while generics can communicate natural kind beliefs, they communicate other information and are open to alternative interpretations as well" (2020, p. 301). Our results identify an additional communicative role for causal generics. Namely, they show that causal generics can be used to communicate information about which features of an individual amount to decision-theoretically valuable information.

Experiments

Experiment 1

Experiment 1 tested Prediction 1, i.e., the prediction that when speakers and listeners share a common data set

produced by some social structure, listeners will judge speakers who endorse lower decision-theoretic thresholds for causal claims to be more likely to utter claims deriving from simpler causal models summarizing the dynamics of that social system, rather than more complex causal claims. To test this, we presented participants with a hypothetical scenario involving either performance of local schools on a math test (in the style of the example from the introduction), the rates at which people in different neighborhoods in a large city own a bicycle, or the performance of children in swimming classes. We manipulated whether a speaker who was asked to summarize these data endorsed a 30% or 50% decision-theoretic threshold. We then asked participants to choose between possible causal claims that the speaker might actually make in their summary of the data, some of which were simpler and others of which were more complex, to test our prediction that when participants are told that the speaker has a lower decision-theoretic threshold, they are more likely to accept simpler causal claims. The data, stimuli, and pre-registrations for all experiments in this paper are available at: tinyurl.com/3ub6tsjy.

Participants Participants were 290 adults recruited via Prolific. An additional 10 participants were excluded for failing comprehension checks. For all studies reported here, participation was restricted to users with a US-based IP address and a 95% rating based on at least 100 previous studies. All studies described in this paper were preregistered, and IRB approval was obtained from the authors' university.

Materials and Procedures Participants read about one of three novel social systems and were given data about that system as well as information about a hypothetical speaker's decision-theoretic thresholds. To illustrate, in one vignette participants were told that in a fictional county with a number of different schools serving different communities, seventh-grade students (i.e., students who are 12-13 years old) take a test to determine whether they are placed into a more advanced Algebra 1 class or a less advanced pre-algebra class in eighth-grade. Participants are told that the data in Table 1 describe the percentages of children from different schools who are placed into Algebra 1 on the basis of their performance on the math test. They are then told that a county math teacher has been asked to summarize these data as part of a broader report on math performance in the county school district. Participants are also told that this teacher either believes that a school should be regarded as delivering adequate math education if either: i) 30% of students are placed into Algebra 1, or ii) 50% of students are so placed. On the basis of this information, participants were then asked which of the following generic causal claims they thought the teacher would be most likely to include in their summary:

- **SIMPLE:** Attending a school with a small recent immigrant population improves math performance.
- **COMPLEX:** Attending a majority-white school with a small recent immigrant population improves math performance.

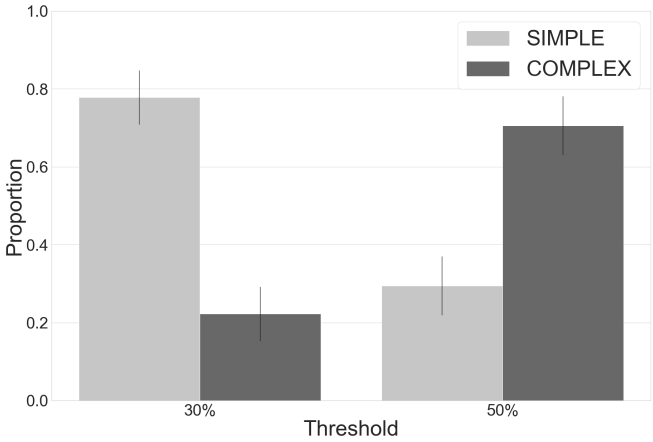

Figure 2**:** Proportion of participants in each threshold condition who selected each generic causal claim as the one most likely to have been made by that speaker, with 95% CIs.

- **CONTROL:** Attending a majority-non-white school with a small recent immigrant population improves math performance.

Participants' choice of claim was the dependent variable in the study. Recall that in the data given to participants, 60% of students in majority-white schools with a small recent immigrant population are placed into Algebra 1, as compared to 45% of students in majority-non-white schools with a small recent immigrant population placed into Algebra 1. For this reason, we took choosing CONTROL over COMPLEX or SIMPLE to indicate a lack of comprehension, and excluded participants who made this choice, along with participants who completed the study in less than 60 seconds.

Results Figure 3 shows the proportion of participants who chose SIMPLE and COMPLEX as the most likely claim made by the speaker for both of the decision-theoretic thresholds that the speaker could endorse. A χ^2 contingency test revealed a significant difference in the proportion of participants choosing either claim across the two threshold conditions ($\chi^2(1) = 66.12, p = 4.25 \times 10^{-16}$). In a test of robustness, we also found significant results when restricting analysis solely to each of the three vignettes (Vignette 1 (Math Performance): $\chi^2(1) = 16.03, p = 6.22 \times 10^{-5}$; Vignette 2 (Bicycle Ownership): $\chi^2(1) = 25.00, p = 5.73 \times 10^{-7}$; Vignette 3 (Swimming Performance): $\chi^2(1) = 21.91, p = 2.85 \times 10^{-6}$). These results are in keeping with our prediction that listeners are more likely to expect a speaker to utter the simpler causal summary of a given data set when the probabilistic threshold is lower.

Experiment 2

Experiment 2 tested Prediction 2, i.e., the prediction that when speakers and listeners share a common data set produced by some social structure, listeners will judge speakers who utter simple, coarse-grained generic causal claims more likely to endorse low decision-theoretic thresholds. To test this, we presented participants with the same hypothetical scenarios as in Experiment 1, with the

Table 2: Possible data sets that participants were told speakers might be viewing.

Data Set 1		Data Set 2		Data Set 3	
	% of Students who Pass the Math Test		% of Students who Pass the Math Test		% of Students who Pass the Math Test
Majority-White Schools with Small Recent Immigrant Population	60	Majority-White Schools with Small Recent Immigrant Population	60	Majority-White Schools with Small Recent Immigrant Population	60
Majority-Non-White Schools with Small Recent Immigrant Population	45	Majority-Non-White Schools with Small Recent Immigrant Population	25	Majority-Non-White Schools with Small Recent Immigrant Population	55
Majority-White Schools with Large Recent Immigrant Population	20	Majority-White Schools with Large Recent Immigrant Population	20	Majority-White Schools with Large Recent Immigrant Population	20
Majority-Non-White Schools with Large Recent Immigrant Population	10	Majority-Non-White Schools with Large Recent Immigrant Population	10	Majority-Non-White Schools with Large Recent Immigrant Population	10

amendment that two speakers offered summaries of the data. One speaker's summary contained only SIMPLE, while the other contained only COMPLEX. Participants were then asked to judge which speaker was more likely to endorse a lower threshold.

Participants Participants were 147 adults recruited via Prolific, with three participants excluded for failing comprehension checks.

Methods and Procedures Participants read about a novel social system (the same systems as in Experiment 1), and were given data about that system as well as information about a hypothetical speaker's decision-theoretic thresholds. To illustrate, in one vignette participants were given the same scenario and data about Algebra 1 placement in a fictional county school district as in Experiment 1. However, in this experiment participants were then told that *two* teachers had produced reports summarizing the data, with one teacher summarizing the data using only the claim SIMPLE and the other teacher summarizing the data using only the claim COMPLEX. Participants were then asked which teacher they believe is more likely to endorse the claim that a school is providing adequate math education as long as 30% of students are placed in Algebra 1; their answer to this binary question is our dependent variable. We excluded participants who spent less than 60 seconds completing the task, or who incorrectly answered a factual question about their vignette.

Results As predicted, a majority of participants (68.03%, $p = 1.47 \times 10^{-5}$) identified the speaker who made the simpler generic causal claim as the one who endorsed the lower decision-theoretic threshold for generic causal claims in the relevant context. Restricting the results to each vignette, we saw 70.59% of participants ($p = .005$) in Vignette 1 (algebra placement) identify the speaker of the simpler causal claim as having the lower decision-theoretic threshold, as compared to 62.75% ($p = .092$) of participants in Vignette 2 (bicycle ownership) and 74.52% ($p = .001$) of participants in Vignette 3 (swimming performance). We take these results to provide evidence in keeping with our hypothesis that listeners are able to infer a speaker's decision-theoretic thresholds for generic causal claims, which in turn reflect that speaker's decision-theoretic values, from the level of granularity at which a speaker makes a generic causal claim.

Experiment 3

In Experiment 3, we aim to show that when listeners are given both: i) a speaker's generic causal summary of unseen data, and ii) the same speaker's decision-theoretic thresholds for generic causal claims, listeners can make inferences about the data that the speaker is most likely to be summarizing, where these inferences are in keeping with the predictions of our theory. To this end, we ran an experiment in which participants were told about the same hypothetical social systems as they were in Experiments 1 and 2, and were told both the generic causal claim summarizing that data made by a speaker and that speaker's decision-theoretic threshold for generic causal claims. The level of granular detail at which the generic causal claim was made (SIMPLE vs. COMPLEX), as well as the speaker's decision-theoretic threshold (30% vs. 50%), was varied between participants. Participants were then asked to choose which of three data sets that could have been generated by the social system in question was most likely to have in fact been generated, on the basis of the information they were provided. Their answer to this question is our dependent variable.

Participants Participants were 416 adults recruited via Prolific. An additional 4 participants were excluded for failing comprehension checks.

Methods and Procedures Participants read about a novel social system (the same systems as in Experiment 1), and were given a hypothetical speaker's generic causal summary of the unseen data produced by that social system as well as information about that speaker's decision-theoretic thresholds for accepting a causal claim. To illustrate, in one vignette participants were given the same scenario about Algebra 1 placement in a fictional county school district as in

Experiment 1, and told either that the speaker made the claim SIMPLE or the claim COMPLEX, and that the speaker either believed that the threshold for adequate math education should be 30% or 50%. Participants were then asked which of the three data sets shown in Table 2 was most likely to contain the data that the teacher's causal summary was based on. Within our framework, the causal claim COMPLEX is always consistent with Data Set 2; a speaker who believes that the system produces Data Set 2 is more likely to produce the causal claim COMPLEX than to produce SIMPLE, regardless of whether the decision-theoretic threshold is 30% or 50%. However, if the speaker's probabilistic decision-theoretic threshold for causal claims is 50%, then Data Set 1 also renders an utterance of COMPLEX more likely than an utterance of SIMPLE. By contrast, the causal claim COMPLEX is never consistent with Data Set 2, but is always consistent with Data Set 3; a speaker who believes that the system produces Data Set 3 is more likely to produce the causal claim SIMPLE than to produce COMPLEX, regardless of whether the decision-theoretic threshold is set to 30% or 50%. However, if the threshold is set to 30%, then the causal claim SIMPLE is also consistent with Data Set 1. This analysis yields the predicted responses for participants shown in Table 3. As in Experiment 2, participants who completed the survey in <60 seconds or incorrectly answered a question about their vignette were excluded from analysis.

Table 3: Predictions for Experiment 3.

Causal Claim	Threshold	Predicted Choice
SIMPLE	30%	Data Sets 1 or 3
SIMPLE	50%	Data Set 3
COMPLEX	30%	Data Set 2
COMPLEX	50%	Data Set 1 or 2

Results Figure 3 shows the results of Experiment 3. As predicted, Data Set 2 is much more commonly selected as the basis for a speaker's reasoning when the speaker makes a more complex causal claim, whereas Data Set 3 is more commonly selected when the speaker makes a simpler causal claim. Moreover, and also as predicted, when the speaker makes a simpler causal claim, Data Set 1 is more commonly selected when the decision-theoretic threshold for generic causal claims is 30% than when it is 50%.

To test the significance of these patterns, we constructed binary variables denoting whether a participant chose Data Set 1, 2, or 3 as the most likely data set that a speaker used when making their prediction. We performed a logistic regression for each these binary variables against: i) whether the speaker's claim was SIMPLE or COMPLEX, ii) whether the speaker's stated decision-theoretic threshold was 30% or 50%, iii) the vignette shown to participants, and iv) all interaction effects between all variables. For Data Set 1, we found significant effects of: i) the granularity of the speaker's claim ($\beta = 0.33, p = .028$) (i.e., Table 1 is less likely to be selected when the speaker makes the simple causal claim and adopts a 50% decision-theoretic threshold), and ii) the speaker's decision-theoretic threshold ($\beta = 0.30, p = .047$) (i.e., Data Set 1 is more likely to be selected when the speaker

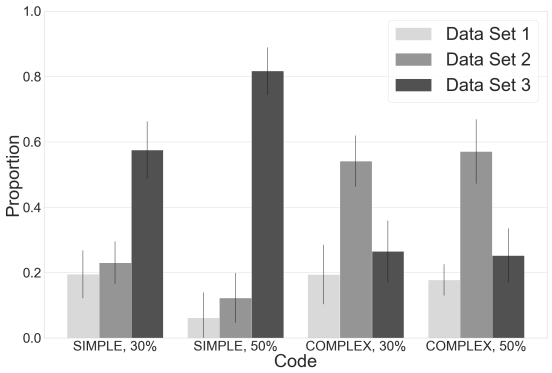

Figure 3: Proportion of participants in each condition who chose each data set, with 95% CIs.

makes the claim SIMPLE and has a decision-theoretic threshold of 30%). In a limitation of our results that warrants further investigation, we did not find a significant interaction effect between the granularity of the speaker's causal claim and the speaker's choice of decision-theoretic threshold ($\beta = -0.29, p = .053$). For Data Set 2, we found a significant effect of the granularity of the speaker's claim, with speakers who made the more complex causal claim more likely to choose Data Set 2 ($\beta = 0.93, p < .001$). For Data Set 3, we found significant effects of: i) the granularity of the speaker's claim ($\beta = -0.99, p < .001$) (i.e., Data Set 3 is more likely to be selected when the speaker makes a simple causal claim), ii) the speaker's decision-theoretic threshold ($\beta = -0.28, p = .015$) (i.e., Data Set 3 is more likely to be selected when the speaker adopts the 50% threshold), and iii) the interaction between granularity and the speaker's chosen decision-theoretic threshold ($\beta = 0.33\ p = .004$) (i.e., speakers who make the simpler causal claim are more likely to be judged to base their claim on Data Set 3 when their stated decision-theoretic threshold is 50%).

General Discussion

Generic causal claims involving social categories are often ethically fraught. We may question the motivations of a speaker who cites social categories that have historically been used as mechanisms for unfair discrimination (e.g., race, gender, and socio-economic status) when making causal claims about their social world. In light of our results, we believe that these judgements may be partially based on the conative attitudes reflected in a speaker's choice of generic causal claims. This explanation differs from the more familiar idea that generic claims about social categories can convey essentialist assumptions; our claim concerns not (only) the representation of social kinds, but the speaker's *decision-theoretic values*. We have shown that people can make relatively sophisticated inferences between causal claims, decision-theoretic values, and data. While our results here are not explicitly brought to bear on agents' *moral* values, we hope in future work to further investigate the relationships between moral judgments and the conative attitudes revealed by the choice of granularity in causal representations of the social world.

References

Benitez, J., Leshin, R. A., & Rhodes, M. (2022). The influence of linguistic form and causal explanations on the development of social essentialism. *Cognition*, *229*, 105246.

Bigler, R. S., & Clark, C. (2014). The inherence heuristic: A key theoretical addition to understanding social stereotyping and prejudice. *Behavioral and Brain Sciences*, *37*(5), 483.

Cimpian, A. (2015). The inherence heuristic: Generating everyday explanations. *Emerging Trends in the Social and Behavioral Sciences: An Interdisciplinary, Searchable, and Linkable Resource*, 1–15.

Cimpian, A., & Markman, E. M. (2011). The generic/nongeneric distinction influences how children interpret new information about social others. *Child Development*, *82*(2), 471–492.

Cimpian, A., & Salomon, E. (2014a). Refining and expanding the proposal of an inherence heuristic in human understanding. *Behavioral and Brain Sciences*, *37*(5), 506–527.

Cimpian, A., & Salomon, E. (2014b). The inherence heuristic: An intuitive means of making sense of the world, and a potential precursor to psychological essentialism. *Behavioral and Brain Sciences*, *37*(5), 461–480.

Foster-Hanson, E., Leslie, S.-J., & Rhodes, M. (2022). Speaking of Kinds: How Correcting Generic Statements can Shape Children's Concepts. *Cognitive Science*, *46*(12), e13223. https://doi.org/10.1111/cogs.13223

Foster-Hanson, E., & Rhodes, M. (2020). The psychology of natural kind terms. *The Routledge Handbook of Linguistic Reference*, 295–308.

Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. Oxford Cognitive Development.

Gelman, S. A. (2013). Artifacts and essentialism. *Review of Philosophy and Psychology*, *4*(3), 449–463.

Gelman, S. A., & Roberts, S. O. (2017). How language shapes the cultural inheritance of categories. *Proceedings of the National Academy of Sciences*, *114*(30), 7900–7907.

Haslam, N., Rothschild, L., & Ernst, D. (2000). Essentialist beliefs about social categories. *British Journal of Social Psychology*, *39*(1), 113–127.

Hussak, L. J., & Cimpian, A. (2018). Memory accessibility shapes explanation: Testing key claims of the inherence heuristic account. *Memory & Cognition*, *46*(1), 68–88.

Kinney, D. (2019). On the explanatory depth and pragmatic value of coarse-grained, probabilistic, causal explanations. *Philosophy of Science*, *86*(1), 145–167.

Kinney, D., & Lombrozo, T. (2022). Evaluations of Causal Claims Reflect a Trade-Off Between Informativeness and Compression. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *44*(44).

Kinney, D., & Watson, D. (2020). Causal feature learning for utility-maximizing agents. *International Conference on Probabilistic Graphical Models*, 257–268.

Leslie, S.-J. (2014). Carving up the social world with generics. *Oxford Studies in Experimental Philosophy*, *1*.

Noyes, A., & Keil, F. C. (2019). Generics designate kinds but not always essences. *Proceedings of the National Academy of Sciences*, *116*(41), 20354–20359.

Pearl, J. (2000). *Causality: Models, Reasoning and Inference*. Cambridge University Press.

Prasada, S., Khemlani, S., Leslie, S.-J., & Glucksberg, S. (2013). Conceptual distinctions amongst generics. *Cognition*, *126*(3), 405–422.

Ritchie, K. (2021). Essentializing inferences. *Mind & Language*, *36*(4), 570–591.

Rosch, E. (1978). *Principles of categorization*.

Salomon, E., & Cimpian, A. (2014). The inherence heuristic as a source of essentialist thought. *Personality and Social Psychology Bulletin*, *40*(10), 1297–1315.

Tessler, M. H., & Goodman, N. D. (2019). The language of generalization. *Psychological Review*, *126*(3), 395.

Vasilyeva, N., & Lombrozo, T. (2020). Structural thinking about social categories: Evidence from formal explanations, generics, and generalization. *Cognition*, *204*, 104383.

Wodak, D., Leslie, S.-J., & Rhodes, M. (2015). What a loaded generalization: Generics and social cognition. *Philosophy Compass*, *10*(9), 625–635.

Woodward, J. (2016). The problem of variable choice. *Synthese*, *193*(4), 1047–1072.