Tell Me Your (Cognitive) Budget, and I'll Tell You What You Value

Forthcoming in Cognition

David Kinney[1] and Tania Lombrozo[2]

[1]Department of Psychology, Yale University

[2]Department of Psychology, Princeton University

Abstract

Consider the following two (hypothetical) generic causal claims: "Living in a neighborhood with many families with children increases purchases of bicycles" and "living in an affluent neighborhood with many families with children increases purchases of bicycles." These claims not only differ in what they suggest about how bicycle ownership is distributed across different neighborhoods (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., neighborhood vs. affluent neighborhood) 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 values, and 3) the data available to the speaker. We present results of four experiments (N=1,323) in the domain of social categories that provide evidence in keeping with these predictions.

Keywords: causation; granularity; generics; social categories.

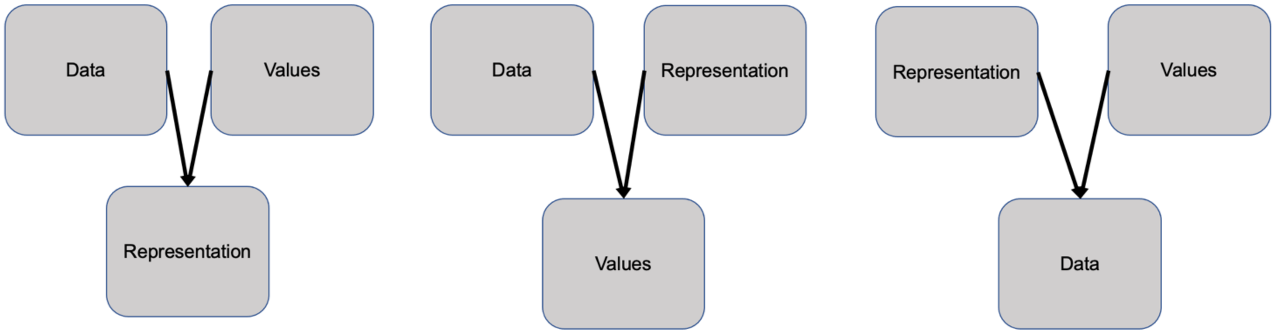
Introduction

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." His point was that an agent's choices about how to spend their resources often provide better evidence of that agent's preferences or desires than explicit claims about their preferences or desires. Just as governments have fiscal spending budgets, we as agents have *cognitive* budgets in our representations of our environment. That is, *what* we represent, and *how* we represent it, can have costs in terms of memory, processing, or other cognitive resources. So, while there is often value in being able to represent our environment at a very fine level of detail, such fine-grained representations also come at a cost – and thereby place a greater burden on our cognitive budgets.

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 information from the environment that they deem valuable to preserve), and the data they are representing. We report three experiments that support these relationships of mutual constraint, such that knowing any two of an agent's *representation* of the environment, their *values*, or their *data* about the environment places strong constraints upon the third (see Figure 1). We then argue that our results support a central role for the value of information 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 inferences we draw about the values of individuals who use particular causal representations.

Figure 1

Schematic representation of the inferences that we predict that participants in each of our experiments will draw. Figure 1a shows the pattern of inference wherein information about a speaker's values and the data available to the speaker allows a listener to predict the level of granularity at which the speaker will represent the causal structure of their environment. Figure 1b shows the pattern of inference wherein information about the level of granularity with which a speaker represents their environment and the data available to the speaker allows a listener to infer the speaker's values. Figure 1c shows the pattern of inference wherein information about the level of granularity with which a speaker represents their environment and the speaker's values allows a listener to infer the data available to the speaker.



(a)
Prediction 1, tested in Experiment 2

(b)
Prediction 2, tested in Experiment 3

(c)
Prediction 3, tested in Experiment 4

The representations that we consider are generic causal claims about the social world. To illustrate, consider the data presented in Table 1. While these data are correlational, they are consistent with several causal relationships that could obtain between the type of school a student attends and their math performance. Two such relationships, which vary in their level of granularity, are expressed by the following claims:

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.

Table 1

Hypothetical data about math performance in a large school district.

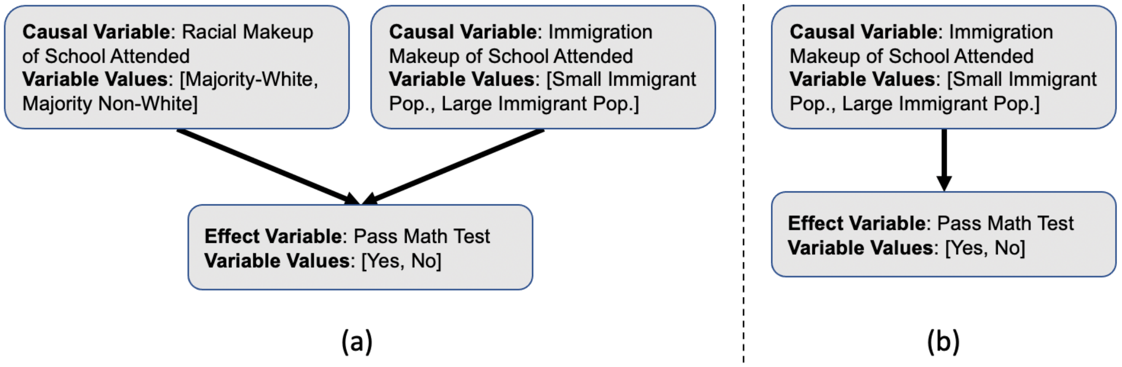
	% of Students who Pass the Math Test
Majority-White Schools with Small Recent Immigrant Population	60
Majority-Non-White Schools with Small Recent Immigrant Population	45
Majority-White Schools with Large Recent Immigrant Population	25
Majority-Non-White Schools with Large Recent Immigrant Population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

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 – in this sense it has a higher cognitive cost. We take an utterance of COMPLEX to be indicative of an agent representing the social dynamics of school math performance using the causal model shown in Figure 2a, whereas an utterance of SIMPLE is indicative of an agent representing the social dynamics of school math performance using the causal model shown in Figure 2b.

We can illustrate our proposed dynamic between representations, values, and data in the context of this example. Suppose that two teachers comment on the same report of students' math performance in a particular school district, corresponding to the data presented in Table 1. The first teacher utters COMPLEX, and the second teacher utters SIMPLE. Both teachers' claims are consistent with the data, but they differ in their level of granularity. In particular, SIMPLE

Figure 2

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


Causal Variable: Racial Makeup of School Attended
Variable Values: [Majority-White, Majority Non-White]
Causal Variable: Immigration Makeup of School Attended
Variable Values: [Small Immigrant Pop., Large Immigrant Pop.]
Effect Variable: Pass Math Test
Variable Values: [Yes, No]
(a)
Causal Variable: Immigration Makeup of School Attended
Variable Values: [Small Immigrant Pop., Large Immigrant Pop.]
Effect Variable: Pass Math Test
Variable Values: [Yes, No]
(b)

collapses across "majority-white schools with small recent immigrant populations" and "majority-non-white schools with small recent immigrant populations," creating the single factor "schools with small recent immigrant populations." In so doing it omits the representation of the school's racial composition entirely. We might correspondingly infer that for this teacher, the information lost in collapsing across racial composition is of low value. Perhaps this teacher thinks a passing rate of over 30% is adequate, such that immigrant status alone should guide decisions, and that tracking a school's racial composition is irrelevant. By contrast, the teacher who utters COMPLEX is willing to incur some cognitive cost to preserve information about racial composition, and so we might infer that for this teacher, information about racial composition is of higher value. Perhaps this teacher wants to see schools of all demographics reach a threshold of over 50%, or perhaps this teacher would approach decisions about appropriate interventions differently depending on the racial composition of a given school. In this way, the teachers' causal representations – reflected in a generic claim about what causes better math performance (COMPLEX or SIMPLE) – reveal something about their values. In the next section, we make the line of reasoning supporting this intuition more precise, and we explain how similar lines of reasoning support the other inferences depicted in Figure 1. In so doing, we put forward a novel theoretical framework explaining the connections between the data available to an agent, the granularity of their causal representation of the environment, and their values.

The remainder of this paper proceeds as follows. First, we summarize the formal framework on which our work is based, drawing heavily on Kinney and Lombrozo (2023), and offering a qualitative derivation of our three predictions (Figure 1). Next, we discuss the connections between our proposal and previous work on the psychology and philosophy of causal cognition, especially with respect to causal modeling of the social world. We then present findings from four experiments that are in keeping with our predictions. We conclude by discussing the significance of our findings for the value judgments that we make about speakers who utter certain generic causal claims about their social world, and for the psychology and ontology of social categories.

Framework for Deriving Key Predictions

At what level of granularity should an agent represent the causal structure of their social world? For example, should an agent represent schools differently depending on the racial composition of their student population? 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, 2016b), with decisions about variable granularity regarded as an instance of this more general problem. Kinney (2019) and Kinney and Watson (2020) provide a framework showing how this problem can be resolved, at least in part, by supposing that agents seek causal models that maximize simplicity (and thus reduce cognitive cost), but while retaining all of the information that is valuable to them. In line with these theoretical models, Kinney and Lombrozo (2022) offer empirical evidence that all else being equal, people have a preference for more compressed causal models (so, for example, favoring the model in Figure 2b over that in Figure 2a, provided that the models both retain all valuable information). Kinney and Lombrozo (2023) further develop and test these basic ideas, arguing that an agent’s propensity to encode a more coarse-grained representation over a more fine-grained representation is: i) an increasing function of how much more costly it would be to encode the more fine-grained representation, and ii) a decreasing function of the value of the information lost in the move from the more fine-grained to more coarse-grained representation. In schematic form:

Propensity to Encode Coarse-Grained Representation ∝ Cost of Encoding a More Fine-Grained Representation – Value of the Information Lost in Compression.

These ideas can be cashed out quantitatively, and Kinney and Lombrozo (2023) in fact provide a numerical measure of the value of the information lost in a compression of a causal model. However, to derive the predictions of the current paper, it only matters that this measure is always non-negative. Intuitively, the idea is that information is valuable to an agent to the extent that it informs that agent’s decisions, given their utilities and the context in which they find themselves. But because encoding more information has some cost (for example, in terms of increased demands on memory or on processing), agents need to engage in a process of “cognitive budgeting” that balances this cost against the value of the information contained in a more fine-grained representation.

To make this concrete, consider the case in which a speaker is choosing whether to summarize the data in Table 1 using either SIMPLE or COMPLEX. We assume that there is a cost to encoding a representation consistent with the more fine-grained claim COMPLEX, such that, all else being equal, the speaker is more likely to represent and utter the claim SIMPLE. However, all else may not be equal. In particular, the information preserved in the causal model corresponding to COMPLEX could have some value to the speaker. For example, suppose the speaker believes that in order for a school to be classified as providing adequate math education, it must be the case that at least 50% of its students pass the math test. We will call this HIGH THRESHOLD:

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

Moreover, suppose that this threshold governs how the speaker thinks resources should be allocated, such that knowing whether a school is above or below the threshold is relevant to their decision-making, but knowing (for example) by how much a given school exceeds the threshold is *not*. For this speaker, the information lost in the compression from COMPLEX to SIMPLE comes at a cost: it obscures the difference between majority-white schools with small recent immigrant populations (which meet the threshold) and majority-non-white schools with small recent immigrant populations (which do not).

By contrast, consider a speaker who instead endorses a threshold of 30%:

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

Once again, assume that this threshold captures differences in how an agent might value the information found in data, such that for the speaker who endorses LOW THRESHOLD, the only information relevant to their decision making is whether a school meets or falls short of the 30% threshold. For this speaker, the information lost in the move from COMPLEX to SIMPLE has no value: schools with a small immigrant population are expected to meet the threshold regardless of whether they are majority-white or majority-non-white.

We refer to the thresholds reflected in HIGH THRESHOLD and LOW THRESHOLD as "value-of-information thresholds," as they dictate the information that is valuable to preserve from the perspective of an agent in a given context. Given these thresholds and the assumption that encoding COMPLEX over SIMPLE has some non-negligible cognitive cost (and moreover, one that does not itself depend on the threshold), we can derive predictions about how our hypothetical speakers will tend to represent the causal structure responsible for the data in Table 1. Specifically, we should expect the speaker who endorses HIGH THRESHOLD to be more likely than the speaker who endorses LOW THRESHOLD to incur the extra cost of a fine-grained causal representation, and therefore to be the speaker more likely to utter COMPLEX. This line of reasoning supports Prediction 1 (see Figure 1).

Using a similar logic, we can derive Prediction 2: that listeners will make differential judgments about the values of a speaker (e.g., whether the speaker endorses LOW THRESHOLD versus HIGH THRESHOLD) depending on the speaker's causal representation (e.g., whether they utter SIMPLE versus COMPLEX). Specifically, if we assume that two speakers incur a similar cost when encoding a less-compressed causal representation, then we derive that the speaker who utters SIMPLE is likely to assign less value to the information lost in compression than the speaker who utters COMPLEX. In the context of our example, this difference in how a speaker values information corresponds to a difference in the threshold that the speaker endorses. Listeners should therefore judge that a speaker who utters COMPLEX is more likely to endorse HIGH THRESHOLD (vs. LOW THRESHOLD), and that a speaker who utters SIMPLE is more likely to endorse LOW THRESHOLD (vs. HIGH THRESHOLD).

Finally, suppose that a speaker utters a generic causal claim at a particular level of granularity (e.g., they utter SIMPLE or COMPLEX in our running example), and the speaker is known by the listener to adopt either LOW THRESHOLD or HIGH THRESHOLD. However, it is unknown to the listener what data the speaker has access to. Using the schema above, the listener can determine the value of the information lost in coarse-graining the claim COMPLEX into the claim SIMPLE given a candidate data set that the speaker might have access to. Given multiple candidate data sets, the listener can evaluate which data set would minimize the value of information lost in opting for SIMPLE over COMPLEX. To illustrate, a speaker who utters SIMPLE while endorsing HIGH THRESHOLD will be judged by a listener as most likely to have had access to data such that all schools with a low immigrant population, regardless of their racial makeup, are likely to perform above the 50% threshold. More generally, we predict that when told that a speaker who endorses either LOW THRESHOLD or HIGH THRESHOLD utters either SIMPLE or COMPLEX, participants will judge that speaker to have been more likely to have had access to a data set that maximizes the speaker's propensity to produce an utterance at the level of their chosen level of granularity, as opposed to data sets that would make their chosen level of granularity less likely. This is Prediction 3.

While the current work builds on the framework and findings from Kinney and Lombrozo (2023), it goes beyond that work in three ways. First, and most significantly, that paper only considers and tests Prediction 1; here we also consider and test Predictions 2 and 3. The latter predictions are especially important in the context of social judgments, as people routinely draw inferences about others on the basis of the representations they share in the form of linguistic utterances. Second, while that paper elicited participants' judgments about the optimal granularity of causal variables in hypothetical workplace or gameplaying scenarios, our experiments apply their framework to cases involving realistic social categories, such as gender and race. This allows us to address the important question of why people choose to use more fine-grained or coarse-grained categorization schemes for other people or objects in their social world (for instance, choosing to make a claim about *male* children versus children, or *East-Asian* immigrants rather than immigrants). Third, while that paper's experiments measured participants' own preferences as to the granularity with which they represent the causal structure of their environment, we measure participants' judgements about the granularity with which *other speakers* are likely to represent the causal structure of their environment. While we generate the same predictions for the case of self and other, testing inferences about others is especially relevant to accounts of social judgment.

Prior Work on Causal Representations of Social Kinds

While there has been relatively little work applying formal models of variable selection to the social domain, there is a long tradition within social cognition of taking the categorization schemes that agents use to understand their social world to be closely connected to how those same agents represent the causal structure of that world. Campbell (1958) uses the term 'entativity' to refer to the extent to which a group of people all belonging to a social category can be understood as a unified entity. He argues that a group's entativity is closely correlated with the degree to which group members are taken to share a "common fate" across time. The idea that members of a group have a common fate is often justified on causal grounds; individuals in the same group share a common fate because they are subject to the same causal dynamics.[1] The current paper contributes to this discussion by identifying a role for the value of information in explaining why agents take certain social variables to form part of their causal model of their environment.

In more contemporary work on social cognition, discussions of entativity have informed further theorizing about the role of generic language in communication about social categories. As mentioned above, causal claims involving social categories are often generic claims, such as "attending a majority-white school improves math performance." The role of causal generics in

[1] It is worth noting that the causal dynamics believed to unite members of social groups need not stem from a common internal cause or "essence." The inductive potential of group membership can instead be underwritten by stable structural constraints (Author, 2020) and social roles or institutions (Noyes et al., 2021).

thinking about the reality of social groupings comes to the fore in work on psychological essentialism (Gelman, 2003; Haslam et al., 2000; Rhodes et al., 2012) 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). The current paper contributes to this discussion by identifying a further communicative role for causal generics involving social categories: a role in conveying the speaker's values, by which we mean not just their moral values, but the values that more broadly govern their decisions.

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; for other interpretations of generics, see also Noyes and Keil, 2019; Prasada et al., 2013; Vasilyeva and Lombrozo, 2020). Our results provide evidence of a role for generics in communicating other information of a very particular sort. Because generics reflect an agent's choices in allocating their "representational budget," they also convey which social categories the agent regards as containing valuable information. By "valuable information," we mean information that should inform an agent's decisions, given their particular evaluation of possible states of affairs and their theory of rational choice. Thus, our results here highlight a previously unexplored connection between generic causal language and the variety of instrumental rationality modeled in decision theory.

In more computational work on generics, Tessler and Goodman (2019) hold that listeners accept a causal claim '*c* causes *e*' as true (and that speakers expect listeners to accept a causal claim as true) when the probability of an instance of event-type *c* causing an instance of event-type *e* is sufficiently high, where what counts as 'sufficiently high' is context-dependent. We are not concerned here with the truth of causal claims, but rather with how an agent might choose a causal representation from a set of causal claims that are all true, but which differ with respect to their granularity. Nevertheless, our approach coheres with Tessler and Goodman's in that we take a causal claim to be a more apt utterance when a particular probability (in our case, the probability of the cause given the effect) is above a certain context-sensitive threshold. In our framework, the relevant context fixing this threshold is determined by the data available to an agent, and how that context shapes the ways in which agents value information.

Overview of Experiments

In sum, prior work on variable selection suggests that people balance an all-else-being-equal preference for simpler or more compressed representations with a desire to preserve valuable information. This connection between representational choices and the value of information explains how an agent's representational "budget" can reveal their values. Across three experiments (Experiments 2-4), we test the predictions depicted in Figure 1: i) that listeners will reliably identify a speaker with access to a given data set to be more or less likely to represent that data in a compressed way depending on their values (Experiment 2), ii) that listeners will reliably infer that a speaker with access to a given data set has different values depending on the level of granularity with which they represent that data set (Experiment 3), and iii) that listeners will reliably identify a speaker as more likely to have had access to a particular data set depending on both the speaker's values and the level of granularity with which they represent the causal structure of their environment (Experiment 4).

Before testing these predictions, we present an initial experiment, Experiment 1, that verifies two crucial presuppositions that underwrite the predictions and interpretations for Experiments 2-4: i) that people do in fact operate with a cognitive budget (such that they trade off an all-things-considered preference for more coarse-grained causal summaries against a desire to preserve informative cause-effect relationships when summarizing data), and ii) that causal generic claims are regarded as apt summaries of a data set like that shown in Table 1.

The results of these four experiments go beyond prior work on variable choice both in their application to a social domain and in the systematic investigation of all three inferences supported by the posited relationships of mutual constraint between representations, values, and data. This work also goes beyond prior research on social categories and generic language in offering a new way to understand the inferences about an agent that might follow from their choice of generic causal representation. Finally, our results augment prior work on generics by offering an interpretation of the thresholds that govern generic expression and interpretation in Tessler and Goodman's framework in terms of the value of the information provided by generic summaries of data.

Experiment 1

The aim of Experiment 1 was to verify two presuppositions that inform our subsequent experiments. First, our framework assumes the existence of a cognitive budgeting mechanism such that participants should favor more coarse-grained causal summaries of data over more fine-grained summaries, all else being equal. This is because more fine-grained representations incur some cost that must be traded off against the value of information. The cost could take the form of additional demands on memory, processing, time, effort, or any other cognitive resource. Our framework is agnostic with respect to the implementation details that generate this cost, but does require there to be some cost that trades off against information. In Experiment 1, we test two predictions that follow from this assumption: that participants will favor a more coarse-grained summary of data over a more fine-grained summary when the former involves no information loss, and that this preference will be attenuated as information loss increases.

The second presupposition of our subsequent experiments concerns the use of generic causal claims. In our introductory example involving SIMPLE and COMPLEX, we used such claims to express causal representations of the process generating a set of data. However, it is possible that participants would resist a causal interpretation of correlational data, or that they would more naturally express causal assumptions in the form of quantified rather than generic claims. To test whether participants would regard the causal and generic language used to summarize the data sets in Experiments 2-4 as apt summaries of the data, Experiment 1 had participants generate their own summaries of data and evaluate alternative causal interpretations, as we explain below. We predicted that participants would frequently generate generic claims and endorse causal interpretations whereby the factors used to aggregate data were regarded as causes of the effect variable.

To test these predictions, we presented participants with a hypothetical scenario in the style of the example from the introduction (in particular, one scenario involved the performance of local schools on a math test). After seeing data relevant to their assigned scenario, participants were asked, without access to the data, to complete two tasks in counter-balanced order. The first involved choosing the best causal graph to represent the data-generating process, with different options corresponding to different levels of granularity for such a representation. We also asked participants to generate their own verbal summary of the data, which we subsequently coded for the granularity with which they describe the data and the use of generic language. By varying the data that participants were presented, we were able to vary whether the information lost in moving from a more fine-grained representation to a more coarse-grained representation was zero, low, or high. These tasks thus allowed us to assess whether participants' responses involved a trade-off between compression (favoring coarse-grained representations) and informativeness (favoring fine-grained representations), and also whether participants spontaneously produced generic language.

In an additional task, participants were asked to select between three different possible augmentations of the causal graph they initially selected. These augmented graphs either maintained or eliminated the causal connections between social categories and social outcomes represented in the originally chosen graph, such that a participant's choice would indicate whether they regarded the correlations shown in the data as due to a causal pathway between categories and outcomes, or due to mere association. This final task allowed us to assess participants' causal interpretation of the correlational data with which they were presented.

Methods

Participants

Participants were 469 adults recruited via Prolific. No participants were excluded for failing comprehension checks, but one participant was excluded for completing the survey in less than one minute (following pre-registered criteria for exclusion). 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. Data, analysis code, pre-registrations and stimuli can be found here: https://osf.io/a65vs/?view_only=7b12ee0785174deba06bbd0d809da6d5.

Materials and Procedures

Participants read about a novel social system and were given data about that system. Our vignettes involved the performance of local schools on a math test, the rates at which people in different neighborhoods in a large city own a bicycle (where the data were presented as a function of neighborhood affluence and the prevalence of families with children), or the performance of local children in swimming classes (where the data were presented as a function of the gender composition of swimming classes and whether children in the class attended public or private schools). The vignette shown was varied *between* participants, such that each participant saw only one vignette.

In all cases, social category labels were applied to social institutions, rather than directly to people (e.g., "majority-white schools" or "affluent neighborhoods," vs. white children and affluent families). We did so out of concern that some participants would resist making generalizations over individuals on the basis of racial or gender categories, thereby introducing a systematic bias in their preferred representations. Nevertheless, our experiments still concern an important way in which social category labels are often used in causal language.

To illustrate the structure of our stimuli, 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 were then told that the data in Table 2 describe the percentages of children from different schools who are placed into Algebra 1 on the basis of their performance on the math test.

The percentage of students who passed the math test in majority-non-white schools with a small recent immigrant population was manipulated between participants, and set to either 45, 55, or 60.

Table 2

Hypothetical data about math performance in a large school district.

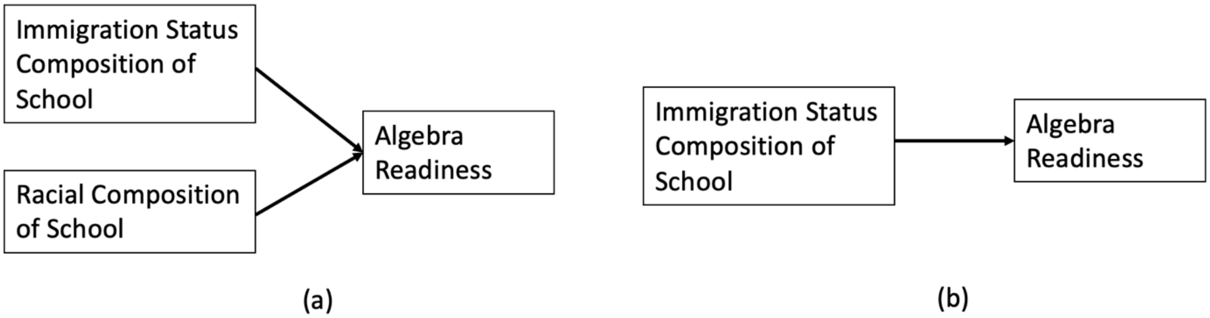
	% of Students who Pass the Math Test
Majority-White Schools with Small Recent Immigrant Population	60
Majority-Non-White Schools with Small Recent Immigrant Population	[45/55/60]
Majority-White Schools with Large Recent Immigrant Population	10
Majority-Non-White Schools with Large Recent Immigrant Population	10

When this manipulated percentage was set to 60%, giving a more coarse-grained summary of the data that only mentions the immigration status composition of schools while ignoring the racial composition of schools would not result in any loss of information about the factors associated with passing the math test.[2] In this condition, specifying the racial composition of a school does not provide any information about a student's likelihood of being placed in Algebra 1. As the manipulated percentage is decreased to 55 and then 45, the amount of information about Algebra 1 that is lost when giving a coarse-grained summary of the data that ignores the racial composition of schools increases. Assuming that the *value* of the information lost in coarse-graining varies monotonically and positively with the *amount* of information lost in coarse-graining, our cognitive budgeting framework predicts that as the percentage is increased between

[2] For these statements, one might worry that whether a school is majority white or not is in part a consequence of recent immigration (if the immigrant population is assumed to be non-white, for example). This was not a feature of our other items, and does not seem to have changed any qualitative features of the data. See Footnote 4 below for more details on this point.

Figure 3

Initial graph Selection task in Experiment 1.


Immigration Status Composition of School
Racial Composition of School
Algebra Readiness
(a)
Immigration Status Composition of School
Algebra Readiness
(b)

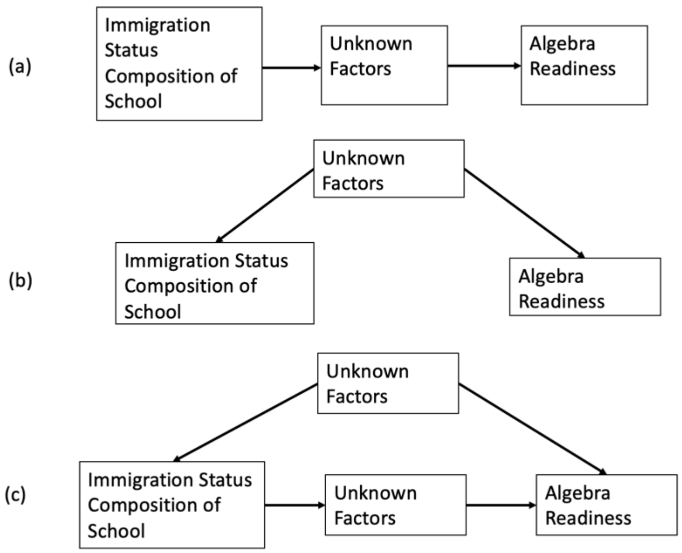
participants from 45 to 60, we will see more participants give coarse-grained summaries of the data.

To test this, we asked participants to navigate to a screen where they no longer had access to the data. They were then told that a teacher had been tasked with giving a summary of the data, and needed to choose, between the two graphs in Figure 3, which one "best represents their understanding of the factors affecting algebra placement." Choosing Figure 3a would indicate that a participant expected the teacher to choose a more fine-grained representation, while choosing Figure 3b would indicate that a participant expected the teacher to choose a more coarse-grained representation. Next, participants were told that the same teacher had been asked to write a 1-2 sentence summary of the factors affecting algebra placement. Participants were given a text block in which to write what they expected the teacher would produce (they were required to use at least 50 characters in their open-ended summary). The order of these two tasks (i.e., the graph selection task and the open-ended response task) was counterbalanced between participants.

The final task tested how participants would choose to revise their initially selected graph when given the opportunity to do so. Specifically, we wanted to give participants the opportunity to clarify whether they saw the correlation between nodes in the graph they initially selected as due to a causal mechanism connecting the putative causal node(s) to the putative effect node, or if

Figure 4

Choices between augmented causal graphs in Experiment 1.


(a)
Immigration Status Composition of School
Unknown Factors
Algebra Readiness
(b)
Unknown Factors
Immigration Status Composition of School
Algebra Readiness
(c)
Unknown Factors
Immigration Status Composition of School
Unknown Factors
Algebra Readiness

they would choose to represent the putative cause(s) and effects as correlated due to a non-causal association. To test this, we told participants that the same teacher who used the graph that they first chose now had the opportunity to choose a second causal graph that included nodes representing unknown factors. For example, a participant who first chose the graph in Figure 3b was presented with the option to choose one of the three graphs in Figure 4 as one that the teacher would choose, if given the chance to augment their original choice. Choosing Figure 4b would indicate that the participant took the teacher to endorse the claim that the correlation observed in the data table is due to a spurious correlation between the immigration status of the school and algebra placement. By contrast, choosing Figure 4a or 4c would indicate the existence of a causal mechanism at least partially generating the observed correlations. Thus, higher-than expected rates of choosing Figure 4a or 4c would indicate that participants tended to treat a causal summary of the data as apt. Note that the actual graphs shown in this task differed between participants depending on which graph they chose during the first graph selection task.

As an attention check for exclusion, participants also answered an easy question about their vignette. For example, participants shown the vignette described above were asked the multiple-choice question: 'Which of the following topics were the previous questions about?' The correct answer, in this case, was 'Algebra placement.'

Response Coding

Participants' open-ended responses were coded both for their level of granularity and their use of generic language. Granularity was coded in two ways. On the "liberal encoding," participants' responses were coded as coarse-grained if they either: i) mentioned only the immigration composition of a school (or the analogous factor in other vignettes) and not the racial composition of the school (or the analogous factor in other vignettes) (e.g., "A larger portion of students were ready for algebra 1 in schools where the recent immigrant population was lower"), or ii) mentioned racial composition only to say that this factor is not relevant to the effect or otherwise causally inert (e.g., "It appears that a little over half of the students in a school without a recent immigrant population get placed in the higher math course, while a school with a recent immigrant population has a much higher percentage going into the lower math course. This percentage doesn't seem to be affected by if the school has a majority of white or non-white students"). Participants' responses were coded as "null" if they were off-topic or not understandable. Otherwise, participants' responses were coded as fine-grained. The "conservative coding" included only criterion (i) above, such that mentioning the racial composition of the school (or the analogous factor in other vignettes) in any way at all resulted in a response being coded as "fine-grained." For both codings, conjunctive constructions like "majority-white and majority non-white schools with small recent immigrant populations had higher rates of Algebra 1 placement" were counted as *not* giving coarse-grained summaries, since they could be plausibly read as making fine-grained, conjunctive causal claims. All coding was done by two independent coders, with disagreements resolved via discussion. Cronbach's alpha for the liberal and conservative encodings was .89 (CI: [.87, .91]) and .91 (CI: [.89, .92]), respectively.

To code responses for the use of generics, we adopted the following guidelines for coding noun phrases (NPs) as generics, due to Gelman et al. (2008):

> Generics were defined as NPs that refer to general categories and are not tied to a particular situation or point in time. They were identified by a combination of morphological, syntactic, semantic, and pragmatic cues. For example, generics could not be examples of particular individuals or instances, and so numbers, pronouns, the word "some," and the word "the" were used as indications that an NP was not generic. They also usually could not be in sentences in the past or future tense or in the progressive form (p. 9).

A written response was coded as generic if and only if all its NPs that referenced any of the putative causes in the data set (e.g., immigration composition or racial composition of a school) were generics (e.g., "Schools with a heavy immigrant population have very low scores"). Cronbach's alpha between the two coders for this task was .78 (CI: [.74, .82]).

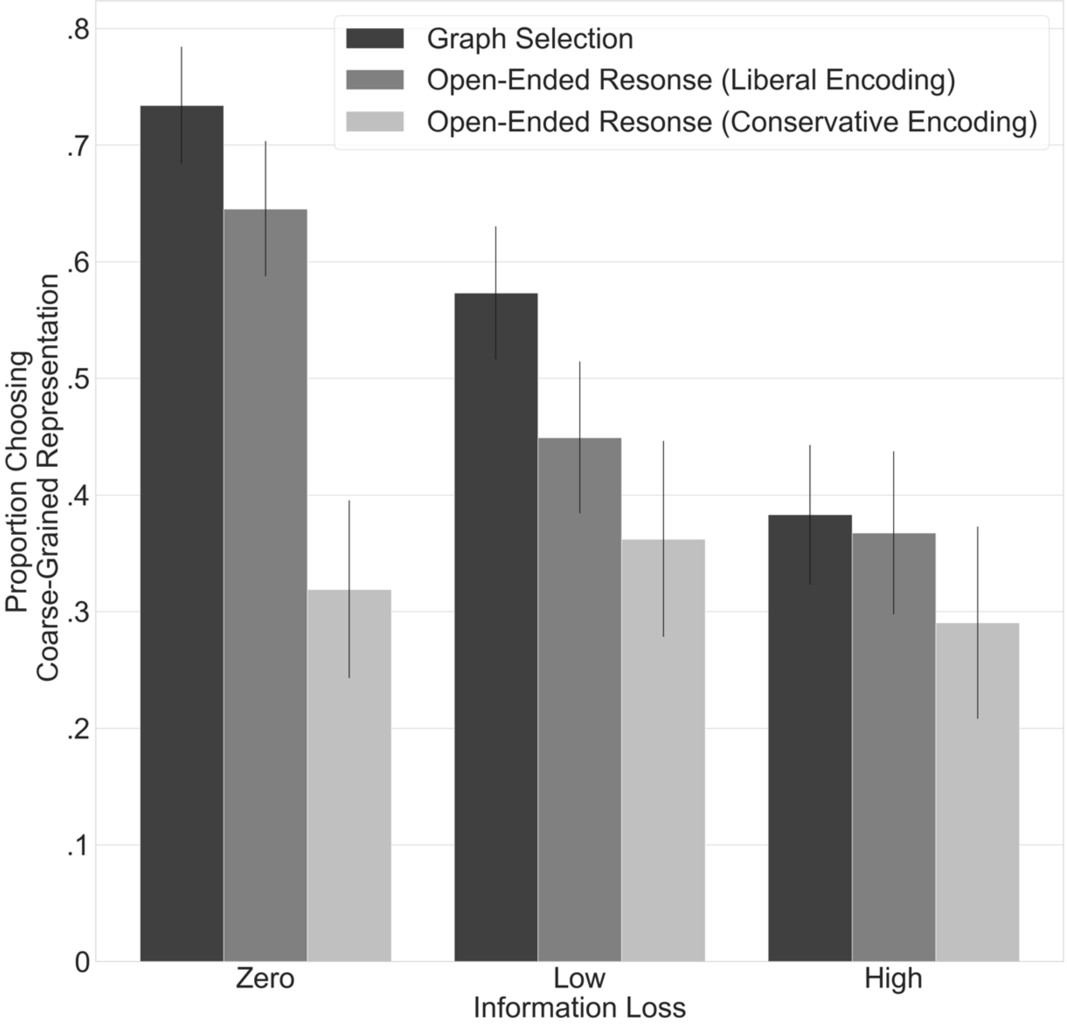
Results

Figure 5 displays the proportion of participants who selected or generated a coarse-grained representation at each level of information loss (zero, low, high), and for each of our three measures: the causal selection task, the liberal coding of the open response, and the conservative coding of the open response. In each case we predicted that coarse-graining would be frequent when information loss was zero, but that it would become less frequent as information loss increased.

We analyzed the graph selection task with a binary logistic regression that regressed the binary variable representing whether or not a participant chose the coarse-grained graph against: 1) the amount of information lost in compression (coded ordinally), 2) the vignette shown to participants, 3) the order in which the graph selection task and the open-ended response task were

Figure 5

Proportion of participants coarse-graining in each causal summary task in Experiment 1.



shown to participants, and 4) all interactions between all independent variables. This analysis found only a main effect of the amount of information lost in compression. As predicted, participants were more likely to choose the coarse-grained graph when less information was lost in compression ($\beta = -0.766, p < .001$). Moreover, when information loss was zero, participants selected the course-grained graph more often than chance ($p = 3.29 \times 10^{-9}$).

We analyzed the open-ended response task with binary logistic regressions predicting a coarse-grained coding from the same independent variables as the previous regression. For the liberal coding, we found a main effect of information loss, as predicted, with participants more likely to choose the coarse-grained graph when less information was lost in compression ($\beta = -0.667, p < .001$). This regression also found an interaction between the amount of information lost and the order in which the graph selection task and the open-ended response task were presented: information loss had a larger effect on the granularity of participants' open-ended responses when those responses were given *after* selecting a graph ($\beta = 0.269, p = .038$), but even when open-ended responses were elicited first, information loss had the predicted effect on coarse-graining, ($\beta = -.398, p = .027$).[3]

The same regression for the conservative encoding scheme found only a main effect of the vignette shown to participants, with participants most likely to coarse-grain when viewing the algebra placement vignette ($\beta = 0.313, p = .020$).[4] This coding scheme, which was decided *a priori*, does not seem to have captured much meaningful variation in participants' open-ended responses, plausibly because the demands of the task discouraged participants from omitting information altogether. This may seem to undermine the idea that participants were budgeting the

[3] In an exploratory analysis to check whether there was a relationship between the amount of time participants spent on the task and the granularity of the representations they selected, we regressed the duration of time (in seconds) that participants took to complete the task against both the granularity of the graph they selected and the granularity of their verbal summary of the data on the liberal encoding. We removed participants whose time spent in the survey was above the 99th percentile and below the 1st percentile. In neither case did we find a significant relationship (Graph Granularity: $\beta = 5.40, p = .700$; Verbal Summary Granularity: ($\beta = -24.60, p = .094$). We thank an anonymous reviewer for suggesting that we investigate this relationship.

[4] This result, while not predicted, is nevertheless interesting in the context of our study. It is reasonable to assume that many participants would hold the prior belief that recent immigrants to a school district would tend to be non-white. However, this result shows that this prior belief, if it exists, is not driving participants to mention the non-operative causal factor in this case. Indeed, participants were significantly more likely to completely elide mention of the secondary causal factor, so that their open-ended response would be coded as coarse-grained on the conservative coding scheme, in the algebra placement vignette than in other vignettes. In the other vignettes, the two causal factors were: 1) the prevalence of families with children in a neighborhood and the income level of the neighborhood, and 2) the majority gender of students in a swimming class and whether students in the swimming class came from public or private schools. Intuitively, an interaction effect between factors in either of these two cases is less likely, and yet participants are more likely to mention the secondary factor in these vignettes than they are involving race and immigration status. This shows that, at least in our sample, a prior belief in a high likelihood of interaction effects is not driving participants to mention surprising null effects more often.

granularity of their representations, since even in cases where coarse-graining would result in no or little information loss, many participants seemingly maintained more complex representations that tracked null effects. However, even though responses on the liberal coding could mention additional causal factors, and in that sense represented them, they are still less demanding in terms of cognitive resources in that they would not require separately tracking or reasoning about the factor stipulated not to matter. We also found significant correlations between the granularity of the selected graph and the granularity of the open-ended response for both the liberal encoding ($R^2 = .474, p = 1.84 \times 10^{-24}$) and the conservative encoding ($R^2 = .262, p = 6.79 \times 10^{-7}$). This result is especially important in the case of the liberal encoding, since it shows that even participants who explicitly mentioned null effects tended to see this as consistent with a more coarse-grained causal model of the relevant system, in keeping with our broader hypothesis of cognitive budgeting.[5] This is true even when we restrict our analysis of the liberal encoding to just those cases in which information loss is non-zero: here too we see a correlation between the granularity of the encoding received for an open-ended response and the granularity graph selected by the participant for both the liberal encoding ($R^2 = .489, p = 9.96 \times 10^{-18}$) and the conservative encoding ($R^2 = .396, p = 1.23 \times 10^{-11}$), suggesting that when we restrict our attention to just those cases where there is no notable null effect to comment on, we still see

[5] While it is noteworthy that we do not see a significant effect of information loss on coarse-graining when we consider the more conservative encoding of granularity, we take this result to nevertheless be in keeping with our proposal. Specifically, in the zero-information-loss condition, we see very high rates of coarse-graining on the liberal encoding as compared to the conservative encoding. One reason why this may be the case is that when the racial composition of a school makes *no* difference to algebra placement, this fact is worth explicitly commenting on in an open-ended response. By contrast, in the low-information-loss and high-information-loss condition, participants who *do* coarse-grain their representations seem to be more likely to elide any mention of race altogether, rather than falsely claim that there is no racial difference in algebra placement. Since any response that is deemed coarse-grained on the liberal encoding is also coarse-grained on the conservative encoding, but not vice-versa, overall rates of coarse-graining still decrease as information loss increases. However, a greater percentage of those coarse-grained responses are coded as such according both the liberal and the conservative criterion as information loss increases.

correlations between participants' open-ended descriptions of a system and their choice of graphical representation for that system.

Turning to coding for generic claims, we found that 55.3% of participants spontaneously used exclusively generic language to describe the social category labels in their open-ended responses. Examples of responses coded as generics include the following:

Schools with more recent immigrants tend to do worse in Algebra placement.

The less white and the more immigrants, the worse the placement.

This suggests that many participants found it natural to summarize data in the form of generic claims.

Finally, for the final causal graph selection task, we found that when asked to augment their originally chosen causal graph with nodes representing additional causal factors, only 11% of participants chose Figure 4b or an analogous graph for other vignettes ($p = 7.31 \times 10^{-29}$), indicating that the associations between variables shown in their data sets derived from a common cause rather than any direct causal relationships. The remaining 89% selected graphs consistent with a causal interpretation of the relationship between categories and outcomes.

The results of additional pre-registered analyses are reported in the Supplemental Materials.

Discussion

The results displayed in Figure 5 provide strong support for our theoretical framework in which an all-things-considered preference for representing and summarizing data in a coarse-grained way is budgeted against a desire to represent and describe informative relationships between causes and effects. In general, we saw high rates of coarse-graining when information loss was zero or low. These same rates of coarse-graining quickly decreased as information loss increased. This suggests, as hypothesized above, that participants are budgeting a preference for coarse-graining against a desire to preserve informative cause-effect relationships when representing a data-generating process.

Participants' verbal summaries of the data often included exclusively generic language (about half the time), suggesting that they found this to be a natural and appropriate way to express causal representations in this task. Moreover, when participants were given the opportunity to clarify whether, when selecting a graph from Figure 4, they would have preferred the option of a graph in which there were no causal paths between social categories and social outcomes, participants overwhelmingly chose *not* to select such a graph. These findings suggest that for many participants, generic causal claims like "attending a school with a small recent immigrant population improves performance on math tests" or "attending a majority-white school with a small recent immigrant population improves performance on math tests" would seem an apt summary of the data presented to them in this experiment.

Finally, we note that the significant correlation between the granularity of participants' choices in the graph selection task and the granularity of their open-ended summary of the data suggests that there is a close connection between the level of granularity with which a speaker summarized a data set and their judgment as to the best representation of the causal dynamics of the process generating that data set. Having established both the plausibility of our cognitive budgeting framework and the aptness of generic causal claims as summaries of data in these scenarios, we are now in a position to test the three predictions of our model regarding how agents will choose a level of granularity for generic causal claims about their social world. These predictions are tested below, in Experiments 2-4.

Experiment 2

Experiment 2 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 more likely to endorse causal representations (in this case, causal generic claims) that balance compression with the preservation of valuable information. To test this, we presented participants with one of the same three hypothetical scenarios used in Experiment 1, again manipulating the vignette between participants. In an augmentation of Experiment 1, we also manipulated whether a speaker who was asked to summarize this data set endorsed a 30% or 50% value-of-information 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 value-of-information threshold, they are more likely to accept simpler causal claims.

Methods

Participants

Participants were 289 adults recruited via Prolific. An additional 13 participants were excluded for failing comprehension checks, and one participant was excluded for completing the survey in less than 60 seconds (following pre-registered exclusion criteria).

Materials and Procedures

Participants once again read about a novel social system, and were given data about that system as well as information about a hypothetical speaker's value-of-information threshold. To illustrate, in one vignette participants were told, as in Experiment 1, that in a fictional county with a number of different schools serving different communities, seventh-grade students 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.

Unlike in Experiment 1, participants were told that the data in Table 1, rather than the data in Table 2, describe the percentages of children from different schools who are placed into Algebra 1 on the basis of their performance on the math test. Participants were then told the following:

> County administrators vary with respect to what they consider to be the minimum percentage of students placed into Algebra 1 for a school to be classified as providing adequate math education. Some believe it should be 30%, others believe it should be 50%. This classification matters because schools that are classified as providing adequate math education do better on school rankings, are perceived more favorably by the community, can attract new teachers more effectively, and even have a positive impact on local real estate values. Based on input from teachers, the county has decided to say that a school is delivering "adequate math education" for 7th graders if at least [30/50]% of students are placed in Algebra 1.

Participants were then asked to state whether four schools, each with a different racial and immigration-status composition and a certain percentage of students being placed in Algebra 1, would be classified as providing adequate math education according to the threshold adopted by the teacher. Participants who answered incorrectly were corrected. They then received the following prompt:

> The county is preparing a document to share with teachers and administrators summarizing what they've learned about student performance over the last 5 years. The document will summarize data concerning a variety of topics (math and language arts achievement, extracurricular involvement, etc.), with a short section focusing on performance on the 7th grade algebra readiness test. The county administrator preparing the document must select one, and only one, of the following three claims to include in the report. Which claim would be best to include?
>
> **SIMPLE:** Attending a school with a small recent immigrant population improves performance on math tests.
>
> **COMPLEX:** Attending a majority-white school with a small recent immigrant population improves performance on math tests.
>
> **CONTROL:** Attending a majority-non-white school with a small recent immigrant population improves performance on math tests.

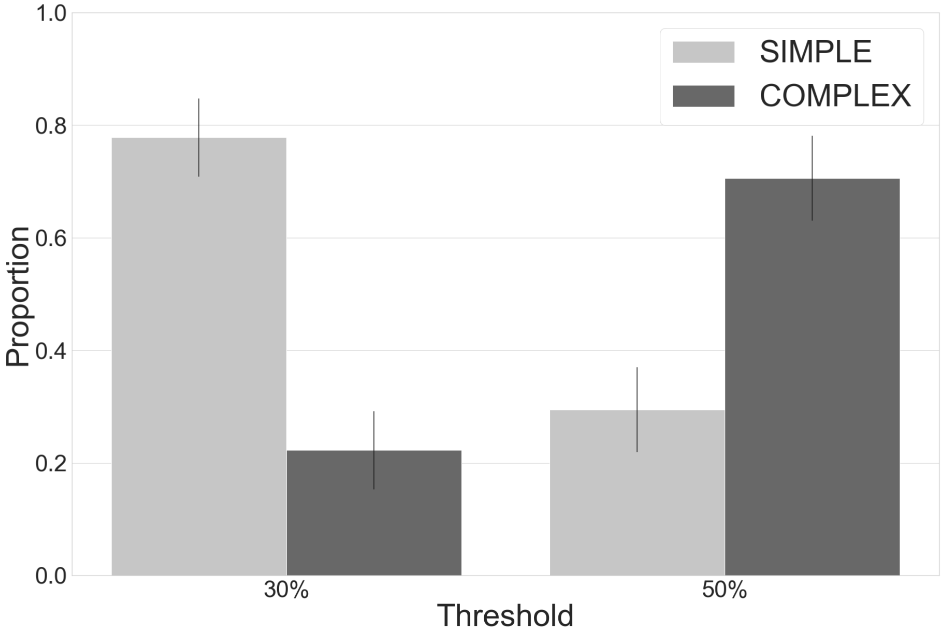
Participants' choice of claim was the dependent variable in the study. Recall that in the data given to participants, 60% of students in majority-white schools with a small recent immigrant population are placed into Algebra 1, as compared to 45% of students in majority-non-white schools with a small 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 so excluded participants who made this choice.

Results

Figure 6 shows the proportion of participants who chose each of SIMPLE and COMPLEX as the most likely claim made by the speaker for both of the value-of-information thresholds that the speaker could endorse. As pre-registered, we performed a χ^2 contingency test, which 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 all in keeping with our prediction that listeners can draw systematic

Figure 6

Proportion of participants in Experiment 2 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.



inference about the generic causal claims that a speaker is more likely to utter based on the data available to that speaker and the speaker's value-of-information threshold. This result also amounts to a conceptual replication of Experiment 4 in Kinney and Lombrozo (2023), but applied to the utterances of others (versus participants' own summary of a causal system) and in a novel, more socially relevant domain.

Experiment 3

Experiment 3 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 endorse low value-of-information thresholds for generic causal claims to be more likely to utter simple, coarse-grained generic causal claims. To test this, we presented participants with the same hypothetical scenarios as in Experiment 2, with the amendment that instead of one speaker summarizing the data, two speakers offered summaries. One speaker's summary contained only a simpler generic causal claim, while the other contained only a more complex causal claim. Participants were then asked which speaker they felt was more likely to endorse a lower threshold, to test the hypothesis that listeners will judge the speaker who made the simpler claim as more likely to endorse the lower threshold.

Participants

Participants were 147 adults recruited via Prolific. An additional participant was excluded for failing comprehension checks, and two participants were excluded for completing the survey in less than 60 seconds (following our pre-registration).

Methods and Procedures

Participants read about a novel social system (the same systems as in Experiment 2), and were given data about that system as well as two hypothetical speakers' generic causal claims at varying levels of granularity. 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 2. 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] think is most likely to be the one who believes that the minimum percentage should be 30%"; their answer to this binary question is our dependent variable. As an attention check for exclusion, participants also answered an easy question about their vignette.

Results

As predicted, a majority of participants (68.03%) identified the speaker who made the simpler generic causal claim as the one who endorsed the lower value-of-information threshold for generic causal claims in the relevant context. Assuming a null hypothesis in which participants are equally likely to identify either speaker as having a lower value-of-information threshold, this amounts to a p-value in a binomial significance test of $p = 1.47 \times 10^{-5}$. 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 value-of-information 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 value-of-information threshold for generic causal claims from the data available to that speaker along with the level of granularity at which the speaker makes a generic causal claim. These findings go beyond Kinney and Lombrozo (2023) by showing that in addition to data and the value of information constraining the granularity of causal representations, the data available to a speaker and the granularity of their causal representations also constrain their values. They also go beyond prior work on generic claims about social categories by establishing an explicit connection between said claims and the implicit values of the speaker.

Experiment 4

Experiment 2 showed that given both: i) the data generated by a social system, and ii) speakers' value-of-information threshold, listeners can make inferences about the generic causal claims that speaker is most likely to utter. Experiment 3 showed that listeners who are given both the data generated by a social system and the level of granularity with which a speaker described that system can infer that speaker's relevant value-of-information threshold. The patterns of inference seen in both of these experiments are consistent with our formal theory of the relationship between data, values, and the granularity of causal claims. In Experiment 4, we aim to complete this inferential circle, and show that when listeners are given both: i) a generic causal summary of unseen data, and ii) a speaker's relevant value-of-information threshold, listeners can make inferences about the data that the speaker is most likely to be summarizing. To this end, we ran an experiment in which participants were told about the same hypothetical social systems as in Experiments 2 and 3, and were told both the generic causal claim summarizing that data made by a speaker and that speaker's value-of-information threshold. The level of granular detail at which the generic causal claim was made, as well as the speaker's value-of-information threshold, were varied between participants. Participants were then asked to choose, on the basis of the information they were provided, which of three data sets that could have been generated by the social system in question was most likely to have been actually generated. Their answer to this question is our dependent variable.

Participants

Participants were 416 adults recruited via Prolific. An additional two participants were excluded for failing comprehension checks, with two also excluded for completing the survey in less than 60 seconds (as pre-registered).

Methods and Procedures

Participants read about a novel social system (the same systems as in Experiment 2), 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 value-of-information threshold. To illustrate, in one vignette participants were given the same scenario about Algebra 1 placement in a fictional county school district as in Experiment 2, and told either that the speaker uttered the claim SIMPLE or that the speaker uttered the claim COMPLEX, and either that the speaker believed that the threshold for adequate math education should be 30% or that the speaker believed that the threshold should be 50%. Participants were then asked, of the Data Sets shown in Table 3, which one they "believe contains the data that the teacher based their report on."

Table 3

Possible data sets that participants were told speakers might be viewing in Experiment 4.

Data Set 1		**Data Set 2**		**Data Set 3**	
	% of Students Placed in Algebra 1		**% of Students Placed in Algebra 1**		**% of Students Placed in Algebra 1**
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	25
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

Table 4

Predictions for Experiment 4

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

Table 4 summarizes the predicted responses across conditions, which we explain briefly in what follows. 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 value-of-information threshold is set to 30% or 50%. However, if the speaker's value-of-information threshold is 50%, then Data Set 1 also renders an utterance of COMPLEX more likely than an utterance of SIMPLE. By contrast, the causal claim SIMPLE 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 value-of-information 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.

As in Experiment 2, participants also answered an easy question about their vignette as an attention check for exclusion.

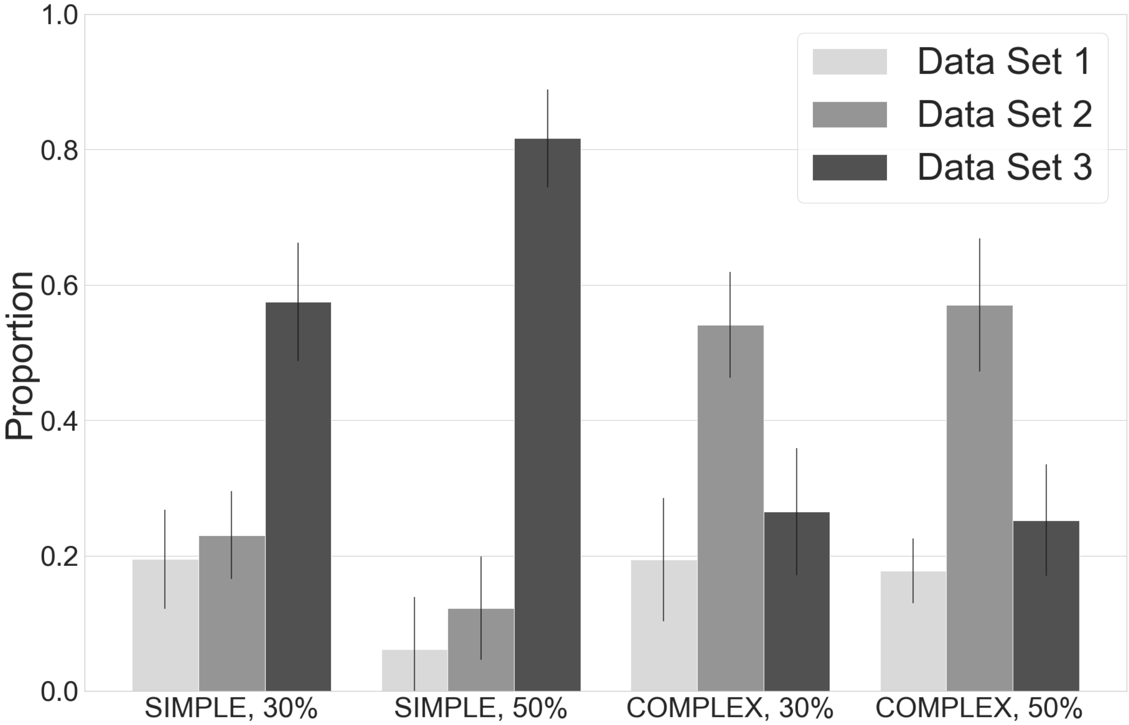
Results

Figure 7 shows the results of Experiment 4. As predicted, Data Set 2 was much more commonly selected as the basis for a speaker's reasoning when the speaker made a more complex causal claim, whereas Data Set 3 was more commonly selected when the speaker made a simpler causal claim. Moreover, and also as predicted, when the speaker made a simpler causal claim, Data Set 1 was more commonly selected when the value-of-information threshold was 30% than when it was 50%.

For significance testing, we constructed binary variables representing 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 these variables against: i) whether the speaker's claim was SIMPLE or COMPLEX ("granularity"), ii) whether the speaker's stated value-of-

Figure 7

Proportion of participants in each condition of Experiment 4 who chose each data set, with 95% CIs.



information threshold was 30% or 50% ("threshold"), iii) the vignette shown to participants, and iv) all interaction effects between all variables.

For Data Set 1, we found main effects of granularity ($\beta = 0.33, p = .028$) and of threshold ($\beta = 0.30, p = .047$), qualified by a marginal interaction ($\beta = -0.29, p = .053$). These findings are consistent with our predictions that Data Set 1 would be more likely to be selected when the speaker made a complex causal claim than a simple causal claim (especially when the speaker adopted a 50% value-of-information threshold), and when the speaker's threshold was 30% rather than 50%.

For Data Set 2, we found a significant effect of granularity, with participants more likely to choose Data Set 2 when the speaker's claim was complex rather than simple ($\beta = 0.93, p < .001$). We also found a significant interaction between granularity and threshold, with participants especially unlikely to select Data Set 2 when the claim was simple and the threshold was 50%

($\beta = -0.26\ p = .039$). This second finding is not specifically predicted by our theory, but does not qualify the key finding that Data Set 2 was more likely to be selected for complex than for simple claims, regardless of threshold.

For Data Set 3, we found a significant effect of granularity, such that participants were more likely to choose Data Set 3 when the speaker made the simpler causal claim ($\beta = -0.99, p < .001$). We also found a significant effect of threshold ($\beta = -0.28, p = .015$); this is consistent with an increased probability of selecting Data Set 3 when the threshold was 50%. Finally, we also found a significant interaction between granularity and threshold, with participants more likely to select Data Set 3 when speakers made the simpler causal claim with a threshold of 50% ($\beta = 0.33\ p = .004$). This too is in keeping with our prediction, which states that under these conditions, but not others, Data Set 3 is the choice that is most consistent with the speaker's granularity and value-of-information threshold.

To summarize, for all three logistic regressions we found significant or marginal results that confirm or are consistent with our predictions. We also failed to find any significant or marginal effects of the vignette shown to participants, or any significant or marginal interaction effects involving the vignette shown to participants; this is consistent with our results being robust to changes in the specific context in which participants are asked to make inferences.

General Discussion

As highlighted in the introduction, there is an intuitive connection between budgets and values: how we spend our resources sends a signal as to what we care about. The experiments reported in the current paper establish that this intuition is accurate when it comes to the relationship between the resources that a person spends to build a detailed causal representation of their social world and their values. Moreover, this result is implied by a more general theory of why we represent our world at a given level of granularity. This theory also entails all three of the mutually constraining relationships between variable granularity (representation), values, and data that are observed in our experiments.

We take our findings to have several potentially interesting implications. First, 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. One possibility is that these negative evaluations stem from the assumption that the speaker endorses an essentialist construal of the relevant categories (e.g., that the speaker takes a category like "race" to pick out internal properties of category members that explain observed differences across racial groups). While this may explain negative evaluations in some cases, our results suggest that an additional factor may be at play: the signal about a speaker's values conveyed by their generic causal claim. That is, we might take a speaker's use of a particular causal generic claim (especially one made at a particular level of granularity) to be evidence that they value (or fail to value) certain states of affairs over others. We may object on moral grounds to these implicit evaluative attitudes on the part of the speaker, independently of whether we ascribe to the speaker any particular essentialist beliefs. In future work, we hope to directly test whether these two sources of negative judgments about the speakers of objectionable generic causal claims can be isolated.

Second, and on a more philosophical note, there is a connection between our results and the idea, found in Haslanger (1995), that at least some social categories can be "pragmatically constructed," meaning that "their use is determined, at least in part, by social factors" (p. 100). If we take an agent's evaluations of their social environment, represented as a utility function, to be "social factors," then within our framework, social categories featuring in generic causal claims ought to be understood as pragmatically constructed, since whether or not they are used in a given scenario depends on the shape of the utility functions that represent these particular conative attitudes. More precisely, on our analysis the social categories reified through their use in generic causal claims can be said to be "weakly pragmatically constructed," in the sense that while the social factor of agents' utility functions plays a role in explaining when and why they are used, agent-independent aspects of social reality (i.e., the data produced by social processes) also constrain their use. This stands in contrast to "strongly pragmatically constructed social kinds," which Haslanger describes as "illusions projected onto the social world" that do not represent any agent-independent facts (1995, p. 100-101). This weakly pragmatic approach to social ontology is also defended in Lauer (2022), and stands in contrast with more explicitly realist approaches to social ontology such as Kincaid (2016, 2018).

Alternatively, one might instead claim that peoples' use of generic social categories to summarize data does not entail any commitment to the ontological reality of those categories at all. On such a reading, our results are consistent with the less philosophically controversial idea that people have a large ontology of social categories that they can potentially use to describe the world and to explain phenomena, with both pragmatic and empirical concerns dictating which categories they happen to use (Jenkins, 2023). For our part, we are comfortable with either reading of these implications of our work for social ontology.

A third implication of our findings also strikes a philosophical note. Work in the philosophy of causation by Yablo (1992) and Woodward (2010, 2016a, 2021b, 2021a) has advanced the notion of "proportionality" as a property of causal relationships, describing a relationship between cause and effect as "proportional" to the extent that changes in the state of that cause are informative about counterfactual changes in the state of the effect. This is relevant to the question of the level of granularity at which causal claims are made; if a more coarse-grained causal claim describes an equally proportional, if not more proportional, cause-effect relationship than a fine-grained causal claim about the same subject matter, then this may be said to speak in favor of using the more coarse-grained causal claim. Our results here suggest a reading of proportionality in terms of the value of information. A coarse-grained claim can be as proportional, if not more proportional, than a fine-grained claim that summarizes the same data when and because the fine-grained claim does not provide an agent with any *valuable* information about the cause-effect relationship. Thus, our account yields an understanding of proportionality wherein the proportionality of a causal claim is indexed to the interests of some agent.

Fourth, the idea that our generic causal claims about the social world reflect our cognitive budgets has roots in the literature on the psychology of categorization, e.g., Rosch (1978). Given the close connection between compression and more general cognitive processes like sense-making and understanding (Kirfel et al., 2021; Marzen & DeDeo, 2017; Pacer & Lombrozo, 2017; Wilkenfeld, 2019; Wojtowicz et al., 2021), it is likely the case that even if the granularity of our causal claims about the social world is determined largely by our agential interests, the particular compressed social categories that we use in these explanations can be taken as evidence for how agents make sense of their social world more broadly. Thus, our results here suggest a more general connection between decision theory, representation, and sense-making in human cognition.

Alternative Explanations and Limitations of the Current Studies

One might worry that throughout our experiments, a dynamic similar to the conjunction fallacy (Kahneman et al., 1982) is influencing our results. In the classic example of the conjunction fallacy, Linda, a 31-year-old single woman who majored in philosophy and was active in the anti-war movement in college, is deemed more likely to be a bank teller and active in the feminist movement than she is to be a bank teller. This violates the axioms of probability theory. We follow Tentori et al. (2013) in holding that these judgements are explained by the fact that our evidence about Linda provides a greater degree of *confirmation* for the conjunctive hypothesis that she is a bank teller and active in the feminist movement than it does the hypothesis that she is a bank teller. One could think that a similar dynamic might influence how participants generally evaluate the claims SIMPLE and COMPLEX in Experiments 1-4. For instance, participants might judge that a school having a high rate of algebra placement lends greater confirmatory support to it being majority white and having a small population of recent immigrants than it does to it being majority white. While we find it plausible that this could affect some participants' evaluations of our claims, we do not find it plausible that the degree of such influence would vary systematically with the value of information threshold manipulated in Experiments 2-4. Correspondingly, we do not think the conjunction fallacy offers a viable alternative explanation of our results.

Another potential concern is whether our predictions differ from what one might predict simply by assuming that listeners take speakers to follow Grice's *maxim of quantity*, which states that speakers should utter all and only the information relevant in a given context (Grice, 1975). In particular, if we assume that our manipulation of value-of-information thresholds affects what speakers take to be relevant, then speakers should utter less detailed claims (i.e., more coarse-grained claim) when the details they omit are irrelevant, and more detailed claims (i.e., more fine-grained claims) otherwise. This is consistent with what we observe in Experiment 2, and offers a basis for generating the predictions for Experiments 3 and 4, as well. However, the findings from Experiment 1 highlight important ways in which our framework goes beyond – though is not inconsistent with – this basic Gricean idea. First, Experiment 1 demonstrates a graded effect of information loss on the granularity of causal claims, with more fine-grained representations becoming increasingly frequent as information loss goes from zero to low to high. Information loss offers a continuous quantification of the less precise and typically binary term "relevance," and our findings lend it support. Second, Experiment 1 shows effects of information loss not only on verbal outputs, but on graph selection as well. This is consistent with our focus on causal *representation*, quite generally, as opposed to *linguistic communication*, in particular. Finally, we hold that, at least sometimes, people utter coarse-grained causal claims in part *because* they have represented the causal structure of a system in a coarse-grained manner. This is supported by another result from Experiment 1: participants were more likely to generate a coarse-grained verbal response *after* (vs. before) being asked to select a causal graph representing their understanding of the causal structure of the data-generating process. This suggests that participants first adopted a more coarse-grained representation of causal structure, and then provided a summary of the data in keeping with this structure.

While we take our experimental results to confirm the theoretical framework presented, we acknowledge that they have some important limitations. Our experiments were conducted solely on U.S.-based participants recruited through on-line platforms, and so we are limited in the extent to which we can generalize from our results here to agents in other contexts. This is an especially salient worry in light of the socially-relevant nature of our study. It is possible that social and cultural norms could be a moderating factor on the effects we observe. In addition, our results are constrained to relatively simple causal scenarios that can be presented and understood in a matter of minutes, such that our results are of less significance in understanding deliberative attempts to represent the causal structure of the world (as in science), attempts that can unfold over much longer periods of time than those studied in the experiments reported here.

Another important limitation is that our studies shed little light on the mechanisms by which participants produced responses. For instance, our findings remain silent on the question of what, precisely, increases the cost associated with more complex representations (candidates for such a factor include demands on working memory, processing time, etc.). Our findings also remain silent concerning the process by which participants generated or selected a given representation. In particular, it is psychologically implausible that they first generated and considered all possible representations and then selected between them; it is much more plausible that various heuristics guided which representations were even constructed or entertained. Thus, our talk of "choosing" a representation must be read loosely, and the mechanisms at work in constructing and evaluating causal representations remain an important area for future research.

Directions for Future Work

As mentioned above, an immediate avenue for future work is to explicitly test whether judgments about speakers of generic causal claims can be decomposed into judgments about the essentializing attitudes of those speakers towards members of certain groups and judgments about the speaker's values that can be inferred from the generic causal claims in question. If such a decomposition is possible, one could then test how each of these judgments contributes to a listener's overall moral evaluation of a speaker who makes a generic causal claim.

Relatedly, as demonstrated in our discussion of previous work, others in this literature have argued that speakers use generics in order to communicate to listeners that a certain feature of a person is a natural kind or an essential property of that person. Our results here identify a further use of generic language: to communicate that a person's membership in a particular category is relevant information. It remains for further work to explore whether, and to what extent, the evidence currently taken to support the hypothesis that generics communicate beliefs about the naturalness or essence of categories can also be explained by the hypothesis that generics communicate evaluative information, and whether there are any reasons to prefer the latter explanation over the former. In other words, it remains to future work to vitiate as to whether our model identifies a previously undiscussed, additional role for causal generics, or amounts to a rival to existing accounts of the role of generic language in perpetuating existing attitudes towards social groups.

From the perspective of social theory, an intriguing avenue for future work concerns the interaction between agential values and a propensity to provide intersectional explanations for social outcomes. Intersectional theories explain patterns of social advantage and disadvantage in terms of the non-additive interactions between the multiple, distinct social categories to which a person can simultaneously belong (Crenshaw, 1991). In the context of our framework, intersectional explanations can be understood as more fine-grained than non-intersectional explanations, as the interactions between social identities that they cite add strictly more detail to their causal description of the social environment (see Bright et al., 2016 for a formal articulation of intersectional theories as causal theories). One could then argue that agents are more likely to provide intersectional explanations when their value-of-information thresholds render information about certain interactions between social inequalities valuable. Further empirical work could then investigate whether people are actually more likely to provide intersectional explanations when placed in decision scenarios such that these explanations contain more valuable information.

Finally, we note that in some cases, agents may value fine-grained information about social categories precisely because they seek to ameliorate pernicious social inequalities. For instance, someone might represent both race and sex when it comes to math performance not because they take these categories to offer valuable information regarding intrinsic aptitude for math, but because they want to track and correct systematic sources of inequality (see also Vasilyeva & Lombrozo, 2020, for relevant discussion). Our framework offers a basis for inferring agential values from representations and data, but does not tell us *why* individuals value the information that they do. It remains to future work to investigate when and how such further inferences are drawn.

Conclusion

Summing up, we find evidence for mutually constraining relationships between listeners' inferences regarding the level of granularity at which speakers make generic causal claims, the values of those speakers, and the data available to those speakers. These relationships are found to hold in cases where the generic causal claims in question feature social categories. This suggests that listeners' expectations about which social categories speakers are likely to invoke in describing the causal structure of the world may be just as shaped by listeners' judgments about the goals of that agent as they are by listeners' judgments about the essentializing attitudes of the speaker. It also suggests that people are guided by both pragmatic and empirical concerns when they selectively choose to use particular elements from a wider ontology of social categories to describe the world and explain phenomena. Taken together, our results suggest an intriguing and potentially deep connection between our pragmatic goals and our representations of the social world that we inhabit.

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.

Bright, L. K., Malinsky, D., & Thompson, M. (2016). Causally interpreting intersectionality theory. *Philosophy of Science*, *83*(1), 60–81.

Campbell, D. T. (1958). Common fate, similarity, and other indices of the status of aggregates of persons as social entities. *Behavioral Science*, *3*(1), 14.

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.

Crenshaw, K. (1991). Mapping the Margins: Intersectionality, Identity Politics, and Violence against Women of Color. *Stanford Law Review*, *43*(6), 1241–1299. https://doi.org/10.2307/1229039

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., Goetz, P. J., Sarnecka, B. W., & Flukes, J. (2008). Generic Language in Parent-Child Conversations. *Language Learning and Development*, *4*(1), 1–31. https://doi.org/10.1080/15475440701542625

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.

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

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

Haslanger, S. (1995). Ontology and social construction. *Philosophical Topics*, *23*(2), 95–125.

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

Jenkins, K. (2023). *Ontology and oppression: Race, gender, and social reality*. Oxford University Press. https://books.google.co.uk/books?hl=en&lr=&id=Chq1EAAAQBAJ&oi=fnd&pg=PP1&dq=ontology+and+oppression&ots=R_VaJnXvkH&sig=mfM3yh2GefcyLY71FvQAVeH7xiU

Kahneman, D., Tversky, A., & Slovic, P. (1982). *Judgments of and by representativeness*. https://books.google.co.uk/books?hl=en&lr=&id=YfjwCAAAQBAJ&oi=fnd&pg=PA25&dq=Judgments+of+and+by+representativeness&ots=eJ8abobBqo&sig=hNKtBFVT2ZRcvW_LWDv3qbMDppQ

Kincaid, H. (2016). Debating the reality of social classes. *Philosophy of the Social Sciences*, *46*(2), 189–209.

Kincaid, H. (2018). Debating the reality of race, caste, and ethnicity. *Philosophy of the Social Sciences*, *48*(2), 139–167.

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. B., & Lombrozo, T. (2023). *Building Compressed Causal Models of the World*. PsyArXiv. https://doi.org/10.31234/osf.io/2f7x6

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.

Kirfel, L., Icard, T., & Gerstenberg, T. (2021). Inference from explanation. *Journal of Experimental Psychology: General*.

Lauer, R. (2022). Motivating a Pragmatic Approach to Naturalized Social Ontology. *Journal for General Philosophy of Science*, *53*(4), 403–419. https://doi.org/10.1007/s10838-021-09581-3

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

Marzen, S. E., & DeDeo, S. (2017). The evolution of lossy compression. *Journal of The Royal Society Interface*, *14*(130), 20170166.

Noyes, A., Dunham, Y., Keil, F. C., & Ritchie, K. (2021). Evidence for multiple sources of inductive potential: Occupations and their relations to social institutions. *Cognitive Psychology*, *130*, 101422.

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

Pacer, M., & Lombrozo, T. (2017). Ockham's razor cuts to the root: Simplicity in causal explanation. *Journal of Experimental Psychology: General*, *146*(12), 1761.

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

Rhodes, M., Leslie, S.-J., & Tworek, C. M. (2012). Cultural transmission of social essentialism. *Proceedings of the National Academy of Sciences*, *109*(34), 13526–13531. https://doi.org/10.1073/pnas.1208951109

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.

Tentori, K., Crupi, V., & Russo, S. (2013). On the determinants of the conjunction fallacy: Probability versus inductive confirmation. *Journal of Experimental Psychology: General*, *142*(1), 235.

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.

Wilkenfeld, D. A. (2019). Understanding as compression. *Philosophical Studies*, *176*(10), 2807–2831.

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

Wojtowicz, Z., Chater, N., & Loewenstein, G. (2021). *The motivational processes of sense-making*.

Woodward, J. (2010). Causation in biology: Stability, specificity, and the choice of levels of explanation. *Biology & Philosophy*, *25*(3), 287–318.

Woodward, J. (2016a). Causal cognition: Physical connections, proportionality, and the role of normative theory. *Of Psychology: Causality and Psychological Subject*, 105.

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

Woodward, J. (2021a). *Causation with a human face: Normative theory and descriptive psychology*. Oxford University Press.

Woodward, J. (2021b). Explanatory autonomy: The role of proportionality, stability, and conditional irrelevance. *Synthese*, *198*(1), 237–265.

Yablo, S. (1992). Mental causation. *The Philosophical Review*, *101*(2), 245–280.