sociological science

Measuring Memberships in Collectives in Light of Developments in Cognitive Science and Natural-Language Processing

Michael T. Hannan

Stanford University

Abstract: Which individuals and corporate actors belong in a collective, and who decides? Sociology has not had good analytical tools for addressing these questions. Recent work that adapts probabilistic representations of concepts and probabilistic categorization to sociological research opens opportunities for making progress on the measurement of memberships. It turns out that the probabilistic cognitive-based reformulation reveals unexpected connections to language models and natural-language processing. In particular, the leading probabilistic classifier BERT provides new and powerful ways to measure core concepts.

Keywords: categorization; group membership; machine learning

Sociological research on culture, economic life, and organizations has begun in recent years to take cognition seriously. Although this changed focus has yielded interesting new insights, it has not yet reached foundational issues, but it should. I consider one such issue here: how to establish memberships of agents in collectives such as artistic movements, sects, and political parties. The issue of what entities "belong" to a collective and to what degree shapes both how agents think about them and how sociologists ought to study them. Analyzing how agents "see" boundaries and memberships in light of what we have learned about cognition presents interesting opportunities and challenges.

This article takes on this issue. To make the arguments and proposals concrete, it narrows its focus to consider collectives in terms of people's mental representations of them. In thinking about the membership of a movement, say, it regards the relevant agents as having a mental representation of the movement. Such a mental representation (concept) expresses what it means to be a member or instance of the collective, a movement in this example. We can gain insight about membership of entities in collectives by using the metrics of typicality and ambiguity. Typicality refers to judgements of how well an entity fits what is expected of instances of a concept. Ambiguity refers to the level of confusion that arises in a focal agent's mind about which of several potentially relevant concepts apply to an entity. Typicality and ambiguity are individual judgements. One of the most pressing analytical challenges for sociological applications involves representing how such judgements aggregate into social judgements.

The initial sections of the article concentrate on developments in one line of sociological research: organizational ecology. This choice is helpful because this style of work has been especially self-conscious about justifying measurements of memberships (as instances of organizational forms, in this case) and has also tried

Citation: Hannan, Michael T. 2022. "Measuring Memberships in Collectives in Light of Developments in Cognitive Science and Natural-Language Processing." Sociological Science 9: 473-492.

Received: August 8, 2022

Accepted: September 28, 2022

Published: December 16, 2022

Editor(s): Ari Adut, Ray Reagans

DOI: 10.15195/v9.a19

Copyright: © 2022 The Author(s). This open-access article has been published under a Creative Commons Attribution License, which allows unrestricted use, distribution and reproduction, in any form, as long as the original author and source have been credited. 🅭🅯

to employ tools from cognitive science. Organizational ecology took a cognitive turn that was spurred by a concern that the original notion of *organizational form* proposed by Hannan and Freeman (1977, 1984) needed to be sharpened to provide a better guide to empirical research. The 1977 proposal sought an analogue in the organizational domain of biology's species concept. The idea advanced was that an organizational form was a blueprint for organizing and that blueprints could be inferred from organizational designs and normative orders ("the ways of organizing that are defined as right and proper by both members and relevant sectors of the environment" [P. 935]). This sparse specification was made more concrete by Hannan and Freeman (1984): "four properties provide a possible basis on which to classify organizations into forms for ecological analysis" (P. 156). These are (1) stated goals, the bases on which organizations seeks to gain legitimacy and other symbolic resources; (2) form of authority; (3) core technology; and (4) the ways of attracting resources from the environment. Subsequent work tried to rework the notion (especially the fourth point) from the perspective of the human agents who control the resources needed to sustain the organizations that exemplify the form.

Hannan and Freeman (1977) argued for an ecology of organizations that relies on analysis of selection processes as a corrective to what they saw as an overly optimistic adaptationist view. The key motivation was the belief that organizations face inertial pressures that preclude major adaptation of "core" features. Such processes justify a focus on *populations* of organizations.

Pursuing this agenda requires defining form and population. I argue here that clarity can be gained by considering form to be an ordinary concept with a high level of agreement in the audience about meaning, specifically about what properties its instances should be expected to possess. If form is framed as a collective concept, then it is natural to define population as the result of a set of categorization judgements. Proceeding along these lines allows strong connections to main lines of research in cognitive science.

But most research on organizational ecology (and most of sociology) did not proceed along these lines. Instead analysts designed explicit rules for determining membership. This approach[1] reflected the so-called classical rendering of concepts, which holds that any entity is either fully an instance of the concept or not an instance at all. In other words there is no fuzziness or uncertainty about membership. As I recount below, research over the past 50 years leaves little doubt that uncertainty prevails in decisions about membership. The work described next tried to reflect this uncertainty directly as partial membership and fuzzy boundaries. Under this interpretation, crisp rules cannot be written for deciding what entities to include in the study of a collective. Surprisingly (at least for me), a fully probabilistic approach can rescue the original approach of using crisp inclusion rules. I discuss this below.

The Audience Turn

A key step toward a modern view on memberships was shifting the burden of deciding what is and is not an instance of a form (or any other concept) from the analyst to the agents in the system being studied (Hannan, Pólos, and Carroll 2007; Hannan 2010).

Prototypes and Graded Membership

A key idea holds that audience judgements do not produce the kind of crisp boundaries that analysts assume. The basic element is a *concept*, a mental representation that tells what to expect, in varying degrees, of instances. Audience members very likely perceive what Eleanor Rosch (1973; 1975) called a structure of *graded membership* in concepts. Some entities are seen as clear instances of a concept, others as clearly not, and others as instances to a greater or lesser degree. In the Roschian view, the structure is anchored by one or more prototypes, ideal versions of instances of the concept.

Fuzzy Forms and Memberships

Hannan, Pólos, and Carroll (2007), hereafter HPC, defined an organizational form as a concept[2] that meets three requirements. First the concept pertains to a kind of organization. Second the concept is widely shared in the focal audience in the sense that the members of the audience agree about what to expect of instances of a form. Third the concept has strong legitimation in the sense of taken for grantedness.

HPC used fuzzy set theory's grade-of-membership function, in notation $\mu_{\mathfrak{c}}(o)$, to define the membership of organization o in the organizational form $\mathfrak{c}$. This function ranges over [0,1] and tells the degree to which an entity is a member of a set. This proposal addressed two key questions. First, what entities should be considered as potential members of a form? HPC's implicit answer is the set of all the objects labeled as organizations. This set, denoted as $\mathcal{O}$, serves as the universe of discourse.

A second, more vexing, issue concerns aggregation over the members of a focal audience, in notation $\mathcal{A}$. The typicality of an entity as an instance of a concept is an individual judgement. Each member of the audience determines how typical is one or more persons or corporate actors as an instance of the concept. So there can be as a many memberships as there are member of the audience. Allowing such full variability makes analysis intractable. How can one proceed? HPC considered the average over the members of the audience of the grades of membership function, $\bar{\mu}_{\mathfrak{c}}(o) = \sum_{a \in \mathcal{A}} \mu^{a}_{\mathfrak{c}}(o)/|\mathcal{A}|$, as determining whether a concept is an organizational form.

Then the membership (or population) associated with the form $\mathfrak{c}$ can be defined as a set of ordered pairs of organizations and average grades of membership:

$$\mathbf{mem}(\mathfrak{c}) = \{\langle x, \bar{\mu}_{\mathfrak{c}}(o)\rangle, o \in \mathcal{O}\}. \tag{1}$$

Fuzzy set theory defines the size (cardinality) of a set as the sum of grades of membership. In other words, an entity's contribution to the size of the set is simply its grade of membership in the set. Using this definition the key notion of the size of the membership can defined as

$$n(\mathfrak{c}) = |\mathbf{mem}(\mathfrak{c})| = \textstyle\sum_{o \in \mathcal{D}} \bar{\mu}_{\mathfrak{c}}(o).$$

This fuzzy-set construction has the appealing property that prototypical instances of a concept contribute more to its size (or density) than do marginal examples. However, this construction does a poor job of representing how people think and talk about persons and corporate actors generally. It would be weird to learn that an entity that is very atypical of some concept (say with grade of membership close to zero) *does* count as a member. Subsequent work using categorization functions offers a better way to represent uncertainty in membership, as I discuss below.

A Fully Probabilistic Approach

The crucial groundwork has been prepared for a new approach that replaces the fuzzy-logic notion of grade of membership with a fully developed probability model. My discussion of these issues has three parts. The first sketches the probability model. The second contains proposals for applying this model to central issues in sociology. The third argues for using modern deep-learning language models as a way to measure key constructs in the model.

The foundations of the probability model are reported in the monograph *Concepts and Categories* (Hannan et al. 2019), hereafter C&C. This work adopted a subjective probabilistic construction that can be used for a wide variety of analytical issues including categorization and valuation. Moreover, this framework allows us to build predictive models so that the accuracy of these predictions can be exposed to empirical tests.

Semantic Space

Ideas about concepts and categorization invariably consider the *features* of entities. The semantic space for a concept is the space defined by the possible values of the features that a focal person regards as relevant for judging the typicality of entities as possible instances of the concept. For example, the traditional[3] wine genre is defined in terms of such features as degree of intervention in fermentation, length of the maceration period, and type of vessel used for aging (Negro, Hannan, and Olzak 2022). The agent's mental representation of a particular entity such as a winery can be regarded as a position in the semantic space.

Concepts

A contemporary rendering treats a concept as a probability density, $\pi_{\mathbb{F}_{\mathfrak{c}}}(x \mid \mathfrak{c})$, defined over a semantic space (in notation, $\mathbb{F}_{\mathfrak{c}}$). This formalism gives the subjective probability (or belief) that an entity believed to be an instance of the concept $\mathfrak{c}$ has some particular combination of values of relevant features. This function tells the subjective probability distribution of positions in semantic space for $\mathfrak{c}$s. In other words, the *meaning* of the concept $\mathfrak{c}$ (in notation $[\![\mathfrak{c}]\!]$) can usefully be represented as given by this probability measure: $[\![\mathfrak{c}]\!] \equiv \pi_{\mathbb{F}_{\mathfrak{c}}}(\cdot \mid \mathfrak{c})$.

Concept likelihood and typicality. How does the concept likelihood relate to typicality and grade of membership, the backbone of the previous approach? In her seminal work on typicality, Eleanor Rosch (1973) proposed that prototype means "an excellent example" of a concept. Objects that differ greatly from the prototype are poor examples. The concept likelihood captures exactly this idea. But explicating the relationship between typicality and concept likelihood required the authors of C&C to fill in the missing details on what exactly these terms mean.

Researchers usually elicit typicality empirically by describing or depicting an entity and asking, "How typical is this entity for the concept $\mathfrak{c}$?" This approach sidesteps the task of specifying the semantic space. (But judgements of typicality surely depend on the choice of that space.) The Roschian idea can be represented by defining the typicality as the similarity of the object's position in the semantic space and the positions in the "center" of the concept.

Unlike concept likelihoods, which as probabilities must fall in [0,1] and sum to one over the space, typicalities, as measured in previous research, are not so constrained. An obvious adjustment rescales typicalities so that they are bounded in [0,1] and sum to one over the space. Specifically, let $\tau_{\mathfrak{c}}(x)$ denote the typicality of a position in concept $\mathfrak{c}$ for an individual, and let $\tau^*_{\mathfrak{c}}(x)$ be defined as the ratio of $\tau_{\mathfrak{c}}(x)$ to the sum of the $\tau_{\mathfrak{c}}(x)$ over the space. With this modification, it is natural to assume that concept likelihoods are equivalent to scaled typicalities:

$$\pi_{\mathbb{F}_{\mathfrak{c}}}(x \mid \mathfrak{c}) \equiv \tau^*_{\mathfrak{c}}(x).$$

Domains and Cohorts of Concepts

Contexts also play a crucial role because they make certain conceptual domains relevant and can allow for different ways of categorizing entities. For example, a discussion of where to have lunch elicits the restaurant domain. C&C introduced the idea of a *cohort of concepts*: a set of concepts (1) that are subconcepts of a common root—the domain concept—and (2) such that no member of the set is a subconcept of any other. The idea of a cohort plays an important theoretical role because it identifies the confusing cases, those in which an entity can be seen as instance of more than one of the class of related concepts. The fact that someone might be labeled as <Italian, woman, sociologist, rower> might not be confusing because the various concepts come from different domains. So all of the arguments below are conditioned to apply to a single conceptual domain and its cohort of concepts, where expectations for concepts likely clash. Clashing expectations are the prime source of conceptual ambiguity, as I discuss below. For instance, if science is the root, then the concepts in the domain include science, social science, chemistry, sociology, linguistics, and so forth. So a context evokes a domain and the set of potentially applicable concepts. In the example given above, the concepts stand at different "levels" of a partial ordering by a subconcept relation; for example, sociology is a subconcept of social science. For many analytic purposes it makes sense to consider only the concepts that stand at the same level.

Categorization

Categorization means assigning entities to concepts, for example, "this group is a terrorist organization." If concepts lack sharp boundaries, then categorization involves some uncertainty. Some objects generally have the expected feature values (are located in the semantic space close to the center of a concept), and there is little doubt about categorization. Others have feature values that locate them very far from the concept center, meaning that there is little doubt that the concept does not apply. Otherwise uncertainty exists to varying degrees. The probabilistic parallel of the (fuzzy) grade of membership is the *categorization probability*.

The key to relating concepts and memberships has to do with explicating the form of categorization uncertainty. C&C follows the modern Bayesian approach by looking at categorization as a problem of statistical inference (Anderson 1991; Tenenbaum and Griffiths 2002). Such analysis considers a person deciding whether to regard an entity—more precisely its mental representation—as an instance of a concept. The instance-of idea can be expressed formally with the predicate $\text{IS-A}(\mathfrak{c}, o, a)$ that reads "the agent a believes the entity o to be an instance of the concept $\mathfrak{c}$."

The probabilistic perspective on categorization interprets the strength of a person's belief in the truth of the hypothesis $\text{IS-A}(\mathfrak{c}, o, a)$ as a subjective probability. Suppose that one has to answer "yes" or "no" to the question "Is the entity o a $\mathfrak{c}$?" when the mental representation of the entity is the position x in the semantic space associated with the concept $\mathfrak{c}$: $R_{\mathbb{F}_\mathfrak{c}}(o) = x$. The Bayesian approach holds that the probability that one would answer "yes" is given by the *Bayesian categorization probability*:

$$P(\text{IS-A}(\mathfrak{c}, o, a) \,|\, x) \equiv P_a(\mathfrak{c} \,|\, x).$$

(The expression on the right in this equivalence gives a convenient notational shorthand.)

The foregoing relationship can be re-expressed using Bayes' theorem (and the notational shorthand) as

$$P\big(\mathfrak{c} \,|\, x\big) = \frac{P_{\mathbb{F}_\mathfrak{c}}(x \,|\, \mathfrak{c})\, P(\mathfrak{c})}{P_{\mathbb{F}_\mathfrak{c}}(x)}.$$

This equation contains priors on the position in semantic space, $P_{\mathbb{F}_\mathfrak{c}}(x)$, and on membership in the concept (independent of position), $P(\mathfrak{c})$. In most applications these priors are set to so-called base rates, the proportion of all potentially relevant objects observed to occupy position x and to belong to the concept $\mathfrak{c}$.

Finally the Bayesian categorization probability can be rewritten in terms of the concept likelihood:[4]

$$P\big(\mathfrak{c} \,|\, x\big) = \pi_{\mathbb{F}_\mathfrak{c}}(x \,|\, \mathfrak{c}) \frac{P(\mathfrak{c})}{P_{\mathbb{F}_\mathfrak{c}}(x)}. \tag{2}$$

This formulation lies at the heart of the modern probabilistic approach in that it ties the abstract notion of concept to judgements about concrete objects. In this view, *a category is a realization of an underlying Bayesian categorization process.* Then an agent's category for the concept $\mathfrak{c}$ is the *crisp* set of objects that the agent judges to be instances. In other words, the membership in the concept is given by

$$\mathbf{mem}(\mathfrak{c}, a) \equiv \{o \,|\, \text{IS-A}(\mathfrak{c}, o, a)\}.$$

Form Reconsidered

The complicated argument just recapitulated offers a revised way of defining organizational form, one that avoids the limitation of the test-code structure and that makes clear exactly what is being taken for granted.

Consider the not-yet-analyzed notion of the audience. Earlier work stipulated that an audience for a domain consists of agents who have an interest in the domain and also control material and symbolic resources necessary for the survival of the organizations "in" the domain. Domains generally contain multiple sub-audiences. These can include actual and potential organizational members, active and potential customers or patrons, critics and other market intermediaries, controllers of capital, and regulators. The multiple audiences need not agree about the meaning of the concepts in the domain. Even within sub-audiences we expect something like a division of labor with some members paying much closer attention to the organizations and their actions. Koçak, Hannan, and Hsu (2014) refer to such segments as *audience vanguards*, and they argue that the emergence of consensus about meanings generally arises within one or more vanguards and then spreads to the rest of the audience. So it might be necessary to work audience-by-audience in analyzing particular cases.[5]

A further condition must be met: audience members must have conceptualized the root concept that establishes the domain. So, for instance, the audience members for various film genres are those whose inventory of concepts includes film. What does it mean to "have" a concept? In the context of the probabilistic model, having a concept labeled $\mathfrak{c}$ means associating the meaning of $\mathfrak{c}$ with a probability distribution over a feature (semantic) space.[6]

Next consider the issue of aggregating from the concepts of individuals to the audience level. A crucial question asks whether the audience members agree about the meaning of concepts. We can address this question by considering the *distances* among the concepts of the audience members. Specifically C&C uses the Kullback–Leibler divergence, a measure of the distance between two probability measures[7] to define a distance between a pair of concepts, in notation $\vec{D}(\mathfrak{c}_a, \mathfrak{c}_b)$.[8]

With this definition of directed distance, the dissensus about a concept within an audience can be defined simply as the average dissensus about the concept over all pairs of members of the audience.

HPC's definition of organizational form, as noted above, requires that a collective concept have high taken for grantedness to qualify as a form. In hindsight, conflating the two notions (collective agreement and taken for grantedness) appears to be a strategic mistake in theory building. Taken for grantedness might reach a high level long after an audience has reached collective agreement about meaning. If population is defined a concrete instantiation of form, then it is undefined for the period from the onset of agreement until (and if) the concept becomes highly taken for granted according to this definition. This problem would vitiate attempts to model the dynamics of taken for grantedness within a population, a crucial component of the theory of density dependence.

I suggest returning closer to the original idea, to define an organizational form as a collectively agreed-upon concept referring to the world of organizations. As a first step, consider the more general case. Following Negro et al. (2022), I use the term genre to refer to an agreed-upon concept generally.

Definition 1 (genre)**.** *A concept, labeled by $\mathfrak{c}$, is a genre for an audience if its members generally agree about the meaning of the concept; that is, dissensus in the audience about the concept does not exceed some small (positive) constant.*

Suppose that $\mathfrak{c}$ is a concept for the members of the focal audience.

$$\forall\, A, \mathfrak{c}\, \exists\, \delta\, [(\delta > 0) \wedge \textsc{genre}(\mathfrak{c}, A) \longleftrightarrow \delta > \overline{\mathcal{D}}_A(\mathfrak{c})], \tag{3}$$

where $\overline{\mathcal{D}}_A(\mathfrak{c})$ denotes the average dissensus in the audience about the concept.

Unfortunately the definition rests on the value of the (unknown) constant (that might depend on the concept/audience as the definition is stated). Empirical research might experiment with alternative values of the constant.

With genre defined, we can regard an organizational form as a genre within the domain whose root is organization.

Membership in a Collective from Categorizations

From the perspective of an individual, the membership of a collective is an ordinary category. Moving to the level of the audience means dealing with aggregation of individual judgements.

One straightforward—but possibly extreme—way of way of aggregating individual judgements is as to the *superset* of all of the categories of the members of the audience. Then the collective category consists of the set of all objects that at least one member of the audience considers to be an instance of the focal concept. Notice that this form of aggregation preserves crispness. Again the size (density) of the collective is the cardinality of the set of members. But because the membership as defined above is crisp, the cardinality is simply the number of unique members.

If the audience is fairly homogenous, for example, a set of professional critics or enthusiasts, then this superset construction might make sense for empirical applications. But if there is considerable heterogeneity, then the membership that results from this construction might be one whose composition does not make sense to any of the members of the audience.

Other options for aggregation might have more empirical promise. Consider a concrete example. Goodreads.com reports for each book aggregated assignments by users to one or more literary genres. For each book, the data reveal the distribution of categorizations over the website's list of 36 major genres. Think of the data as a matrix with books as rows and genres as columns with each cell giving the number of assignments. The superset proposal sketched above defines the membership for a genre as the set of books with a nonzero entry in for that genre's column, a minimum criterion. Alternatively one could use a maximum criterion: set the membership in a genre to the collection of books for which that genre is its modal categorization. This rule effectively converts (or truncates) multiple categorizations of objects to single categorizations. Obviously some intermediate rules can be designed, for example, by varying the minimum level used in the first procedure, for example, all books with N or more categorizations in a genre or all with proportion P or higher assignments to the genre.

These alternatives also yield crisp memberships. However, the crispness results from the use of arbitrary cutoffs. Nonetheless, all of these constructions appear to be more reasonable from a cognitive perspective than the fuzzy-set construction discussed above. The latter requires that people retain in memory not only a set of objects they have experienced in the domain but also the grades of membership in concepts of all of these objects. In other words, the cognitive load imposed on the agents by this requirement would be exceedingly high. Under the revision I propose, people need recall only the set of objects that they categorized as instances. This approach thus respects the tendency toward what Klaus Fiedler (2012) calls *metacognitive myopia*. Extensive research shows that people generally recall the decisions that they made but not the cognitive processes involved in making the decision. For the case at hand, this would mean that people tend to remember their categorizations but not the uncertainty involved in those decisions.

Implications and New Research Directions

Social Concept Learning

The key notion genre (collective concept) depends on intensional consensus, which will be difficult, if not impossible, to measure directly, because intensions are not observable, as noted above. One possible way forward is to proceed indirectly using categorizations. Unlike concepts, categorizations are observable. That is, people can share their extensions and even make them public, as when professional critics publish lists of rated objects.

Greta Hsu (2006) was the first to use observed categorizations to measure extensional consensus about objects. Archival records provided genre categorizations of film by several highly visible professional critics. Hsu calculated consensus about a film's genre as a simple similarity measure for pairs of critics and then averaged over pairs.

I suggest a different calculation from the observed matrix of critics by films by genre: measure the agreement between pairs of critics by genre, then average over pairs to obtain a measure of extensional consensus for a genre. If archival materials provide such matrices for multiple time points, then research can trace the evolution of consensus. In the best case such a record covers the early history of a proto-genre.

For this strategy to work, extensional consensus must linked to consensus about meanings (intensional consensus) and taken for grantedness. Addressing the first issue, the link between extensional and intensional consensus, requires new theory on concept learning. In particular, we need to know how individuals adjust their concepts in response to categorizations by others. Given the Bayesian foundations on which I build, a promising approach is to adapt Bayesian models of social cognition. Such models assume that individuals update their mental representations (concepts) in response to observed categorizations by others. In general, such a model imposes unrealistic computational demands on the agent.

Many modelers have responded by weakening the dependence on Bayesian analysis (producing so-called non-Bayesian models of social learning).

A more attractive approach in my view is a social sampling model for collective learning introduced by Krafft et al. (2021). In the first stage of this model, agents sample a decision among a set of options (such as categorizations) with probability proportional to the popularity of this choice in the population of agents. This stage represents social influence that allows information to be aggregated over time. In the second stage, the agent decides whether to accept or reject this choice by performing a Bayesian calculation of the likelihood of their available information about the situation (feature values of the entity in the C&C framework) given the choice. The stage continues until an option is chosen. The second stage is designed to represent a heuristic decision process of bounded rationality given that the decisions are made locally rather than globally.

The implications of this model have been studied for the special case in which one of the options is "best" (a so-called hide-and-seek environment). For this case there is a tight link between heuristic decision-making at the agent level and the evolution of (extensional) agreement at the population level. Specifically the agent's posterior becomes proportional to the average decision for the population of agents as time unfolds.

The environment for categorization does not have "best" choices. So we do not yet know the implications of the social sampling model for individual categorization behavior and extensional agreement. If something like the implication for hide-and-seek environments holds for categorization contexts, then we have a path toward addressing the sociological problem. Such a suitably generalized model would not lead to a measure of agreement about mental representations. Rather this kind of model treats adjustment of such representations so as to yield collective coordination on decisions (including categorizations). As such a process unfolds, the agents behave as though they agree about the representations. In this sense this kind of model implies that intensional and extensional agreement coevolve. This is what is needed to provide a warrant for treating periods in which extensional consensus increases sharply as a likely time of the emergence of intensional consensus and of genre.

Conceptual Ambiguity

Hsu (2006) began the modern stream of work on conceptual ambiguity with an analysis of the jack-of-all-trades issue. She argued that both professional critics and the general audience find it confusing when organizations behave like jacks of all trades, masters of none, and that they react negatively to such confusing experiences. Hsu, Hannan, and Koçak (2009) proposed a more general treatment along with a measurement strategy (discussed below). People find it easy to make sense of objects that they can definitively associate with one particular concept, but they likely have great difficulty interpreting others that fit partially with many concepts but not one in particular. For example, a movie that has elements of horror and romance is arguably more difficult to interpret than ones that have elements of just one genre.

The surge of research in these issues followed Hsu et al. (2009) in using the language of "category spanning." C&C argued that it is not multiple categorization per se that makes an entity hard to interpret but that having feature values makes it hard to decide which concepts apply. This assumption makes the argument general enough to apply even in situations in which entities have not yet been categorized.

The previous research conceptualized vectors of typicalities in concepts as defining so-called categorical niches in semantic space.[9] An entity with high typicality in one concept and low typicality in others in a cohort has a narrow categorical niche. By virtue of its specialized position, such an organization has strong appeal to a narrow band of the audience. And an entity with equal typicality in each concept has the broadest possible categorical niche.

The probabilistic reformulation of these ideas considers an entity as conceptually ambiguous to the extent that a person finds it hard to make sense of it in terms of their concepts in the cohort for that context. The ambiguity of a mental representation (as a position in the semantic space) depends on the distribution of categorization probabilities for that position for all of the concepts in a cohort. Objects represented as positions with a high categorization probability for only one concept have low conceptual ambiguity; objects represented as positions with an even distribution of categorization probabilities have maximal conceptual ambiguity.

Another way to put it is that an entity is ambiguous if it could likely be considered an instance of more than one concept in a cohort. To represent this intuition, C&C defines conceptual ambiguity as the *entropy* of the vector of the scaled[10] categorization probabilities.

Measurement of Typicality and Ambiguity

Recasting sociological arguments on secure cognitive foundations requires improvement in measurement. Neither organizational form nor the typicality and ambiguity of organizations have yet been measured in a manner that fits knowledge about cognition. Doing so requires empirical measurement of semantic spaces, the positions of entities in the semantic space, and their categorization probabilities. Each is problematic without a major methodological reorientation. To this point research has proceeded using primarily observed categorizations because measurements of relevant feature values have been lacking. So the challenge is to measure semantic space and estimate the categorization probabilities.

Organizational form, typicality, and ambiguity are latent psychological variables that depend on agents' concepts and on their perceptions of objects' positions in semantic space. In many empirical settings these constructs are not directly observable to the researcher. However, taking advantage of the wide availability of textual descriptions of organizations and products and of recent progress in machine-learning classifiers offers a promising approach. Before sketching the new approach, I summarize the current methodology that works with observed categorizations.

From labels (categorizations). Many previous studies on concepts in cultural and economic life have made inferences about typicality from categorizations because the researchers generally have access to categorizations but not to the values of the concept-relevant features. Sometimes the available data report only the categorizations. This can occur when a market intermediary, such as a website curator or regulator, assigns labels. The first (largely implicit) step in building a measure of typicality from labels (categorizations) assumes that objects labeled as instances of only one concept in the cohort generally fit better to that concept than those objects that get two labels. The reasoning then makes a similar assertion about dual labeling versus triple labeling, and so forth. Overall the expectation is that the typicality in any assigned concept decreases monotonically with the number of concepts assigned (subject to the condition that it remain non-negative).

Hsu et al. (2009) proposed a simple functional form for typicality from labels: the typicality of an entity in a concept is zero if it is not labeled as an instance, and otherwise it is one divided by the number of labels (from the cohort) it bears. For example, if an entity gets labeled as an instance of three concepts, then its typicality equals one-third for each of them and its typicality in all other concepts in the cohort equals zero by this measure.[11]

As pointed out above, C&C rendered the distance between concepts in terms of concept likelihoods. Lacking information on the positions of objects, a researcher likely cannot easily recover concept likelihoods. However, distances can be calculated using overlaps of memberships and a measure such as Jaccard distance.

Measurement of ambiguity requires empirical estimates of categorization probabilities. The standard Bayesian categorization rule depends on feature values. When these are not available, Hannan et al. (2019:175) recommends an approximation (see Olzak 2022 and Negro et al. 2022 for applications). However, we lack information about the quality of the approximation. Fortunately the strategy discussed next avoids the need for an approximation.

From feature values. Sometimes available data sources describe entities in terms of a set of feature values, for example, as sets of technical specifications. In such settings, analysts have assumed that agents use these specifications to categorize objects (see, for example, Smith 2011). Well-established dimensionality reduction techniques exist to allow the identification of the features that matter most for categorization decisions.

In many more cases, sources describe entities in natural-language texts rather than as sets of feature values. Consider, for example, the descriptions of several associations included in the *Encyclopedia of Associations: National Organizations of the U.S.* (Atterberry 2018) that are tagged with the subject "Environment" (possibly among others).

Anglers for Conservation:

> Strives to create a new generation of coastal stewards using community-based angling education, habitat restoration and applied conservation science. Educates the public in basic fishing skills and use of conservation-minded methods in order to protect the fish, their habitat and the angler.

Coalition on the Environment and Jewish Life:

> Represents Jewish organizations in their common aim to span the full spectrum of Jewish religious and communal life. Seeks to expand contemporary understanding of Jewish values. Serves as the voice of the organized Jewish community on environmental issues around the country. Aims to extend Jewish traditions as social action to environmental action and advocacy.

Environmentalists against War:

> Represents peace, social justice and environmental organizations. Advocates for environmental preservation and environmental justice. Researches and disseminates information on the human, social, and environmental impacts of war and militarism, at home and abroad.

The Nature Conservancy:

> Strives to prevent climate change and preserve biological diversity through protection of natural areas. Identifies ecologically significant lands and protects them through gift, purchase, or cooperative management agreements with government or private agencies, voluntary arrangements with private landowners, and cost-saving methods of protection. Provides long-term stewardship for 1340 conservancy-owned preserves and makes most conservancy lands available for non-destructive use on request by educational and scientific organizations.

World Peace One:

> Helps people make changes that improve the quality of life for all through various programs. Programs integrate five areas: personal mission and fulfillment; increasing personal capacity; empowering others; creating a world-sustaining lifestyle; and inviting others to participate in continuing this "chain-reaction" process.

Applying Deep-Learning Language Models

Suppose that one wants to assess the typicality of these and other associations as environmentalist as well as their ambiguity in terms of a cohort of other association concepts, for example, fraternal, occupational, religious, ideological, and so forth. In this and similar settings, the relevant feature space is not known a priori—it needs to be inferred from the data. Deep learning provides an effective solution to this challenge.

Deep learning refers to the process according to which the free parameters of a neural-network model are learned from data. Deep learning builds models for a language by constructing functions that take text documents as inputs and representing them as points in a (real-valued) space of hidden features. These functions are often represented as a vertical "stack" of linear functions ("layers") with some nonlinear intermediary steps ("activation functions"). In this context, "deep" means that the model has many layers, and "learning" means that the constants in the functions (often several millions) are learned from the data in a classic gradient-descent optimization. This procedure assumes the truth of a set of categorical assignments ("ground truth"), assigns a loss function such as mean squared error of prediction or cross entropy, and adjusts the constants in each step to minimize the loss.

In forming a prediction, the procedures generate vectors of what can reasonably be regarded as categorization probabilities for each input. If an output vector corresponds well to the judgement that humans make (see below), then this is exactly the information needed for measuring typicality and ambiguity,

A deep-learning language model has many free weights and thus needs to be "trained" on data to learn the parameter values that lead to the best possible categorization performance. For instance, state-of-the-art language models like BERT (Bidirectional Encoder Representations from Transformers) are generally trained on texts in which words are "masked" at random and the prediction task is to choose the correct word. (The text itself provides the ground truth.) The procedure checks the accuracy of the predictions and adjusts parameter settings in the direction of reducing error, and the process continues. Deep learning differs from other kinds of machine learning in that it infers a semantic space from the data, whereas other approaches generally require the analyst to specify the feature space. The core of deep learning is the automatic construction of the high-dimensional feature space in which texts are represented. This occurs via sophisticated algorithms that proceed by trial and error to maximize categorization accuracy for the training data.

In applications, the trained language model is supplemented with a set of human judgements, for example, categorizations of texts into genres. Based on a large amount of training data, the algorithm constructs a candidate feature space and specifies categorization probabilities (close parallels to $P(\mathfrak{c} \mid x)$). The performance of the model is assessed on the validation data in terms of categorization accuracy: for each text in the validation data, the model predicts its type (the one associated with its highest categorization probability). The proportion of correct predictions gives the performance of the model. The learning algorithms adjust the feature space and the concept likelihoods iteratively to improve model performance on the validation data. Once satisfactory performance has been achieved on the validation data, the model can be applied to the test data.

The class of deep-learning algorithms has achieved extraordinary performance on a large range of applications, most notably speech recognition and language translation. The current best of class, BERT and its extensions, have achieved accuracy levels that surpass humans on a number of tasks.

BERT and other contemporary methods develop language models in a space of very high dimensionality. For instance, the base version of BERT constructs its neural net in 768 dimensions, and the large version uses 1,024. It is unlikely that positions in such spaces have the same meaning as positions in an agent's semantic space (of presumably much lower dimensionality). If so, then likelihoods assigned to positions in the high-dimensional space will not serve as good proxies for concept likelihoods.

The fact that BERT and its offspring have achieved such accuracy in translation and reasoning tasks suggest that they construct neural networks that do a good job of representing human cognition. So it might be reasonable to presume that the typicalities that can be extracted from a trained model will closely match human performance. However, it appears that this has not been checked systematically. Le Mens et al. (2022) did such an examination in the realm of book genres. Using a sample of summary descriptions of books from Goodreads.com, they asked human subjects to indicate how typical was each of mystery and (with a different sample) of romance. Then they calculated typicality with respect to genres in several ways. One calculated typicality from categorization probabilities derived from BERT. The second based the calculations on probabilities derived from deep-learning training of a probabilistic classifier applied to the output of the popular GloVe (Global Vectors for Word Embedding) algorithm, used previously in a sociological application by Kozlowski, Taddy, and Evans (2019). The third used a "bag of words" (naive-Bayes) procedure that analyzes only word frequencies, used previously in a sociological application by DiMaggio, Nag, and Blei (2013). Finally they measured typicality with the several label-based measures of typicality discussed above. They found that BERT performed very well. When BERT was trained on the distributions of categorizations of mystery books, the correlation of its predictions with average human judgements was 0.90. For GloVe word embeddings combined with a deep-learning probabilistic classifier, this correlation was 0.79. For the bag-of-words approach and the label-based measure the correlation was 0.76. For romance books the results were very similar. These correlations reflect mainly the performance of these alternatives as classifiers; the results show that they all did a good job of distinguishing mysteries from other books. Importantly, BERT was superior to the other alternatives in matching human judgements about typicality among books whose dominant categorization by Goodreads.com users is mystery. In other words BERT excels at capturing the nuance of typicality judgements in this setting.

These preliminary results support my contention that measuring the values of relevant features gets us much closer to the audience perspective. In this investigation, the benchmark used was directed human judgements of typicality. Do the implications carry over to categorization probabilities? This has not been studied directly. However, use of the underlying probability model connects categorization probabilities to typicalities and estimates of the priors on the concepts. So it seems likely that a study that compared categorization probabilities calculated from BERT with human judgements of the likelihood that a description of a book or other entity is an instance of a genre would yield results similar to those obtained by Le Mens et al. (2022). If this is the case, then the measurement of conceptual ambiguity from categorization probabilities produced by BERT will also be sociologically meaningful.

Because the machine-learning approach is directed at predictions about entities positioned in a metric semantic space, less attention has been paid to the use of the positions themselves. The positions are unlikely to be directly interpretable because they are located in a very high-dimensional space and do not reflect the weighting of the dimensions learned in training. Nonetheless, Euclidean distances between entities are likely to bear sociological interpretation. One can locate the centers of clusters of entities categorized as instances of a genre and calculate their dispersion over the semantic space. So many standard sociological tools can be deployed with such data.[12]

Further development of this proposal for sociological analysis requires attention to several issues and complications.

First there is the issue of comprehensive labeling. The sketch of deep-learning language models might be seen as implying that the method can be applied only in situations in which all of the entities to be analyzed are associated with texts *and* categorizations. But this is not so. In translation, for instance, a trained language model can interpret texts that were not part of the training data. Typing an English sentence into Google Translate yields a translation to Chinese, say, even when the algorithm has never seen the sentence before. In the case of the experiments just discussed, all that is required is that the language model be fine-tuned on a sample of human categorizations for it to be applied to not-yet-categorized data.

This capability of machine-learning classifiers provides a strong advantage for the study of long-lived collectives. The meanings of concepts likely change over time. So it would not be useful to use categorizations made today to uncover yesterday's concepts. But, if research can obtain texts that characterize entities in a domain at some earlier time and categorizations made at that time, then the procedures I sketch here can arguably uncover the concepts of the time.

Finally issues of aggregation abound in this approach. Consider user ratings. In many cases, the researcher has access only to aggregated ratings along with textual descriptions and categorizations (possibly also the result of some kind of collective voting procedure, as I noted above in describing the data provided by Goodreads.com). Then one faces the same kinds of alternatives that I described for empirical measurement of memberships. More experimentation along these lines would be very useful in learning the possibilities and limits of such machine learning for sociological applications.

Discussion

The probabilistic approach advocated here offers important benefits. Building an explicit link between concepts and categories, via the Bayesian categorization function, clarifies issues of theory building and measurement. This article has tried to demonstrate this advantage by reconceptualizing form/genre and its membership as well as the paired notions of typicality and ambiguity. As these analyses show, following this approach allows for a more unified treatment of a broad range of sociological issues.

This framework also provides important benefits for improved measurement. It paves the way toward a shift from the prevailing label-based strategy for assessing membership in collectives to one based on the values of the features that matter for the audience. In other words, it points toward ways to measure semantic spaces and the positions of entities in these spaces. I have emphasized that the formal similarity of the structure proposed in C&C to the one at the core of deep-learning language models allows measurement of membership (typicality) from textual descriptions at very large scale.

I believe that this article's lessons for theory building and measurement have broad implications for other work in sociology that seeks to build connections to cognition. This includes such diverse fields such as cultural sociology (Cerulo et al. 2021; DiMaggio 1997; Mohr et al. 2020; Vaisey 2021), economic sociology (Vila-Henninger 2021), political sociology (Bonikowski, Luo, and Sthuler 2022), and stratification (Kozlowski et al. 2019). Pursuing the new approach might not only improve theory and measurement in these fields but also reveal deeper connections among them.

Notes

[1] Hannan (2022) provides a detailed recounting of the development of theoretical thinking about forms and populations in organizational ecology and expands on the proposal for revision discussed below.

[2] The terminology used in HPC differs considerably from that used in the most recent work. To avoid confusion, I use the contemporary terms here.

[3] I follow the convention of expressing terms in the "object language," in this case the language of the focal audience, in sans serif font.

[4] This last step requires the assumption of what C&C calls context independence of the concept likelihood. The idea is that there is nothing in the Bayesian representation of the categorization probability that requires that a concept remain stable over contexts. In each different context, a Bayesian agent could make categorization judgements for the same concept differently. This kind of flexibility might be realistic, but it runs against the idea that concepts simplify cognition. If a concept is stable over contexts (context independent), then one is justified in replacing $P_{\mathbb{F}_\mathfrak{c}}(x \mid \mathfrak{c})$ with $\pi_{\mathbb{F}_\mathfrak{c}}(x \mid \mathfrak{c})$.

[5] This possibility is reflected in the notation as conditioning of predicates by the focal audience.

[6] More precisely, this needs to be a non-uniform distribution because someone whose expectations can be represented as a uniform distribution over the space lacks the concept.

[7] Let P_1 and P_2 denote two discrete probability measures defined over a common space, $\mathbb{G}$. The Kullback–Leibler (KL) divergence of P_1 from P_2 is

$$D_{KL}(P_1 \,\|\, P_2) \equiv \sum_{x \in \mathbb{G}} P_1(x) \ln \frac{P_1(x)}{P_2(x)}.$$

[8] This measure does not satisfy the metric properties of symmetry and the triangle inequality. This does not cause any difficulty when the direction of comparison is given by the analytic purpose.

[9] These articles did not add the qualification that the concepts come from a cohort. Nonetheless, the treatments are consistent with this idea.

[10] The scaling is the division of the categorization probability by the sum of such probabilities over all concepts in the cohort.

[11] Kovács and Hannan (2015) proposed that a suitable measure of typicality should incorporate metric information about the distances among concepts assigned to an entity and suggested a simple generalization of the original measure of typicality.

[12] The most exciting possibilities would use this approach in comparative analysis, for example, comparing the dispersions of members of different groups. Such analysis will be challenging because the natural approach is to train a classifier for each group separately. If this is done, there is no way to constrain the learned space to be the

same for each. And training a classifier on the unions of possible instances of multiple concepts/groups might not yield a good classifier.

References

Anderson, John R. 1991. "The Adaptive Nature of Human Categorization." *Psychological Review* 98(3):409–29. https://psycnet.apa.org/doi/10.1037/0033-295X.98.3.409.

Atterberry, Tara E. (ed.). 2018. *Encyclopedia of Associations: National Organizations of the U.S.* Farmington Hills, MI: Gale Publishing.

Bonikowski, Bart, Yucjen Luo, and Oscar Sthuler. 2022. "Politics as Usual? Measuring Populism, Nationalism, and Authoritarianism in U.S. Presidential Campaigns (1952–2020) with Deep Neural Language Models." Posted January 24, 2022, on SocArXiv; revised July 26, 2022. https://doi.org/10.31235/osf.io/uhvbp.

Cerulo, Karen A., Vanina Leschziner, and Hana Sheperd. 2021. "Rethinking Culture and Cognition." *Annual Review of Sociology* 47:63–85. https://doi.org/10.1146/annurev-soc-072320-095202.

DiMaggio, Paul, Manish Nag, and David Blei. 2013. "Exploiting Affinities between Topic Modeling and the Sociological Perspective on Culture: Application to Newspaper Coverage of U.S. Government Arts Funding." *Poetics* 41(6):570–606.

DiMaggio, Paul J. 1997. "Culture and Cognition." *Annual Review of Sociology* 23:263–87. https://doi.org/10.1146/annurev.soc.23.1.263.

Fiedler, Klaus. 2012. "Meta-cognitive Myopia and the Dilemmas of Inductive Statistical Inference." In *Psychology of Learning and Motivation*, edited by Brian H. Ross, pp. 1–55.

Hannan, Michael T. 2010. "Partiality of Memberships in Categories and Audiences." *Annual Review of Sociology* 36:159–81. https://doi.org/10.1146/annurev-soc-021610-092336.

Hannan, Michael T. 2022. "Rethinking Organizational Ecology in Light of Developments in Cognitive Science and Natural Language Processing." Posted September 11, 2022, on SocArXiv; revised September 13, 2022. https://doi.org/10.31235/osf.io/vmuan.

Hannan, Michael T., and John Freeman. 1977. "The Population Ecology of Organizations." *American Journal of Sociology* 82(5):929–64. https://doi.org/10.1086/226424.

Hannan, Michael T., and John Freeman. 1984. "Structural Inertia and Organizational Change." *American Sociological Review* 49(2):149–65. https://doi.org/10.2307/2095567.

Hannan, Michael T., Gaël Le Mens, Greta Hsu, Balázs Kovács, Giacomo Negro, Lászlo Pólos, Elizabeth G. Pontikes, and Amanda J. Sharkey. 2019. *Concepts and Categories: Foundations for Sociological and Cultural Analysis*. New York: Columbia University Press.

Hannan, Michael T., László Pólos, and Glenn R. Carroll. 2007. *Logics of Organization Theory: Audiences, Codes, and Ecologies*. Princeton University Press.

Hsu, Greta. 2006. "Jacks of All Trades and Masters of None: Audiences' Reactions to Spanning Genres in Feature Film Production." *Administrative Science Quarterly* 51(3):420–50. https://doi.org/10.2189/asqu.51.3.420.

Hsu, Greta, Michael T. Hannan, and Özgeçan Koçak. 2009. "Multiple Category Memberships in Markets: An Integrative Theory and Two Empirical Tests." *American Sociological Review* 74(1):150–169. https://doi.org/10.1177/000312240907400108.

Koçak, Özgeçan, Michael T. Hannan, and Greta Hsu. 2014. "Emergence of Market Orders: Audience Interaction and Vanguard Influence." *Organization Studies* 35(5):765–790. https://doi.org/10.1177/0170840613511751.

Kovács, Balázs, and Michael T. Hannan. 2015. "Conceptual Spaces and the Consequences of Category Spanning." *Sociological Science* 2:252–86. https://doi.org/10.15195/v2.a13.

Kozlowski, Austin C., Matt Taddy, and James A. Evans. 2019. "The Geometry of Culture: Analyzing the Meaning of Class through Word Embeddings." *American Sociological Review* 84(5):905–49. https://doi.org/10.1177/0003122419877135.

Krafft, P. M., Erez Shmueli, Thomas L. Griffiths, Joshua B. Tenenbaum, and Alex "Sandy" Pentland. 2021. "Bayesian Collective Learning Emerges from Heuristic Social Learning." *Cognition* 212:104469. https://doi.org/10.1016/j.cognition.2020.104469.

Le Mens, Gaël, Balázs Kovács, Michael T. Hannan, and Guillem Pros. 2022. "Using Machine Learning to Uncover the Semantics of Concepts: How Well Do Typicality Measures Extracted from a BERT Text Match Human Judgments of Genre Typicality?" Created May 31, 2022, on Open Science Framework; updated November 10, 2022. https://doi.org/10.17605/OSF.IO/TA273.

Mohr, John W., Christopher A. Bail, Margaret Frye, Jennifer C. Lena, Omar Lizardo, Terrence E. McDonnell, Ann Mische, Iddo Tavory, and Frederick F. Wherry. 2020. *Measuring Culture*. New York: Columbia University Press.

Negro, Giacomo, Michael T. Hannan, and Susan Olzak. 2022. *Wine Markets: Genres and Identities*. New York: Columbia University Press.

Olzak, Susan. 2022. "The Impact of Ideological Ambiguity on Terrorist Organizations." *Journal of Conflict Resolution* 66(4–5):836–66. https://doi.org/10.1177/00220027211073921.

Rosch, Eleanor H. 1973. "On the Internal Structure of Perceptual and Semantic Categories." In *Cognitive Development and the Acquisition of Language*, edited by Timothy E. Moore, pp. 111–44. New York: Academic Press.

Rosch, Eleanor H. 1975. "Cognitive Representations of Semantic Categories." *Journal of Experimental Psychology: General* 104(3):192–233. https://psycnet.apa.org/doi/10.1037/0096-3445.104.3.192.

Smith, Edward Bishop. 2011. "Identities as Lenses: How Organizational Identity Affects Audiences' Evaluation of Organizational Performance." *Administrative Science Quarterly* 56(1):61–94. https://doi.org/10.2189/asqu.2011.56.1.061.

Tenenbaum, Joshua B., and Thomas L. Griffiths. 2002. "Generalization, Similarity, and Bayesian Inference." *Behavioral and Brain Sciences* 24(4):629–40. https://doi.org/10.1017/S0140525X01000061.

Vaisey, Stephen. 2021. "Welcome to the Real World: Escaping the Sociology of Culture and Cognition." *Sociological Forum* 36(S1):1297–315. https://doi.org/10.1111/socf.12770.

Vila-Henninger, Luis Antonio. 2021. "A Dual-Process Model of Economic Behavior: Using Culture and Cognition, Economic Sociology, and Neuroscience to Reconcile More and Self-Interested Economic Action." *Sociological Forum* 36(S1):1271–96. https://doi.org/10.1111/socf.12763.

Acknowledgments: I have drawn liberally from joint work with Glenn Carroll, Greta Hsu, Balázs Kovács, Gaël Le Mens, Giacomo Negro, Lászlo Pólos, Elizabeth Pontikes, and Amanda Sharkey. I thank them and Susan Olzak for their comments. They are not, of course, responsible for how I use their work here.

Michael T. Hannan: Graduate School of Business, Stanford University.
E-mail: hannan@stanford.edu.