



Kindhood and essentialism: Evidence from language

Katherine Ritchie^{a,b,*}, Joshua Knobe^c

^aDepartment of Philosophy, City College of New York, New York, NY, United States

^bProgram in Philosophy, The Graduate Center, CUNY, New York, NY, United States

^cProgram in Cognitive Science & Department of Philosophy, Yale University, New Haven, CT, United States

*Corresponding author: e-mail address: kritchie@ccny.cuny.edu

Contents

1. Kinds and essences	135
1.1 What are kinds?	135
1.2 What are essences?	138
1.3 What is the relationship between kinds and essences?	140
2. Semantic kind and category representations	142
2.1 Semantic evidence for a difference in kind between categories and kinds	143
2.2 Kind and category representations in semantic theories	148
3. Psychological kind and category representations	150
3.1 Stereotypes	152
3.2 Essentialism	153
3.3 Stereotypes and essentialism	157
Acknowledgments	160
References	160

Abstract

A large body of existing research suggests that people think very differently about categories that are seen as *kinds* (e.g., women) and categories that are not seen as kinds (e.g., people hanging out in the park right now). Drawing on work in linguistics, we suggest that people represent these two sorts of categories using fundamentally different representational formats. Categories that are not seen as kinds are simply represented as collections of individuals. By contrast, when it comes to kinds, people have two distinct representations: a representation of a collection of individual people and a representation of the kind itself. The distinction between these two representational formats helps to shed light on otherwise puzzling findings about stereotyping and essentialism. Stereotyping appears to involve a representation of a collection of people, while essentialism involves a representation of a kind itself.

It is an obvious and conspicuous fact about human cognition that we can represent people as falling into categories. We can represent a person as falling into the category of women, or the category of Irish people, or the category of conservatives. We can also come up with new categories on the spot that help us to achieve various purposes. Thus, we can represent a person as falling into the category of people hanging out in the park right now or the category of people who have seen the movie *Trainspotting*.

Yet, though we can represent people as falling into any of these categories, an enormous body of research indicates that there are deep and important differences between categories like the category of women and categories like the category of people hanging out in the park right now (Gelman, 2003, 2004; Gelman & Waxman, 2007; Markman, 1989; Prasada, Hennefield, & Otap, 2012; Shipley, 1993). In fact, as we will emphasize below, these categories differ from each other in two important respects. The first is that the category of women is represented as constituting a *kind*; the second is that the category of women is represented as having an *essence*.

Our aim here is to explore the nature of the representations people use for thinking about these different sorts of categories. On the hypothesis we will be defending, different sorts of categories truly are represented in different ways. In other words, it is not just that people hold certain beliefs about the category of women that they do not hold about the category of people in the park right now. Rather, there is a difference in the actual format of the representations used to think about these categories.

To make progress on these questions, we will be drawing on evidence from a somewhat unexpected source. In ordinary language, people refer to social categories by using linguistic expressions like “women” or “Irish people” or “people hanging out in the park right now.” Linguists have made some very surprising discoveries about plural noun phrases, which give us a much deeper understanding of what expressions like these actually mean. We will be arguing that these discoveries offer us real insight into the nature of people’s representations of social categories.

To foreshadow: Work in linguistics reveals that expressions like “Irish people” and “women” can be used in two very different ways. We will propose that this linguistic fact offers us a window into a fact about human cognition. In particular, the suggestion will be that people’s cognition includes two quite different representations of categories. The distinction between these two representations, we will argue, then allows for a deeper understanding of the way people ordinarily think about kinds and essences.



1. Kinds and essences

Although people can divide up human beings into categories in many different ways, they do not seem to treat all such categories equally. Instead, people seem to regard certain categories as getting at something deep or fundamental (e.g., the category of women) while regarding other categories as just superficial collections that do not get at something fundamental (e.g., the category of people hanging out in the park right now). A great deal of research has gone into understanding the differences between various sorts of categories.

Within recent work, a number of researchers have suggested that it may be helpful to adopt a more fine-grained approach (Bastian & Haslam, 2006; Haslam, Rothschild, & Ernst, 2000; Newman & Knobe, 2019; Noyes & Dunham, n.d.; Rhodes & Gelman, 2009; Yzerbyt, Corneille, & Estrada, 2001). On this approach, there is not just a single difference between the different categories. Rather, we need to distinguish several different dimensions on which categories can differ. A single category can then be high on one of these dimensions, but low on another. Research within this approach has identified a number of such dimensions (Haslam et al., 2000; Newman & Knobe, 2019; Noyes & Dunham, n.d.; Rhodes & Gelman, 2009; Yzerbyt et al., 2001).

For the purposes of this paper, we will be focusing especially on two dimensions along which categories can differ: the degree to which they are seen as constituting a *kind* and the degree to which they are seen as having an *essence*.

1.1 What are kinds?

A large quantity of research in cognitive science and linguistics has explored the ways in which people categorize nonhuman objects. This research indicates that although people can divide objects into categories in many different ways, only some of these categories are seen as constituting genuine *kinds* (Gelman, 2003; Markman, 1989; Prasada et al., 2012). For example, the category of raccoons is seen as constituting a kind, and so is the category of oak trees. By contrast, the category consisting of white animals (swans, goats, great white sharks, etc.) is not seen as constituting a kind. Similarly, the category of staplers constitutes a kind, and so does the category of paper clips, but the category of all the things on your desk right now (paper clips, computer, half-eaten muffin, etc.) does not constitute a kind.

An important question arises as to what features people impute to categories they represent as kinds. Within existing research, it has been suggested that kinds are represented as *cohesive*, as having *inductive potential*, as being *informative*, as having relatively *stable* membership, and as figuring in certain forms of *explanation* (Carlson, 1977; Gelman, 2003; Markman, 1989; Noyes & Dunham, n.d.; Prasada et al., 2012; Prasada & Dillingham, 2006, 2009; Rothbart & Taylor, 1992). For instance llamas are represented as a cohesive category with members that have many things in common, so finding out that some llama has a thick fur coat makes one apt to infer that other llamas do as well. Similarly, in learning that something is a llama one thereby seems to learn lots of other information about it. Something's being a llama is a stable property. An animal that is a llama is a llama for life. Finally, answering that something behaves in a particular way *because* it is a llama is a reasonable (albeit not very informative!) explanation (Howard, Wagner, Carey, & Prasada, 2018; Prasada, 2017; Prasada & Dillingham, 2006, 2009). By contrast, the category of things that are on your desk right now is not represented as at all cohesive, and if we learn that a specific object is on your desk right now, this will explain almost none of its properties. For this reason, the category of llamas is seen as constituting a kind, while the category of things on your desk right now is not.

The same distinction can then be applied to the ways people represent categories of human beings (henceforth "human categories"). Consider again the contrast between the category of women and the category of people who are in the park right now. People tend to regard the former as cohesive, inductively potent, informative, stable, and explanatory, while they would presumably regard the latter as having none of these features. Hence, the former will be seen as constituting a kind, and the latter will not. We note that not all kind membership is seen as stable to the same degree (e.g., a person is not born a psychologist and after becoming a psychologist she can change careers), so kind membership is not necessarily immutable. Nevertheless, it is natural to take someone's membership in a kind to be relatively persistent.

Importantly, existing research consistently finds that being perceived as constituting a kind is not just the same as being perceived as having a biological basis. A number of different studies have explored the relationship between these two types of judgment by looking at the way in which people's answers to different questions covary across different categories (Haslam et al., 2000; Noyes & Dunham, n.d.). In these studies, participants are asked to rate a variety of different categories on questions that tap into the

notion of having a biological basis, such as the degree to which membership in each category is natural and the degree to which it is immutable. They are also asked to rate each category on a number of different questions that tap into kindhood, such as the degree to which members of the category are similar to each other and the degree to which learning that someone was a member of the category would be informative. Factor analysis consistently reveals that these two types of questions load on two distinct factors (Haslam et al., 2000; Noyes & Dunham, n.d.). In other words, people seem to have intuitions about the degree to which a category constitutes a kind (as measured by questions about similarity, informativeness, etc.) that are distinct from their intuitions about the degree to which the category has a biological basis.

As we understand the notion, then, people can divide human beings up into kinds in a number of different ways, including genders (men, women), nationalities (Albanians, Koreans), racial groups (Black people, Latinx), and professions (mathematicians, plumbers). Though some of these categories are seen as natural or biological and others are not, they are all seen as constituting kinds. By contrast, the category of all the people eating Cheerios for breakfast right now does not constitute a kind. Similarly, the intersection of a number of different categories that each constitutes a kind will not usually itself be seen as constituting a kind. Thus, people might think that the categories of gay people, Albanians, and mathematicians each constitute kinds, but the category of gay Albanian mathematicians might not itself be seen as constituting a kind.

In some cases, there may be controversy about whether a particular human category constitutes a kind, and these controversies sometimes play a central role in issues surrounding prejudice and intergroup conflict. For example, some people think that the category of Indian people constitutes a kind, while others think that Indian Hindus and Indian Muslims are so radically different from each other that they do not together constitute a single kind. Similarly, developmental data suggest that whether racial categories are seen as kinds might vary by age and race (Kinzler & Dautel, 2012; Roberts & Gelman, 2016). For example, white English-speaking 5–6-year-olds judged that a white English-speaking child was more likely to grow up to be a Black English speaker, than a white French speaker. In contrast African American 5–6-year-olds, like older white children and adults, judged that the child was more likely to grow up to be white, even though that involved speaking a different language (Kinzler & Dautel, 2012). These differences in stability judgments could reveal that whether race is represented as a kind can vary based on age and socialization.

Finally, existing research has documented a connection between the use of nouns and being perceived as a kind. Note first that English has a noun “raccoon” and a noun “plumber,” but it has no noun that means “white animal” or “gay Albanian mathematician.” Existing research suggests that categories picked out by nouns are indeed more likely to be seen as constituting kinds and to serve as an inductive base (Gelman & Coley, 1990; Gelman & Markman, 1986, 1987; Markman, 1989; Prasada et al., 2012; Waxman, 2010; Waxman & Gelman, 2009; Waxman & Markow, 1995). Thus, when a given category is picked out by a noun (vs a longer phrase), people are more likely to describe it as a “kind” of some superordinate category (Prasada et al., 2012). Developmental studies show that children are more inclined to see two objects as similar when they are described with the same noun than when they are described with the same adjective (Gelman & Coley, 1990; Gelman & Heyman, 1999). Studies with adults show that preferences and traits described with nouns were seen as more persistent and stable than those described with adjectives (Reynaert & Gelman, 2007; Walton & Banaji, 2004). Finally, in philosophy, it has recently been argued that this psychological research reveals important features that should be captured by a theory of meaning (Neufeld, 2019; Ritchie, n.d.). In short, existing research points to an important connection between nouns and thinking of a category as constituting a kind.

1.2 What are essences?

Consider the kind tigers. People associate this kind with certain relatively superficial observable features (striped, four-legged, ferocious, etc.). Yet studies consistently find that people think that being a member of this kind is not just a matter of having these observable features. Instead, they think that being a member of the kind is a matter of having some deeper, internal, unobservable property (Bloom, 2004; Gelman, 2003; Keil, 1992). This deeper property is usually known as an *essence*.

Existing research has uncovered a variety of important facts about how people ordinarily understand essences. Essences are normally seen as *hidden* (Medin & Ortony, 1989). In fact, even when people are completely sure that a category has an essence, they usually think that they have no idea at all what that essence actually is. Essences are usually seen as *internal* (Cimpian, 2015; Gelman, 2003; Gelman & Wellman, 1991; Newman & Keil, 2008). Thus, the essence of a tiger is something internal to the tiger; it is not something about, for example, the tiger’s role in a larger natural or social structure. Finally, essences are normally seen as in some way *determining category*

membership (Gelman, 2003; Keil, 1992). In some important sense, tiger essence, rather than superficial features, is what makes some things tigers and other things not tigers.

One important lesson from existing research is that some categories are seen as having essences while others are seen as not having essences (Gelman, 2003; Markman, 1989; Rhodes & Gelman, 2009). For a clear example of the latter sort of category, consider the category of office supplies. People think that certain objects count as office supplies and that others do not. However, people presumably do not think that there is some deeper, unobservable, internal property that ultimately makes something an office supply. Rather, they think that being an office supply is just a matter of having a certain kind of role in offices. In short, this is a category that people do not essentialize.

Research indicates that the distinction between essentialized categories and nonessentialized categories can teach us a lot about how people understand human categories (Allport, 1954; Dar-Nimrod & Heine, 2011; Gil-White, 2001; Haslam et al., 2000; Haslam & Levy, 2006; Haslam & Loughnan, 2014; Mahalingam, 2007; Rothbart & Taylor, 1992; Williams & Eberhardt, 2008). Consider the contrast between the category of women and the category of waiters. Many people seem to think that there is a deeper essence associated with the category of women (Brescoll & Uhlmann, 2008; Brescoll, Uhlmann, & Newman, 2013; Martin & Parker, 1995; Morton, Postmes, Haslam, & Hornsey, 2009; Prentice & Miller, 2006). That is, they think that women can ordinarily be picked out using certain superficial features but that there is some deeper, unobservable biological property internal to women that is what really makes a person a woman. By contrast, people do not tend to think that there is a deeper essence to being a waiter. Being a waiter is simply a matter of holding a particular job, which is a superficial feature. Anyone who has that feature is a full-fledged member of the category, and category membership does not depend in any way on the possession of any deeper, unobservable internal property.

Now, clearly, one way in which a category can be seen as having an essence is by being seen as having a biological basis. If people think that membership in a category is fundamentally a matter of having a certain sort of DNA (or some other similar biological property), then by that very fact, they think that there is a deeper causal factor that is hidden, internal, and determines category membership. A question arises, however, as to whether it is possible for a human category to be seen as having an essence without being seen as having a biological basis. Though different researchers might disagree when it comes to this question, we will suggest that it is indeed possible.

For example, consider the category of Christians. Presumably, people do not think that this category has a biological basis. Yet, all the same, people's attitude toward it does show many of the hallmarks of essentialism. People may think that the category is associated with certain superficial features (going to church, saying grace) but that ultimately, what it really means to be a Christian is a matter of something deeper and less easily observable (embodying the true values of Christianity). In what follows, we will assume that this type of attitude is also an instance of essentialism (see [Newman & Knobe, 2019](#)). However, this assumption plays only a small role in our argument. If it turns out that we are mistaken about this point and essentialism always involves positing a biological or other natural basis, most of what we say in what follows will still go through.

Finally, we should note that essentialism has been implicated not only in people's judgments about categories but also in people's judgments about *individuals* (e.g., [Christy, Schlegel, & Cimpian, 2019](#); [De Freitas, Tobia, Newman, & Knobe, 2017](#)). Thus, just as people think that there is an essence of certain categories (an essence of raccoons, an essence of women), people tend to think that there is an essence of certain individual entities (an essence of the Mona Lisa, an essence of Martin Luther King, Jr.). Judgments about essences of individual entities are in many ways similar to judgments about essences of categories. The essence of Martin Luther King, Jr. is seen as something that is internal, hidden and, in a specific sense, *identity determining*. That is, if MLK got into a car accident and the hidden internal properties that people see as constituting his essence were lost, something might be left over after the accident, but whatever was left over would not be regarded as genuinely MLK himself.

Ideally a theory of essentialism should explain both essentialism about categories and essentialism about individuals. We return to this point below.

1.3 What is the relationship between kinds and essences?

In sum, there is an important distinction between seeing a category as constituting a kind and believing that the category has a deeper essence. Categories that are seen as constituting kinds are taken to have the cohesiveness or regularity required for inferring and explaining. In some cases, people may think that a kind has a deeper essence, but in other cases, they see a category as constituting a kind without believing in any corresponding essence. Take the category of people from a particular nation or the category of people who have a particular profession. Such a category might be

thought to display a great deal of regularity and cohesiveness, but this regularity and cohesiveness might not be thought to arise because of a deeper hidden causal factor within all of the people in these categories. Instead, it might simply be a matter of certain relatively superficial external factors (the culture of the nation, the requirements of the job).

Yet, despite these important respects in which the two notions can come apart, it also seems that there is a deeply important connection between them. Any adequate account must therefore be able to not only help us understand the differences between these two notions but also their interconnections.

As one way into this problem, consider the use of generics. Generics are sentences that express generalizations about an entire category. People can use such sentences to express generalizations about categories of artifacts (Lamborghinis are expensive) or categories of animals (Boa constrictors are dangerous). Importantly, people can also use generics to express generalizations about human categories. A question now arises as to whether generics about human categories have any relation to kindhood or to essentialism.

At least at first, it might seem that the use of generics is tightly connected with perceptions of kindhood but has nothing to do with essentialism. For example, consider the generic sentences:

- (1) Psychologists know more about statistics than they do about calculus.
- (2) Brazilians speak Portuguese.

Neither sentence seems to involve any sort of essentialism. For example, (1) does not in any way suggest that there is some deeper hidden thing within all psychologists that draws them toward statistics rather than calculus. The obvious interpretation of (1) is the nonessentialist one on which the speaker is simply saying that people who have this particular job end up learning more about some mathematical techniques than others.

A number of researchers have, however, taken generic sentences to suggest the existence of a kind (Asher & Pelletier, 2013; Carlson, 1977; Chierchia, 1998; Cohen, 1999; Liebesman, 2011; Teichman, 2016; for exceptions and an alternative view see Nickel, 2016, 2017). For example, suppose that a speaker begins a generic sentence with the words:

- (3) Tall female Ecuadorian mathematicians ...

Regardless of what comes next, it would be natural to infer that this speaker holds the somewhat peculiar belief that tall female Ecuadorian mathematicians constitute a kind. In other words, it would not normally be felicitous to use this sentence if the speaker believed that tall female Ecuadorian

mathematicians did not constitute a kind but did happen to have some particular property in common.

One might expect, therefore, that empirical research would show that hearing generic sentences about a category would lead people to see that category as constituting a kind but would have no effect on essentialism. But that is not, in fact, the case. Strikingly, some research has found that hearing generic sentences about a category actually led to increased essentialism about that category (as measured by, e.g., their responses to questions involving causal explanation) (Cimpian & Markman, 2011; Rhodes, Leslie, & Tworek, 2012). Yet, other recent research has suggested that not all generics increase essentialism (Hoicka, Saul, Prouten, Whitehead, & Sterken, n.d.; Noyes & Keil, 2019).

To explain these phenomena, we will be developing an account on which there is a close connection between kindhood and essentialism. On this account, it is possible to see a category as constituting a kind while not seeing it as having an essence, but all the same, there is an important sense in which seeing a category as constituting a kind makes people more inclined to essentialize it.



2. Semantic kind and category representations

To make progress on these questions, we turn to a source of evidence that has not figured heavily within existing research in psychology on the representation of social categories. Specifically, we turn to work in linguistics on the meanings of expressions used to pick out human categories. Our basic strategy here is to consider what research that is straightforwardly concerned with the meanings of linguistic expressions like “women” or “tall Ecuadorian mathematicians” might reveal about people’s representations of categories like the category of women or the category of tall Ecuadorian mathematicians. Linguists have made some very surprising discoveries about the meanings of these expressions, and we will be arguing that these discoveries shed important new light on the corresponding psychological questions.

Research in this area tends to be fairly technical, but we think that what is emerging out of all of these technical details is a larger message that has important implications for the psychology of people’s social category representations. Accordingly, we begin with a brief and very informal statement of the larger point. We then turn to a slightly more detailed discussion of the

actual linguistic evidence that supports our main point and a more formal framework that has been used to capture it.

2.1 Semantic evidence for a difference in kind between categories and kinds

To begin with, consider expressions that are used to pick out categories that are seen as constituting human kinds. These include expressions like “women,” “Jewish people,” and “Albanians.” Research on these expressions reveals that they can be used in two very different ways.

Used in one way, they basically just pick out the individual human beings within a category. So, for example, consider the individual human beings who fall within the category of women: Hillary Clinton, Serena Williams, Frida Kahlo, and so forth. In a sense that we will later make more precise, one way of using the word “women” just picks out all of these individual human beings.

But these expressions also have another, very different use. In this other use, they do not just pick out a bunch of individual human beings. Rather, they refer to the kind itself. When we use the expressions in this way, we can discuss facts about the kind that are not facts about any of the individual human beings within the kind (e.g., facts about how the kind evolved).

Now, by contrast, consider expressions that pick out categories that are not seen as constituting a kind (e.g., “tall Ecuadorian mathematicians”). These expressions can only be used in one way. They can be used to pick out all of the individual human beings in the category, but they cannot be used to pick out a kind itself. Since people think that there is no underlying kind here, there is no expression in the language that refers to it. There is no *it* to refer to at all! Hence, our language makes it possible to state certain claims about categories that constitute kinds, but our language does not allow us to state in an analogous way the corresponding claims about categories that do not constitute kinds (Carlson, 1982; Chierchia, 1998; Krifka et al., 1995).

Perhaps the best way to bring out what is so striking about all of this is to contrast it with what one might have expected prior to looking at the linguistic data. Prior to looking at the data, one might have expected something like this: Language contains certain expressions that provide a way to refer to human categories. These categories can then be represented as having various different properties, including the property of being a kind.

As we will show below, this is not how things actually work. Instead, things work more like this: Language does not just allow us to refer to

one type of thing that could be called a “human category.” Instead, language allows us to refer to two very different types of things: (a) a collection of individual human beings and (b) a kind itself. When we are referring to a bunch of individual human beings, we can do it in more or less the same way regardless of whether those human beings constitute a kind or not. Separately, language also allows us to refer to kinds themselves. This allows us to make claims about a kind that are not also claims about the individual human beings that fall within it. When a category is not seen as constituting a kind, it is not that we do this in a different way; we simply cannot do it at all.

To back up these claims, we now turn to the linguistic data. We will be focused especially on data concerning bare plural expressions for kinds and categories of human beings (“women,” “people in the park right now”), but we will also talk about expressions for kinds and categories of nonhuman entities (“birds,” “things on my desk right now”). The key points hold in the same way in both cases.

2.1.1 *Predicates that apply to kinds*

Consider the following sentences:

- (4) Human beings use tools to alter their environment.
- (5) Human beings evolved approximately 200,000 years ago.

The truth of (4) relies on individual human beings doing certain things. For instance, that John uses tools to alter his environment, and Sabina does, and Francesca does too. In contrast, (5) does not say anything about John, Sabina, Francesca, or any other particular humans. It is not about individual human beings at all. Rather, it says that the kind *homo sapiens* evolved. This brings us to an important distinction. Some predicates only apply to individuals and others only apply to kinds. “Evolved” is a predicate that only applies to kinds. Other examples of so-called kind predicates are “extinct,” “invent,” “exterminate,” “be a kind of animal/tool/...,” “be a racial group,” and “be a gender category.”^a

Some predicates can apply to either individuals or kinds. For instance, consider “have a short lifespan” as used in (6).

- (6) 8-track tapes had a short lifespan.

This sentence can be interpreted as saying that individual tapes each generally have a short lifespan, but it can also be interpreted as making a claim

^a Some also classify predicates like “rare” and “widespread” as kind predicates (e.g., Leslie & Lerner, 2016). Due to differences with mere category constructions, we take these to fall in a different class (following Carlson, 1982; Krifka et al., 1995).

about the kind itself. On this latter interpretation, it makes a claim about the amount of time the kind had instances. (For example, we might judge that the “lifespan” of the 8-track tape was short as it was less than 20 years.) These data reveal the first of two important features for our purposes. Predicates applied to bare plurals can rely on either *individual members* (as in 4) or to the *kind itself* (as in 5) (Chierchia, 1998; Krifka, 1987; Krifka, 2003; see Krifka et al., 1995; Leslie & Lerner, 2016, for overviews).

Second, kind predicates show that there is a strong contrast between expressions that pick out kinds and those that pick out mere categories. For instance, both (7) and (8) are odd.

- (7) People currently living in New Haven evolved approximately 200,000 years ago.
 (8) Sleepy people who wear tracksuits and love algorithms are a kind of person.

Linguistic examples show that predicates can apply to individuals and to kinds. Kind predicates support two important take-home messages. First, kinds are not just collections of their members. Second, kinds and categories are distinct.

2.1.2 Counting and quantification

Next, consider sentences like (9) and (10).

- (9) We heard four different birds in the park today.
 (10) Sam saw every animal at the zoo.

These sentences could be true in two ways (Krifka et al., 1995). First, (9) might mean that, for instance, the speaker and her bird-loving companions heard four individual birds—one that was perched in a tall tree, one sitting on the back of a bench, and so on. Alternatively, it might mean that they heard four different birdsongs. They heard the call of the mockingbird, the wren, the owl, and the catbird. Similarly (10) could be true if Sam saw each and every individual giraffe and baboon and lion and so on. Or, it could require that she saw at least one instance of each kind of animal at the zoo. The former involves instances of a kind, the latter involves kinds themselves.

In addition, examples involving counting and quantification show that kinds are not mere categories. Sentences (9) or (10) could be followed up by lists of expressions that refer to kinds, like *wrens* and *mockingbirds* or *giraffes* and *zebras*. In contrast, (11) sounds decidedly odd.

- (11) We heard four different birds in the park today—namely flying birds, brown birds, birds sitting in trees, and owls.

The only sensible interpretation of (11) involves supposing that the speaker takes the categories picked out in the list to each constitute a kind. Since it is odd to think of flying birds, brown birds, or birds sitting in trees as categories that constitute kinds, the sentence is strange.

2.1.3 *Pronouns*

Consider the following sentences:

(12) Jehovah's witnesses are a Christian denominational group first established in the 1870s. They are required to proselytize.

(13) A Jehovah's witness knocked on my door last night and tried to talk with me about faith. They are always doing that sort of thing.

In (12), the social kind or group Jehovah's Witnesses is introduced in the first sentence. In the second sentence of (12), "they" is used to pick out that very same group. In semantic jargon, one says that the introduction of the group licenses the use of the pronoun. Sentence (13) is more interesting. The first sentence is about a particular person. It is not about the kind being established or being a certain sort of group. Rather, it is about a specific person carrying out a specific concrete action. Yet, it still licenses a plural pronoun that picks out the social group or kind. This reveals that kinds can be brought to salience even when only a single member has been referred to (for further discussion, see [Carlson, 1977, 1982](#); [Gelman, 2003](#), ch. 8; [Krifka et al., 1995](#); [Roberts, 2004](#); for experimental support see [Waxman, 2010](#)).

Next, consider examples with expressions for members of categories that do not intuitively constitute kinds.

(14) A long-haired Brazilian woman knocked on my door last night and tried to talk with me about faith. They are always doing that sort of thing.

(15) A short person asked me to reach something on a high shelf. They have a lot of trouble in grocery stores.

The pronouns in (14) and (15) do not easily allow for a kind level reading. The only felicitous interpretation of (14) relies on a singular gender-neutral use of "they" in which a particular person is said to have a habit of knocking on the speaker's door. While (15) might allow for a reading on which "they" picks out the category of short people, it is significantly more strained than the kind interpretations in (12) and (13). The contrast between the (12) and (13), on the one hand, and (14) and (15), on the other, further supports the view that kinds are represented as something more than their members and that kinds are not mere categories.

2.1.4 Attitudes reports

Finally, consider the attitude reports in (16) and (17).

(16) Andy hates some gay people.

(17) Andy hates gay people.

Sentence (16) reports on Andy's hatred of some people who happen to be gay. In contrast, (17) is not true if it just so happens that Andy hates several people who are gay. Rather, it seems to require that he hates people who are gay because they are gay. As Andy might put it, he hates that *kind* of person (see [Krifka et al., 1995](#), pp. 70–72 for further discussion).

To further draw out the difference, consider how Andy's knowledge could affect the truth of these attitude reports. Suppose that Andy hates Charlie and Molly and that both are gay, but Andy has no beliefs about their sexual orientations. In this circumstance, it just so happens that there are some gay people Andy hates. That is sufficient for (16) to be true on one reading. Now, add on to the previous story that Andy has no attitudes whatsoever about sexual orientation and that Charlie and Molly are the only gay people. In this scenario Andy hates the only people who are gay, but he does not know that they are gay and has no attitudes about any sexual orientation being good or bad. In this case, (16) is true, but (17) is false.^b

Next, consider attitude reports with expressions for mere categories.

(18) Scott loves several Americans who speak Portuguese and drink black coffee.

(19) Scott loves Americans who speak Portuguese and drink black coffee.

Notice that while (18) sounds natural, (19) would be a rather odd thing to say. Why is that? On the view we suggest, it is because it requires Scott to represent the category of Americans who speak Portuguese and drink black coffee as a kind. While it is possible for one to have this view, it is rather strange. This helps to explain the oddness of (19).

Data involving attitude reports again show that kinds are not identical to their members. If they were, readings on which sentences like (17) and (19) are true when the attitude holder has beliefs about *some* people who happen to have a particular feature would be possible. Yet, we saw that bare plurals in attitude reports require more than that. They require, we suggest,

^b The sentence in (16) might imply that Andy is homophobic, but we note that its truth does not *require* that Andy have homophobic attitudes. It would be natural to conclude that Andy is homophobic from an utterance of (16) on the view that cooperative speakers utter things that are relevant to the conversation ([Grice, 1975](#)). Why, one might ask, would one describe the people Andy hates as "some gay people" if Andy did not dislike them due to their sexual orientation? Nevertheless, (16) can be true, if also misleading, in scenarios in which Andy has no homophobic attitudes.

attitudes toward kinds themselves. These data show yet again that expressions that pick out kinds and those that pick out mere categories differ. In this instance, they differ in their naturalness in attitude reports.

We have seen that semantic data converge on two major points. First, kind expressions rely on members of kinds in some instances and kinds themselves in others. Second, expressions for kinds and for mere categories differ. Importantly for our purposes, while we have evidence that kinds are represented as entities, we do not find similar evidence that mere categories are represented as entities. Next we briefly sketch the outlines of a more formal semantic theory that can capture these two features.

2.2 Kind and category representations in semantic theories

At the core of semantic theorizing is the idea that expressions are assigned semantic values of different *types*. Some just refer straightforwardly to entities. For instance “Barack Obama” refers to Barack Obama. It is said to be of type *e*, which stands for entity. Other expressions are assigned more complex types. For instance consider the predicate “be intelligent.” This predicate does not refer to one individual entity. Rather, it plays a more complex role in our language. At an intuitive level, one might explain how this predicate works by saying that it is true of some entities (Obama, you, Marie Curie, Albert Einstein) and false of other entities (New York, the Golden Gate Bridge, the book closest to you right now).

Within more formal work in semantics, researchers spell out this intuition by saying that the semantic value of predicates like this one are *functions*, relations that takes inputs and deliver outputs. “Be intelligent” is a function that takes entities as inputs (things of type *e*) and returns truth values as outputs (things of type *t*). The function could be represented by listing inputs on the left of an arrow and outputs on the right.

Barack Obama \rightarrow True.

Marie Curie \rightarrow True

New York \rightarrow False.

...

Other expressions have even more complex functional types.

So far, we have the following pieces. First, we have the idea that a semantic theory has expressions that can refer to various entities: you, Manuel, Cillian, Iluka, and so forth. These are possible referents of expressions of type *e*. Second, we have the idea that there are functions that take entities as inputs and yield truth values as outputs. Now comes the key question.

Within this basic framework, how can we understand the two possible ways of using ordinary expressions for human categories?

When it comes to the use that picks out a collection of different individual human beings, the answer is straightforward. Just like we saw with the example “be intelligent,” the semantic value for an expression like “elderly Japanese taxi drivers” specifies a function and outputs true for just those entities that are elderly Japanese taxi drivers. Similarly, when an expression like “women” is used to pick out a collection of individual human beings, its semantic value is a function that outputs true just for those entities that are women.^c

How then are we to understand the use in which these expressions pick out a kind itself? It might at first seem that this use, too, involves some kind of function from entities to truth values, but, strikingly, the linguistic data we saw above suggest that this is mistaken. Rather, as Chierchia (1998) has influentially argued, the linguistic data point to the view that kinds are represented as *entities*. In other words, a kind is not simply represented as a function that determines a collection of category members; it is represented as a specific sort of thing.

In short, a kind expression like “women” can mean two things. It can have a semantic value that is a function from entities to truth values and a semantic value that is an entity. These two meanings are obviously closely related. They are not like the unrelated meanings of “duck” (a crouching action) and “duck” (an aquatic bird). To capture the way kind expressions can mean two things while showing that the meanings are intimately connected, Chierchia (1998) posits special devices that can “shift” an expression from meaning one thing to meaning the other (see also Cohen, 2007; Krifka, 2003). He posits an up-operator (symbolized as \cup) that shifts from a kind, which is of the lower type e , up to the property of being a member of a kind, which is of the same sort of functional type as “be intelligent.” Then he posits a down-operator that goes in the opposite direction (symbolized as \cap). Together these two operators allow for a semantic theory to provide the requisite kind and the collection of members interpretations of bare plural expressions.

^c Technically these would be the semantic values for “elderly Japanese taxi driver” and “woman” in the singular. Semanticists usually take pluralization to involve forming a more complex set that also includes sums or sets of individuals. So, the plural “elderly Japanese taxi drivers” would not just pick out a set of individuals {Haruki, Kazuo, Yu, ...}, but a set that includes both individuals and sums or sets of individuals as in {Haruki, Kazuo, Yu, Haruki + Kazuo, Kazuo + Yu, ...} (see Landman, 1989; Link, 1983; Schwarzschild, 1996).



3. Psychological kind and category representations

In the previous section, we introduced a framework that has proven helpful in understanding the language people use to talk about kinds. Our principal aim, however, is not to understand something about language, it is to understand how people *think* about kinds. Thus, consider a case in which people do not utter any sentences or use any words but simply have a representation of some kind. We want to understand what is happening in such cases.

The hypothesis we want to propose is that the very same framework that has proven so helpful in theorizing about kind reference in language can also give us insight into questions about how people represent kinds in thought. We will be suggesting that, when it comes to psychological questions about people's mental representations of kinds (essentialism, stereotyping, etc.), we can make progress by using ideas that were first introduced to address straightforward linguistic questions about how people use certain words and phrases (bare plurals, quantifiers, pronouns). Of course, in saying this, we do not mean to suggest that people's mental representations of kinds in any way make use of this sort of language. Our point is an entirely different one. We are suggesting that work on these linguistic phenomena gives us valuable evidence about the structure of people's cognition. We want to leverage this evidence to make progress on psychological questions that are not themselves directly about language.

Let us turn now to these psychological questions. Existing research suggests that people's way of thinking about categories that are seen as constituting kinds (women, Albanians, plumbers) is different from their way of thinking about categories that are not seen as constituting kinds (people in this park, historians who like pizza but do not like beer). How are we to understand the representation of this distinction in people's cognition?

Here is one possibility: People have a way of representing categories. They can represent categories as having various different properties. One such property is the property of being a kind. The more people see a category as having this property, the more inclined they are to see it as also having various other properties. For example, if people see a category as a kind, they might be more inclined to think the category has an essence.

This is the view we want to reject. We will be arguing against it and in favor of an alternative view inspired by linguistic research on semantics. Here is the alternative view: People have a way of representing collections

of human beings. They think that some collections of human beings constitute kinds. In such cases, they actually have two different representations. One is a representation of the various human beings; the other is a representation of the kind itself. These two representations make it possible to represent different sorts of properties. People can represent the collection of human beings as having certain properties, and they can also represent the kind itself as having certain properties (see also [Prasada et al. \(2012\)](#) for arguments for distinct representations of kinds and categories).

Now, we noted above that existing evidence suggests that people are more inclined to regard a category as a kind to the extent that they believe the category to have certain properties (cohesiveness, stability, explanatoriness). But on the view we are proposing here, it would be a mistake to say that regarding something as a kind simply *is* a matter of believing it to have these properties. Rather, while having such beliefs increases the probability that something will be regarded as a kind, they are not necessary. Indeed, a person need not even have the ability to track these other features in order to have a kind representation.

In explaining what a kind representation is, it might be helpful to introduce a simple metaphor. Imagine that a person's mind contains a folder for each thing she regards as an entity. So she has folders for various individual human beings in which she keeps track of the properties of those human beings (e.g., a folder for Manuel in which she represents that he is a human being). Then she also has folders for various kinds in which she keeps track of the properties of those kinds themselves (e.g., a folder for the kind *human beings* in which she represents that they first evolved 200,000 years ago). And then, of course, there will be many categories for which she has no folder at all. For example, she will not have a separate folder labeled *mathematicians who like pizza but do not like beer* in which she records the properties of that category.^d

Within this metaphor, how would one understand the notion of representing something as a kind? Our answer is that it would not have anything to do with the properties one represented something as having (cohesiveness, stability, explanatoriness, or whatever). Rather, to represent a category as a kind would simply be to have a folder for it. Or, to put it in less

^d In this metaphor individual and kind representations are both represented with folders. This does not mean that we take there to be no differences between these sorts of representations. Just as there are accordion folders and file folders, there might be important differences between kind and individual "folders." Our point is that kinds and individuals pattern together in a way that sets them apart from mere categories.

metaphorical terms, regarding a category as a kind is a matter of representing it in the distinctive format characteristically used for entities.

On this view, if people do not see a collection of human beings as constituting a kind, certain questions simply do not arise. Suppose there is some property that people can attribute to kinds. Now suppose that people think that the set of all Albanians constitutes a kind. A question will then arise as to whether they represent this kind—the kind *Albanians*—as having this property. By contrast, suppose people think that the set of mathematicians who like pizza but do not like beer does not constitute a kind. In that case, there just is not any question as to whether they represent this kind as having the property. They do not represent it either as having the property or as lacking the property—they simply do not represent it as an entity at all.

If this framework is on the right track, we face new theoretical questions that might not otherwise have been salient. Existing psychological work has investigated various representations that people can have about human categories. For example, people can have *stereotypes* about categories and they can *essentialize* categories. For each of these, we can now ask a new question. Does it involve a representation of a collection of entities or a representation of a kind itself?

3.1 Stereotypes

Let's begin with a question about stereotypes. Paradigm cases of the phenomenon we are trying to understand here include the stereotype that women are bad drivers or the stereotype that Jewish people are neurotic. What we want to know is whether these stereotypes involve representations of the individual human beings in the kind or representations of the kind itself.

To better understand what question we are asking, recall example 6, repeated below, and the point it was used to illustrate.

(6) 8-track tapes had a short lifespan.

We noted that (6) could be interpreted in two ways—as being about instances of the kind or about the kind itself. In considering the question of what sort of representation stereotypes involve, we are asking whether having stereotypical thoughts about a kind involves individual kind members (like individual 8-track tapes) or the kind itself (like the kind 8-track tapes).

Our answer to this question is that a stereotype involves beliefs about individual human beings within a kind. We are not claiming that stereotypes

are simply beliefs about the statistical prevalence of certain traits among a category members (existing data strongly suggest that this simple account is incorrect; [Hammond & Cimpian, 2017](#)). Nevertheless, we take having a stereotype of women or Jewish people not to involve representing the kind itself as having some feature, but rather to thinking that the individual members have certain features.

To see the force of this point, consider an especially simple example. Suppose you encounter a new group of people—the Daxes—and find that every Dax you meet is physically weak. So Sarah is weak, and so is Harry, and Samantha is too, and so forth. You might then arrive at a representation of Daxes as weak. It is important to distinguish this representation from any further views you might have about the relationship between this feature and the nature of the kind itself. When it comes to that latter question, you could have any number of different views. You might think that the weakness of the Daxes was due to some deep fact about the nature of the kind itself. Alternatively, you might think that the explanation was environmental or structural ([Taylor, 1996](#); [Vasilyeva, Gopnik, & Lombrozo, 2018](#)). For example, you might think that Daxes in this society happen to be oppressed, and that they are weak because they do not get enough food. Or you might never think about these issues at all. You might never develop any opinion about why Daxes are weak, or about the nature of this kind, and just have a representation of the generalization itself. Clearly, there are important differences between these different views—and we will be discussing those differences below—but there is also a certain sort of representation that you could hold regardless of which of these other views you adopted. That representation is a stereotype (namely, the stereotype that Daxes are weak). The aim of this section is just to understand that representation.

Our account of that representation is relatively straightforward. A stereotype is not a representation of the kind itself. As we saw in our simple example, a given stereotype could be compatible with a number of very different views about the nature of the kind, or with having no view at all about the nature of the kind. The stereotype is simply a representation of the individual human beings who fall within a kind.

3.2 Essentialism

Essentialism, we claim, involves a very different type of representation from the one involved in stereotyping. It is a representation not of the individual human beings in a kind but of the kind itself.

While people rarely have explicit debates about the essences of kinds (Leslie, 2013; Medin & Ortony, 1989), it can be instructive to look at some of the cases in which people do have such debates and examine the form they take. Consider the debate about what it is to be a woman. Some people argue that the essence of being a woman is a matter of biology, others argue that it is a matter of distinctive psychological properties, and still others say that there is no essence of being a woman. Or consider the long-standing debate about what it is to be a Jew. One view is that the Jewish people should be understood as a racial group (as Hitler thought), another is that the essence of being a Jew is having a certain kind of deeper spiritual state (as St. Paul suggests in Romans 2:29), and yet another is that there is no essence of being a Jew.

The thing to notice now is that all of these views are about a fundamentally different topic from the one we find in stereotypes. If you think that Jewish people are neurotic, you are thinking that individual Jewish people have a particular property. By contrast, if you think that Jewish people are a racial group, you are not attributing a property to the individual Jewish people. Rather, what you are doing is attributing a property to a kind. Just as you can represent an individual person as having the property of being neurotic, you can represent a kind as having the property of being a racial group. If one thinks that what it is to be a woman depends on biology, one is thinking that the kind has a biological essence that determines who is a woman.

Clearly there is a close connection between taking a kind to have an essence and taking kind members to have some features. We do not deny that. Our claim is that when a kind is essentialized, it is the kind itself, rather than the individual members who are taken to have an essence.

If our hypothesis is on the right track, there is a very tight relationship between seeing a category as constituting a kind and essentializing that category. What we are suggesting is that essentialism is itself a way of representing a kind. Thus, if people see a category as constituting a kind, they can represent that kind either as having an essence or as having no essence. However, if people do not see the category as constituting a kind, they cannot represent it in either of these ways. Just as you cannot represent an individual person either as being a racial group or as being a nonracial group, you cannot represent a collection of individual people either as having an essence or as having no essence. The sorts of representations to which essentialism can be applied are entity representations.

Putting this point in a slightly different way, the claim is that attributing an essence to a kind is radically different from just believing that a collection

of individual human beings share some deeper property. For example, consider the anti-Semite who believes that there is a deeper biological essence to the kind *Jewish people*. This is a prototypical case of essentialism, and a theory of essentialism should help us to understand its causes and consequences. Now, by contrast, consider the medical doctor who looks through her records and notices that some collection of her patients—say, Julian, Mario, and Fiona—all share a certain deeper biological property. On the view we are proposing here, this is a very different type of mental state. It does not involve attributing an essence to a kind, or indeed attributing an essence to anything at all, and we should not expect it to have the causes or consequences identified by a theory of essentialism.

The account proposed here is not the sort of claim that easily lends itself to direct experimental testing. Nonetheless, we will argue that additional considerations both allow us to further clarify the view and show that there is empirical evidence in support of it. In particular, the account allows us to make sense of three different phenomena observed in the existing empirical literature that would otherwise have been quite difficult to explain.

First, as we saw above, there seems to be a puzzle about the different types of things that can be essentialized. On one hand, people essentialize whole categories of entities (e.g., thinking that there is an essence of women); on the other, people essentialize individual entities (e.g., thinking that there is an essence of Martin Luther King, Jr.). Part of what is puzzling here is the nature of the representations involved. Specifically, it seems that people have representations of whole collections of entities and also representations of individual entities, and it initially seems reasonable to suppose that these are representations of very different sorts. Reflection supports that they are different. Semantic data and semantic theories support this as well. If these types are very different it seems strange or at least unexplained as to why people apply essentialism to representations of both sorts.

The present account dissolves this puzzle. On our account, people do indeed have these two types of representations (representations of collections of individuals, representations of individual entities), but essentialism is never actually applied to representations of collections. Instead, essentialism is *always* applied to representations of individual entities. Some of these are representations of individual people; others are representations of kinds. Yet, despite this difference, our claim is that they are representations of the same basic type. In people's cognition, the representation of a kind is deeply unlike the representation of a collection and is instead best understood as being of the same representational format as representations for individual entities.

Second, consider again the way people represent, for example, the category of Christians. In most ordinary circumstances, people might determine membership in this category by looking to certain relatively superficial features (going to church, saying grace), but all the same, people might think that, ultimately, being a Christian is really about something much deeper (e.g., showing true compassion). Judgments with this same basic structure have been observed for a variety of other categories, including what it is to be a scientist, what it is to be an artist, and what it is to be a man (Foster-Hanson & Rhodes, 2019; Guo, Dweck, & Markman, 2018; Knobe, Prasada, & Newman, 2013; for a review, see Reuter, 2019). As we noted above, some recent research suggests that this pattern of judgment is best understood as a form of essentialism (Newman & Knobe, 2019).

If this claim is correct, it provides additional reason to conclude that essentialism involves an attitude toward the kind itself rather than an attitude toward each of the individual members. To see this, note the difference between the stereotype that Jewish people are neurotic and the essentialist view that being a Christian is fundamentally about having true compassion. In the former case, we are simply looking at a set of individual human beings (Jewish people) and saying that they generally have a particular property (being neurotic). In the latter case, by contrast, we seem to be doing something radically different. It is not that we are looking at the set of all Christian people and saying that they generally have the property of being compassionate. Rather, we seem to be looking at something else—the nature of being Christian—and trying to characterize that other thing. Thus, the claim that compassion is the essence of being Christian does not seem to be a claim about the collection of individual Christian people; it is better understood as a claim about the kind *Christians* itself.

Third, to further draw out the central feature of our view, consider two versions of essentialism (Kripke, 1980; Witt, 1995). *Kind Essentialism* is the view that a kind has an underlying essence that determines kind membership across various times and situations. *Individual Essentialism* is the view that an individual has an underlying essence that determines the individual's identity across times and situations.

These two forms of essentialism are independent. To draw this out suppose that people categorize Maria as a lesbian. If they believe there is an essence of the kind lesbians, then they would believe that there is some hidden property such that no one can truly count as a lesbian unless she has the property. Analogously, if they believe that there is an essence of Maria, they will believe that there is some hidden property such that no one could ever

be this individual person—Maria—unless she had this property. But there is no reason why the essence of the kind has to be part of the essence of each individual person within the kind. People might believe that Maria just happens to be a lesbian but that this is not at all essential to her and that she could just as easily not have been a member of this kind. In other words, if people think that a certain property is the essence of a kind, they might think that every individual person in the kind has this property, but they do not necessarily have to think that this property is essential to every individual person in the kind.

The distinction between kind and individual essentialism and the view that kinds can have essences that rely on values also reveal new possibilities about the way people represent kind membership as immutable or variable. On the framework developed here, when a kind is essentialized, an essence is posited at the level of the kind itself, rather than at the level of collections of individuals. The difference between essentializing a kind and essentializing an individual is important in considering how to test for essentialism. For instance, experiments testing for inflexible or immutable membership might focus on whether an individual is essentially a member of a kind, rather than whether the kind itself has an essence (Gelman, 2003; Haslam et al., 2000; Rothbart & Taylor, 1992; Yzerbyt, Rocher, & Schadron, 1997). In certain cases, these may pattern together. For instance, people might believe that some individual creatures are essentially and immutably tigers and represent the kind *tigers* as having an essence. Yet, in other cases, like the kinds *Christians* and *lesbians*, a kind might be taken to have an essence while allowing for variable membership. The framework explored here makes salient the possibility of representing a kind as having an essence while allowing for variations in kind members.

3.3 Stereotypes and essentialism

The core message of the chapter thus far has been that there is a big difference between thinking about categories of two sorts. Some categories that are seen as constituting kinds (e.g., men), while others are not (e.g., young humorless men). Building on evidence from linguistics, we argued that people have distinct representational formats for kinds and mere categories. When representing mere categories people simply represent a collection of individuals. In contrast, representations of kinds involve two formats: a representation of the kind itself and a representation of a collection of kind members. We argued that these shed light on the representations at issue in

stereotyping and essentialism. Stereotypes involve representations of individual members, while essentialism involves a representation of a kind. We now want to end with one final question: What is the relationship between stereotyping and essentialism?

One possibility is that essentialism actually increases the total amount of stereotyping about a category. On this hypothesis, people might have stereotypes about both essentialized kinds (e.g., women) and nonessentialized kinds (e.g., bartenders), but they would have *more* stereotypes about the former than about the latter. There has been a substantial amount of research designed to test this hypothesis, but at this point, it is still hard to know whether it is true or false. To begin with, there is substantial evidence that stereotyping is indeed correlated with essentialism. A number of studies with adults find a correlation between essentialism and stereotype endorsement (Bastian & Haslam, 2006, 2007; Levy, Stroessner, & Dweck, 1998; Meyer & Gelman, 2016; Tadmor, Chao, Hong, & Polzer, 2013), and some developmental research indicates that this correlation can be found even in children (Pauker, Ambady, & Apfelbaum, 2010; Pauker, Xu, Williams, & Biddle, 2016).

These correlational studies lead immediately to a question as to whether essentialism actually causes stereotyping. Some research with adults suggests that inducing essentialism does increase stereotype endorsement (Brescoll & LaFrance, 2004), but other studies point in a different direction. Adult studies with novel categories find that inducing essentialism does not need to lead to increased stereotyping (Bailey & Knobe, *in preparation*), and developmental research finds a causal effect in the opposite direction, with presentations of stereotypes leading to increased essentialism (Pillow, Allen, Low, & Vilma, 2019; Pillow, Pearson, & Allen, 2015). Moreover, some developmental research does not even find a correlation between children's essentialist views and their endorsing stereotypes (Meyer & Gelman, 2016). In short, the state of existing research is complicated (for further discussion, see Mandalaywala, 2020; Pauker, Tai, & Ansari, 2020). It is not clear that the framework we have proposed here can make headway on this vexed issue.

However, the framework does have the potential to illuminate another important relationship between stereotyping and essentialism. Specifically, a second way in which essentialism and stereotyping might be related has to do with explanation. On this view, it would not necessarily be the case that people have *more* stereotypes about essentialized kinds than about non-essentialized kinds, but it would be the case that people can have stereotypes

about essentialized kinds that are *different* from the ones they have about non-essentialized kinds. Some researchers have argued that stereotypes connect pieces of information via underlying explanations (Yzerbyt et al., 1997, 2001). As we discussed above, essentialism is one theoretical structure that could provide an underlying explanation. But explanations need not rely on essentialism; social and cultural patterns could also do this explanatory work (Ritchie, 2019; Taylor, 1996; Vasilyeva et al., 2018).

Consider the stereotypes that women are bad at math, British people have bad teeth, and rock stars do drugs. Even among people who hold these very stereotypes, there could be radical differences in beliefs about what explains the generalization. Some might believe in explanations that do not in any way involve essentialism. For example:

Women are bad at math because cultural forces discourage them to study math and they are not given adequate educational opportunities and encouragement.

Others might believe in explanations that directly involve a category essence, such as:

Women are bad at math because of the biological nature of women themselves.

The framework we have been developing does shed light on the difference between these two different types of explanations. More specifically, the framework suggests that the difference between these two types of explanations is far deeper than it might initially appear. The difference is not just in the actual content of the explanations; it is a difference in the very format of the representations used.

In either case people might hold the very same stereotype (say, that women are bad at math), and this stereotype would involve the same type of representation (a representation of the individual kind members). Still, there is a fundamental difference between the two types of explanations. The first explains a generalization about individual women in terms of purported facts about those individual women. This woman and that woman and so on have each failed to be adequately supported, encouraged, and given the opportunities needed to excel in mathematics. The second explanation is different. It explains a generalization about kind members in terms of a claim about the essence of the kind. On the view we are developing, this involves a deeply different type of representation—a representation not of the individual members but of the kind itself.

This point leads immediately to a new hypothesis about the relationship between stereotypes and essentialism. Regardless of whether essentialism

about a category increases the total amount of stereotyping about that category, essentialism might transform the representations people use in their stereotype explanations. For categories that are not essentialized (e.g., bartenders), people might tend to explain generalizations in terms of facts about the individual category members, whereas for categories that are essentialized (e.g., women), people might be more inclined to explain generalizations in terms of properties of the kind itself.

Future research should continue to explore the differences between representations of kind members and representations of kinds themselves. As we have seen, the difference between these two types of representation plays an important role in people's social cognition, providing insight into stereotyping, essentialism, and the relationship between the two. If we can gradually come to a better general understanding of these types of representations, we are sure to arrive at further insights into these social phenomena and the cognitive processes that underlie them.

Acknowledgments

We thank Yarrow Dunham, Eleonore Neufeld, Alexander Noyes, Sandeep Prasada, Marjorie Rhodes, Rachel Sterken, and Daniel Wodak for valuable feedback on earlier versions of this paper.

References

- Allport, G. (1954). *The nature of prejudice*. Reading, MA: Addison-Wesley.
- Asher, N., & Pelletier, F. J. (2013). More truths about generic truth. In A. Mari, C. Beyssade, & F. Del Prete (Eds.), *Genericity* (pp. 313–333). Oxford, UK: Oxford University Press.
- Bailey A.H. and Knobe J., (in preparation). Does essentialism increase prejudice, stereotyping, and perceptions of group boundaries?.
- Bastian, B., & Haslam, N. (2006). Psychological essentialism and stereotype endorsement. *Journal of Experimental Social Psychology, 42*, 228–235.
- Bastian, B., & Haslam, N. (2007). Psychological essentialism and attention allocation: Preferences for stereotype-consistent versus stereotype-inconsistent information. *The Journal of Social Psychology, 147*(5), 531–541.
- Bloom, P. (2004). *Descartes' baby: How the science of child development explains what makes us human*. New York: Basic Books.
- Brescoll, V., & LaFrance, M. (2004). The correlates and consequences of newspaper reports of research on sex differences. *Psychological Science, 15*(8), 515–520.
- Brescoll, V. L., & Uhlmann, E. L. (2008). Can an angry woman get ahead?: Status conferral, gender, and expression of emotion in the workplace. *Psychological Science, 19*(3), 268–275.
- Brescoll, V. L., Uhlmann, E. L., & Newman, G. E. (2013). The effects of system-justifying motives on endorsement of essentialist explanations for gender differences. *Journal of Personality and Social Psychology, 105*(6), 891–908.
- Carlson, G. N. (1977). *Reference to kinds in English*. Doctoral dissertation Amherst: University of Massachusetts.

- Carlson, G. N. (1982). Generic terms and generic sentences. *Journal of Philosophical Logic*, 11(2), 145–181.
- Chierchia, G. (1998). Reference to kinds across language. *Natural Language Semantics*, 6(4), 339–405.
- Christy, A. G., Schlegel, R. J., & Cimpian, A. (2019). Why do people believe in a “true self”? The role of essentialist reasoning about personal identity and the self. *Journal of Personality and Social Psychology*, 117(2), 386.
- Cimpian, A. (2015). The inherence heuristic: Generating everyday explanations. In R. Scott & S. Kosslyn (Eds.), *Emerging trends in the social and behavioral sciences* (pp. 1–15). Hoboken, NJ: Wiley and Sons.
- Cimpian, A., & Markman, E. M. (2011). The generic/non-generic distinction influences how children interpret new information about social others. *Child Development*, 82, 471–492.
- Cohen, A. (1999). *Think generic! The meaning and use of generic sentences*. Stanford, CA: CSLI.
- Cohen, A. (2007). Between kinds and properties: Bare plurals across languages. In T. Friedman & M. Gibson (Eds.), *Proceedings of the 17th conference on semantics and linguistic theory (SALT)* (pp. 53–70). Ithaca, NY: Cornell University.
- Dar-Nimrod, I., & Heine, S. J. (2011). Genetic essentialism: On the deceptive determinism of DNA. *Psychological Bulletin*, 137(5), 800–818.
- De Freitas, J., Tobia, K. P., Newman, G. E., & Knobe, J. (2017). Normative judgments and individual essence. *Cognitive Science*, 41, 382–402.
- Foster-Hanson, E., & Rhodes, M. (2019). Normative social role concepts in early childhood. *Cognitive Science*, 43, e12782.
- Gelman, S. A. (2003). The essential child: Origins of essentialism in everyday thought. In *Oxford series in cognitive development*. Oxford University Press.
- Gelman, S. A. (2004). Psychological essentialism in children. *Trends in Cognitive Sciences*, 8(9), 404–409.
- Gelman, S. A., & Coley, J. D. (1990). The importance of knowing a dodo is a bird: Categories and inference in 2-year-old children. *Developmental Psychology*, 26(5), 796–804.
- Gelman, S. A., & Heyman, G. D. (1999). Carrot-eaters and creature-believers: The effects of lexicalization on children’s inferences about social categories. *Psychological Science*, 10, 489–493.
- Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition*, 23, 183–209.
- Gelman, S. A., & Markman, E. M. (1987). Young children’s inductions from natural kinds: The role of categories and appearances. *Child Development*, 1532–1541.
- Gelman, S. A., & Waxman, S. R. (2007). Looking beyond looks comments on Sloutsky, Kloos, and Fisher (2007). *Psychological Science*, 18(6), 554–555.
- Gelman, S. A., & Wellman, H. M. (1991). Insidess and essences: Early understandings of the non-obvious. *Cognition*, 38(3), 213–244.
- Gil-White, F. J. (2001). Are ethnic groups biological “species” to the human brain? Essentialism in our cognition of some social categories. *Current Anthropology*, 42(4), 515–553.
- Grice, P. (1975). Logic and conversation. In M. Ezcurdia & R. J. Stainton (Eds.), *The semantics-pragmatics boundary in philosophy* (pp. 47–59). Broadview Press.
- Guo, C., Dweck, C. S., & Markman, E. M. (2018). Gender categories as dual-character concepts? In: *Proceedings of the Fortieth Annual Conference of the Cognitive Science Society*. 1754–1759.
- Hammond, M. D., & Cimpian, A. (2017). Investigating the cognitive structure of stereotypes: Generic beliefs about groups predict social judgments better than statistical beliefs. *Journal of Experimental Psychology: General*, 146(5), 607–614.

- Haslam, N., & Levy, S. R. (2006). Essentialist beliefs about homosexuality: Structure and implications for prejudice. *Personality and Social Psychology Bulletin*, 32(4), 471–485.
- Haslam, N., & Loughnan, S. (2014). Dehumanization and infrahumanization. *Annual Review of Psychology*, 65, 399–423. <https://doi.org/10.1146/annurev-psych-010213-115045>.
- Haslam, N., Rothschild, L., & Ernst, D. (2000). Essentialist beliefs about social categories. *British Journal of Social Psychology*, 39(1), 113–127.
- Haward, P., Wagner, L., Carey, S., & Prasada, S. (2018). The development of principled connections and kind representations. *Cognition*, 176, 255–268.
- Hoicka E., Saul J., Prouten E., Whitehead L. and Sterken R., Language signalling high proportions, not just generics, leads to essentializing for novel social kinds, n.d., under review.
- Keil, F. C. (1992). *Concepts, kinds, and cognitive development*. MIT Press.
- Kinzler, K. D., & Dautel, J. (2012). Children’s essentialist reasoning about language and race. *Developmental Science*, 15, 131–138.
- Knobe, J., Prasada, S., & Newman, G. E. (2013). Dual character concepts and the normative dimension of conceptual representation. *Cognition*, 127(2), 242–257.
- Krifka, M. (2003). Bare NPs: Kind-referring, indefinites, both, or neither? In R. B. Young & Y. Zhou (Eds.), *Proceedings of the 13th conference on semantics and linguistic theory (SALT)* (pp. 180–203). Ithaca, NY: Cornell University.
- Krifka, M. (1987). *An outline of genericity, partly in collaboration with C. Gerstner, SNS-Bericht* 87–23. University of Tübingen.
- Krifka, M., Pelletier, F. J., Carlson, G. N., ter Meulen, A., Chierchia, G., & Link, G. (1995). In G. N. Carlson & F. J. Pelletier (Eds.), *Genericity: An introduction* (pp. 1–124). The generic book: University of Chicago Press.
- Kripke, S. (1980). *Naming and necessity*. Oxford: Basil Blackwell.
- Landman, F. (1989). Groups, I. *Linguistics and Philosophy*, 12(5), 559–605.
- Leslie, S. J. (2013). Essence and natural kinds: When science meets preschooler intuition. *Oxford Studies in Epistemology*, 4, 108–166.
- Leslie, S. J., Lerner, A., & Zalta, E. N. (2016). *Generic generalizations*. In *The Stanford encyclopedia of philosophy*. Winter 2016 edition, Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/win2016/entries/generics/>.
- Levy, S. R., Stroessner, S. J., & Dweck, C. S. (1998). Stereotype formation and endorsement: The role of implicit theories. *Journal of Personality and Social Psychology*, 74(6), 1421–1436.
- Liebman, D. (2011). Simple generics. *Noûs*, 45(3), 409–442.
- Link, G. (1983). The logical analysis of plurals and mass terms: A lattice-theoretic approach. In P. Portner & B. H. Partee (Eds.), *Formal semantics—The essential readings* (pp. 127–147). Blackwell.
- Mahalingam, R. (2007). Essentialism, power, and the representation of social categories: A folk sociology perspective. *Human Development*, 50, 300–319.
- Mandalaywala, T. M. (2020). Does essentialism lead to racial prejudice? It is not so black and white. In M. Rhodes (Ed.), *Vol. 59. Advances in child development and behavior* (pp. 195–245). Elsevier.
- Markman, E. M. (1989). *Categorization and naming in children: Problems of induction*. Cambridge, MA: MIT Press.
- Martin, C. L., & Parker, S. (1995). Folk theories about sex and race differences. *Personality and Social Psychology Bulletin*, 21, 45–57.
- Medin, D. L., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 179–195). New York, NY: Cambridge University Press.
- Meyer, M., & Gelman, S. A. (2016). Gender essentialism in children and parents: Implications for the development of gender stereotyping and gender-typed preferences. *Sex Roles: A Journal of Research*, 75(9–10), 409–421.

- Morton, T. A., Postmes, T., Haslam, S. A., & Hornsey, M. J. (2009). Theorizing gender in the face of social change: Is there anything essential about essentialism? *Journal of Personality and Social Psychology*, *96*(3), 653.
- Neufeld, E. (2019). An essentialist theory of the meaning of slurs. *Philosophers' Imprint*, *19*(35), 1–29.
- Newman, G. E., & Keil, F. C. (2008). Where is the essence? Developmental shifts in children's beliefs about internal features. *Child Development*, *79*(5), 1344–1356.
- Newman, G. E., & Knobe, J. (2019). The essence of essentialism. *Mind & Language*, *34*, 585–605.
- Nickel, B. (2016). *Between logic and the world: An integrated theory of generics*. UK: Oxford University Press.
- Nickel, B. (2017). Generics. In B. Hale, C. Wright, & A. Miller (Eds.), *A companion to the philosophy of language* (2nd ed., pp. 437–462). Malden, MA: Wiley Blackwell.
- Noyes A. and Dunham Y., Separating kindhood from naturalness: Kinds are diverse in causal structure, n.d., under review
- Noyes, A., & Keil, F. (2019). Generics designate kinds but not always essences. *Proceedings of the National Academy of Sciences*, *116*(41), 20354–20359.
- Pauker, K., Ambady, N., & Apfelbaum, E. P. (2010). Race salience and essentialist thinking in racial stereotype development. *Child Development*, *81*(6), 1799–1813.
- Pauker, K., Tai, C., & Ansari, S. (2020). Contextualizing the development of social essentialism. In M. Rhodes (Ed.), *Vol. 59. Advances in child development and behavior* (pp. 65–94). Elsevier.
- Pauker, K., Xu, Y., Williams, A., & Biddle, A. M. (2016). Race essentialism and social contextual differences in children's racial stereotyping. *Child Development*, *87*(5), 1409–1422.
- Pillow, B. H., Allen, C., Low, N., & Vilma, T. (2019). Young children's use of gender for inductive generalizations about biological and behavioral characteristics: The influence of gender categories and gender stereotypes. *Journal of Educational and Developmental Psychology*, *9*(2), 37–44.
- Pillow, B. H., Pearson, R., & Allen, C. (2015). Young children's inductive generalizations about social categories: When is gender essential? *Merrill-Palmer Quarterly*, *61*(4), 441–467.
- Prasada, S. (2017). The scope of formal explanation. *Psychonomic Bulletin & Review*, *24*, 1478–1487.
- Prasada, S., & Dillingham, E. (2006). Principled and statistical connections in common sense conception. *Cognition*, *99*(1), 73–112.
- Prasada, S., & Dillingham, E. (2009). Representation of principled connections: A window onto the formal aspect of common sense conception. *Cognitive Science*, *33*, 401–448.
- Prasada, S., Hennefield, L., & Otap, D. (2012). Conceptual and linguistic representations of kinds and classes. *Cognitive Science*, *36*(7), 1224–1250.
- Prentice, D. A., & Miller, D. T. (2006). Essentializing differences between women and men. *Psychological Science*, *17*(2), 129–135.
- Reuter, K. (2019). Dual character concepts. *Philosophy Compass*, *14*(1), e12557.
- Reynaert, C. C., & Gelman, S. A. (2007). The influence of language form and conventional wording on judgments of illness. *Journal of Psycholinguistic Research*, *36*(4), 273–295.
- Rhodes, M., & Gelman, S. A. (2009). A developmental examination of the conceptual structure of animal, artifact, and human social categories across two cultural contexts. *Cognitive Psychology*, *59*(3), 244–274.
- 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.
- Ritchie, K. (2019). Should we use racial and gender generics? *Thought: A Journal of Philosophy*, *8*(1), 33–41.
- Ritchie, K. Essentializing inferences, n.d.

- Roberts, C. (2004). Pronouns as definites. In M. Reimer & A. Bezuidenhout (Eds.), *Descriptions and beyond* (pp. 503–543). Oxford: Oxford University Press.
- Roberts, S. O., & Gelman, S. A. (2016). Can White children grow up to be black? Children's reasoning about the stability of emotion and race. *Developmental Psychology, 52*(6), 887–893.
- Rothbart, M., & Taylor, M. (1992). Category labels and social reality: Do we view social categories as natural kinds? In G. R. Semin & K. Fiedler (Eds.), *Language, interaction, and social cognition* (pp. 11–36). Newbury Park, CA: Sage.
- Schwarzschild, R. (1996). *Pluralities*. Dordrecht: Kluwer Academic Publishers.
- Shipley, E. F. (1993). Categories, hierarchies, and induction. In D. Medin (Ed.), *Vol. 30. The psychology of learning and motivation* (pp. 265–301). New York: Academic Press.
- Tadmor, C., Chao, M., Hong, Y. Y., & Polzer, J. (2013). Not just for stereotyping anymore: Racial essentialism reduces domain-general creativity. *Psychological Science, 24*(1), 99–105.
- Taylor, M. G. (1996). The development of children's beliefs about social and biological aspects of gender differences. *Child Development, 67*, 1555–1571.
- Teichman, M. (2016). The sophisticated kind theory. *Inquiry: An Interdisciplinary Journal of Philosophy, 1*–47.
- Vasilyeva, N., Gopnik, A., & Lombrozo, T. (2018). The development of structural thinking about social categories. *Developmental Psychology, 54*, 1735.
- Walton, G. M., & Banaji, M. R. (2004). Being what you say: The effect of essentialist linguistic labels on preferences. *Social Cognition, 22*, 193–213.
- Waxman, S. R. (2010). Names will never hurt me? Naming and the development of racial and gender categories in preschool-aged children. *European Journal of Social Psychology, 40*, 593–610.
- Waxman, S. R., & Markow, D. B. (1995). Words as invitations to form categories: Evidence from 12- to 13-month-old infants. *Cognitive Psychology, 29*(3), 257–302.
- Williams, M. J., & Eberhardt, J. L. (2008). Biological conceptions of race and the motivation to cross racial boundaries. *Journal of Personality and Social Psychology, 94*(6), 1033–1047.
- Waxman, S. R., & Gelman, S. A. (2009). Early word-learning entails reference, not merely associations. *Trends in Cognitive Science, 13*(6), 258–263.
- Witt, C. (1995). Anti-essentialism in feminist theory. *Philosophical Topics, 23*(2), 321–344.
- Yzerbyt, V., Corneille, O., & Estrada, C. (2001). The interplay of subjective essentialism and entitativity in the formation of stereotypes. *Personality and Social Psychology Review, 5*(2), 141–155.
- Yzerbyt, V., Rocher, S., & Schadron, G. (1997). Stereotypes as explanations: A subjective essentialistic view of group perception. In R. Spears, P. J. Oakes, N. Ellemers, & S. A. Haslam (Eds.), *The social psychology of stereotyping and group life* (pp. 20–50). Cambridge: Blackwell.