

The Essence of Essentialism¹

George E. Newman and Joshua Knobe

Yale University

[In press at *Mind & Language*]

Over the past several decades, *psychological essentialism* has been an important topic of study, incorporating research from multiple areas of psychology, philosophy and linguistics. At its most basic level, essentialism is the tendency to represent certain concepts in terms of a deeper, unobservable property that is responsible for category membership. Originally, this concept was used to understand people's reasoning about natural kind concepts, such as TIGER and WATER, but more recently, researchers have identified the emergence of essentialist-like intuitions in a number of other contexts, including people's representation of concepts like SCIENTIST or CHRISTIAN. This paper develops an account that aims to capture how essentialism may operate across these varied cases. In short, we argue that while there is diversity in the forms essentialism can take, these varied cases reflect the same underlying cognitive structure.

1. Introduction

A growing body of evidence suggests that adults and even young children do not represent concepts solely in terms of their observable features, but rather in terms of a deeper 'essence' that is not readily observed (Ahn et al., 2001; Gelman, 2003; Dar-Nimrod & Heine, 2011; Keil, 1989; Medin & Ortony, 1989; Xu, & Rhemtulla, 2005). Existing work has explored the nature of this tendency toward psychological essentialism and used it to explain behavior in a wide variety of domains (e.g., Ahn et al., 2006; Bastian & Haslam, 2006; Bloom, 2010; Gelman, 2003; Haslam, Rothschild & Ernst, 2000; Inagaki & Hatano, 2002; Keller, 2005; Kim & Ahn, 2002).

¹ The authors would like to thank Paul Bloom, Andrei Cimpian, Samia Hesni and two anonymous reviewers for their helpful comments and feedback.

To date, the majority of the research on essentialism has focused on patterns of judgment found for natural kind concepts like TIGER or WATER. For example, even very young children appear to represent tigers in terms of an unobserved characteristic that is shared by all tigers and causes tigers to have their distinct tiger properties (e.g., Keil, 1989). In exploring this kind of thinking, researchers have focused on the ways in which it compares to systematic scientific reasoning. Of course, there are important respects in which people's intuitive essentialism fails to capture the complexities revealed by scientific work on these issues (Hull, 1965; Keil & Richardson, 1999; Mayr, 1982; Newman & Keil, 2008; Shtulman, 2006; Sober, 1994). Nonetheless, people's essentialism does seem to provide an intuitive appreciation of important scientific concepts, such as heredity and unobservable similarities among category members (Cimpian & Markman, 2009; Gelman & Wellman, 1991; Gelman & Markman, 1987; Gelman & Heyman, 1999; Opfer & Siegler, 2004; Gelman, 2003; Inagaki & Hatano, 2002; Kalish & Gelman, 1992; Newman & Keil, 2008).

Curiously, however, essentialist thinking does not seem to be restricted to the quasi-scientific role it plays in reasoning about natural kinds. Indeed, as we discuss in the following sections, people also seem to hold essentialist, or essentialist-like, beliefs in many domains that do not directly involve biological or physical categories. For example, there is a great deal of work demonstrating that essentialism plays an important role in adults' and children's intuitive theories about many social categories, such as race, gender, sexual orientation and religion (Bastian & Haslam, 2006; Diesendruck, 2001; Dunham, Baron & Carey, 2011; Gil-White, 2001; Haslam, Bastian & Bissett, 2004; Haslam et al., 2000; Haslam & Levy, 2006; Hirschfeld, 1995; Hirschfeld 1996; Morton, Postmes, Haslam & Hornsey, 2009; Keller, 2005; Kraus & Keltner, 2013; Rhodes, Leslie & Tworek, 2012; Williams & Eberhardt 2008). People also ascribe essences to individual

entities (e.g., the essence of the United States) and even to individual human beings (e.g., the essence of Jim Smith). Finally, people seem to have essentialist intuitions when reasoning with concepts that invoke certain values or ideals (e.g., the essence of being a scientist, the essence of being a Christian).

The existence of essentialism in these instances raises a number of important and puzzling questions, which have become topics of debate in recent years. For example, is the essence that is responsible for individual identity (e.g., the essence that causes Jim to be Jim) the same essence that is responsible for category membership (e.g., the essence that causes Jim to be a person)? Why do people appear to hold essentialist assumptions for categories that are socially constructed such as SCIENTIST or ART? And, perhaps most importantly, is the essentialism in each of these cases reflective of the same underlying psychological phenomenon or many different ones? Haslam and colleagues highlight this issue when they write: “Despite its growing popularity, the concept essentialism suffers from a lack of definition, owing in large part to the diversity of domains in which it has been put to work” (Haslam, Rothschild & Ernst 2000:113; cf. Gelman, 2003; Gelman & Hirschfeld, 1999; Prentice & Miller, 2007).

This paper attempts to make progress on these issues. Drawing on a variety of insights from existing research, we argue that the patterns of judgment observed in these seemingly unrelated cases are actually best understood as instances of the same psychological phenomenon. The common element is the tendency to try to explain observable features in terms of a further unifying principle. This structure can be applied in cases of scientific concepts (as when people reference the essence of being a tiger), but it can also be applied in cases of value-laden concepts (as when people reference the essence of being a Christian). It is this abstract structure, we suggest,

that constitutes the true “essence of essentialism.” We discuss how this approach integrates existing research in this area as well as its implication for future research.

2. *Forms of essentialism*

To get a better sense for the basic strategy employed in this paper, it might be helpful to consider an analogy. Consider research in linguistics on the way people use the word *can*. It is a striking fact that people use this word to talk about physical laws as well as about moral rules. For example:

(1a). [*Physical law*:] An object *can't* go faster than the speed of light.

(1b). [*Moral rule*:] You just *can't* keep treating her like that!

Though these two uses clearly differ in important respects, it appears that they both instantiate the same abstract structure (Kratzer, 1977). Thus, researchers working on these issues do not try to develop one theory for judgments about physical laws and then another completely separate theory for judgments about moral rules. Instead, the usual approach is to try to develop a unified theory of this abstract structure and then to examine the different ways in which that structure can be instantiated in different contexts (for a review, see Portner, 2009).

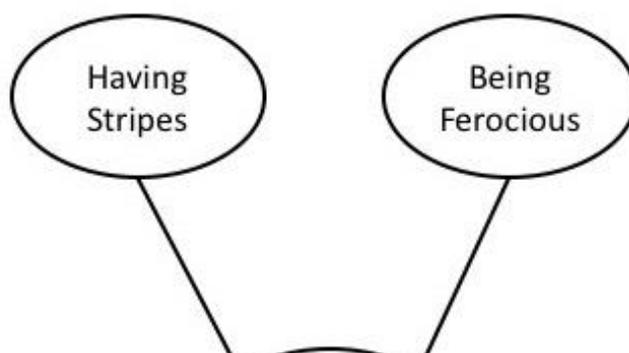
Our aim is to apply a similar approach to the study of essentialism. We begin by looking at people's ordinary understanding of natural kinds. However, our goal is not to develop a theory about natural kind concepts in particular. Instead, our aim is to extract from this research a more abstract structure that might also be used to characterize concepts of other, very different types.

2.1. *Causal essentialism*

Consider the concept TIGER. This concept is associated with certain superficial features (striped, ferocious, furry, etc.), but people do not seem to regard these superficial features themselves as necessary and sufficient for category membership. On the contrary, it is easy to imagine an animal that is truly a tiger but that nonetheless fails to display any of the usual superficial features. For example, consider what would happen if we took a tiger, shaved off all of its fur and then gave it psychotropic drugs that changed its personality. In cases like these, adults and even young children think that the entity would still be a tiger, despite the radical change in observable characteristics (Keil, 1989). For this reason, it is widely acknowledged that people's criteria for category membership are not limited to observable features; they involve some kind of underlying essence that is not as readily observed.

In the particular case of natural kinds, people's understanding of category essence has been studied in impressive detail, and existing work in this area has converged at least to some degree on a widely-accepted account (for a review, see Gelman, 2003). On this account, the essence of a natural kind is understood as the underlying *cause* of its various superficial features. Thus, in the case of tigers, we end up with the structure depicted in Figure 1. The essence of being a tiger is understood as a hidden factor within each tiger—say, its DNA—that causes the features we can actually observe. (Hence, the lines in Figure 1 represent causal relations.) We will refer to this form of essentialism as *causal essentialism*.

Figure 1. Proposed structure of the natural kind concept TIGER, with the essence as a hidden cause of the superficial features.



This account appears to apply not only in the specific case of tigers, but in an enormous variety of others as well. It finds support in studies using numerous different natural kinds, studies that focus on the learning of novel concepts, and even in developmental studies conducted with 3 to 4-year-olds (Carey, 1995; Gelman, 2003). At this point, there is very strong reason to believe that it is at least broadly on the right track.

2.2. *Platonic essentialism*

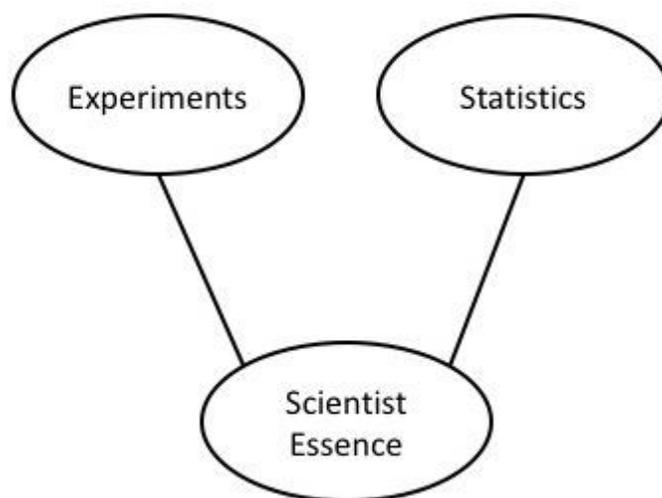
Other categories also seem to be essentialized, but in a very different way. This other form of essentialism is more closely connected with judgments about values and has sometimes been characterized in terms of ‘Platonic ideals’ (Gelman & Rhodes, 2012). We will refer to it as *Platonic essentialism*.

For example, consider the concept SCIENTIST. People associate this concept with certain superficial features (running experiments, analyzing data, writing papers, etc.). However, some people would say that, ultimately, being a scientist is not just a matter of having these features. Instead, they might say that there is some further thing that truly constitutes the essence of what it is to be a scientist—perhaps something like an *impartial quest for truth*. Now imagine a person who has no formal scientific education and therefore has little understanding of how to run

experiments or analyze data, but who nonetheless shows an extraordinary open-mindedness and willingness to change her views in light of empirical evidence. Similar to the tiger example above, when participants are told about such a person, they conclude that even though the person does not possess any of the typical features of a scientist, ultimately, there is a sense in which she is a *true* scientist (Knobe, Prasada & Newman, 2013).

In cases of Platonic essentialism, people appear to believe that what binds together the different features of the category is the fact that they are all ways of embodying the same deeper value. For example, running experiments, analyzing data and writing papers are all ways of realizing an impartial quest for truth. As a result, even if the surface features vary, people will think that individuals are members of the category to the extent that they possess the *essential* value. The core difference from causal essentialism appears to be that instead of understanding the essence as an unobserved cause that gives rise to observable features, in cases of Platonic essentialism, people understand the essence as a value or ideal that the various observable features are ways of realizing (see Figure 2).

Figure 2. Proposed structure of the concept SCIENTIST, with the essence as a value that those superficial features realize.



The fact that Platonic essentialism is not driven entirely by causal judgments is even more apparent when we turn to artifact concepts that are associated with Platonic essences, such as ART, ROCK MUSIC or POETRY (Knobe et al., 2013). For example, consider the concept ROCK MUSIC. People might associate this concept with certain features (electric guitars, repetitive chord progressions, choruses, etc.), but they might also believe that the true essence of ROCK MUSIC lies in some deeper aesthetic value—for example, a certain rebellious nature. Thus, if we want to determine whether a given artifact has this essence, we will not be concerned primarily with questions about the causal factors that generated it; we will be concerned instead with questions about the degree to which its features realize certain values. (Even if a piece of music was randomly generated by a computer, people could listen to the notes and lyrics that were generated and ask whether those characteristics embodied the essence of rock.)

Existing research has aimed to arrive at a better understanding of the specific type of value judgment that plays a role in Platonic essentialism. As a first step, we need to distinguish between the idea of a ‘good’ category member and the idea of a ‘true’ category member. The word ‘good’ can be applied in combination with a variety of concepts, including both those that are seen as having Platonic essences (‘good friend,’ ‘good work of art’) and those that are not (‘good waitress,’ ‘good blog’). By contrast, the use of ‘true’ is more restricted. Participants think that it can be correctly used in combination with concepts seen as having Platonic essences (‘true friend,’ ‘true work of art’) but that it sounds wrong when used with other concepts (‘true waitress,’ ‘true blog’) (Knobe et al., 2013).

The value judgment that plays a role in Platonic essentialism appears to be specifically a judgment of trueness rather than of goodness. To take a specific example, we associate the concept SCIENTIST with certain features (running experiments, developing theories), and it seems that a good scientist would be someone who is proficient with respect to those features (good at running experiments, good at developing theories). However, the Platonic essence of being a scientist is not a matter of being good with respect to all of these features but rather a matter of embodying some deeper value that all of these features serve to realize. Thus, suppose that a person lacked proficiency with respect to each of these features but nonetheless strove in every aspect of her life to pursue an impartial quest for truth. In such a case, we might not consider her to be a "good scientist," but all the same, she would embody the deeper value necessary to be considered a "true scientist," and we would therefore see her as exemplifying the relevant Platonic essence.

1.3 *General essentialism*

Thus far, we have been discussing two phenomena: causal essentialism and Platonic essentialism. We now argue that causal and Platonic essentialism are actually closely related at a more abstract level. Specifically, the two forms of essentialism can be seen as instantiating the same abstract structure, which we will refer to as *general essentialism*.

This abstract structure can be characterized as follows: People associate a concept with certain superficial features, but they do not regard those superficial features as sufficient for category membership; instead, they posit something further that unites them and explains how they are connected. This further thing allows them to answer the question: 'What is it about these specific features that binds them together, so that it makes sense to associate all of them with this

same concept?’ On the view proposed here, it is this more abstract structure that constitutes the essence of essentialism.

Causal and Platonic essentialism can then be seen as two specific forms that this abstract structure can take. In some cases, people think that the features all have the same hidden cause, and in those cases, the essence will be seen as an underlying factor that causes the features (causal essentialism). In others, people think that the features all realize the same abstract value, and in those cases, the essence will be seen as an abstract value (Platonic essentialism).

This framework thereby allows us to pose a new sort of question about the many important findings obtained in the existing essentialism literature. For each of these findings, we can ask the question, is this specific to causal essentialism, or is it revealing something about general essentialism? Presumably, some of the findings will arise only for causal essentialism, but if the present framework is on the right track, we should expect many of them to be more general—these findings should arise not only for causal essentialism but also for Platonic essentialism.

The key empirical test of the present framework is the degree to which we do indeed see striking similarities between causal and Platonic essentialism. If none of the important findings about causal essentialism also arise for Platonic essentialism, the obvious conclusion would be that that there is little value in the more abstract notion of general essentialism. By contrast, if research uncovers numerous respects in which findings about causal essentialism also arise for Platonic essentialism, the conclusion will be that the abstract notion of general essentialism can allow us to detect and understand regularities that would otherwise have been obscured.

In the remainder of the paper, we review existing work on a broad array of different empirical phenomena in which essentialism has been implicated. As we will see, existing results indicate that the two forms of essentialism are strikingly similar when it comes to category

membership intuitions (§2.1), generic language (§2.2), individual concepts (§3.1), and social categories (§3.2). There is also reason to suspect that these two forms of essentialism are triggered by the same causes (§4.1) and that there are cases in which people essentialize a category without having any specific view about which form of essence it has (§4.2). We do not mean to suggest that these results constitute an airtight case for the present framework, but they do provide strong support for the claim that causal and Platonic essentialism are best understood as two manifestations of the same basic phenomenon.

2. *Essentialism in everyday language and thought*

We turn first to evidence from people's everyday reasoning about categories. Within existing research, psychological essentialism has been implicated both in the way people describe category membership (sharp category boundaries) and in the way people express generalizations about categories (generics). As we will see, both of these phenomena reveal striking similarities between causal and Platonic essentialism.

2.1 *Categorization*

One important function of essentialism (if not *the* most important function) is that it provides a basis for categorization. A key question now is whether causal and Platonic essentialism lead to the same patterns in people's categorization judgments.

Much of the existing research on this topic has focused on the question as to whether essentialized categories are seen as having *sharp category boundaries*. More specifically, a variety of studies have suggested that when people see a category as having a causal essence, they tend to see that category as having sharp boundaries (Gelman, 2003). Thus, a person might think:

“Different animals might look like tigers to different degrees, but ultimately, any given animal is either a tiger or it isn’t, and that’s all there is to it.” It seems highly unlikely, however, that participants would have this reaction when it comes to categories that are seen as having Platonic essences. Clearly, there is no sharp line that divides scientists from non-scientists, or rock music from music of other genres. Even if people see a category as having a Platonic essence, they can still see its boundaries as being fuzzy or messy. For this reason, it might at first seem that the pattern of categorization judgments associated with Platonic essentialism is radically different from the one associated with causal essentialism.

Importantly, however, existing studies actually point to a more complex result that makes causal essentialism seem strikingly similar to Platonic essentialism. To begin with, we need to distinguish between a number of different ways in which a category can fail to have sharp boundaries. One way is to have *graded membership*. In cases of graded membership, the concept is associated with criteria that are not purely dichotomous. Thus, some objects are seen as clearly falling within the category, others as clearly falling outside the category, but some are seen as having a more intermediate status. For an especially straightforward example, consider the concept TALL. Some people are clearly tall, some are clearly not tall, but there are also some who have an intermediate status such that it is not clear whether they are tall or not tall.

A question now arises as to whether people attribute graded membership in cases of causal essentialism. This question has actually been the topic of some fascinating research, and there is at least some evidence on either side. On one hand, Kalish (1995) provides evidence that people are more inclined to attribute graded membership in cases of causal essentialism when using causally essentialized concepts than when using defined concepts like PRIME NUMBER. On the other, Diesendruck and Gelman (1999), analyzing many more items (293 in total), find that people

are significantly less inclined to attribute graded membership for natural kind concepts than they are for artifact concepts, and Rhodes & Gelman (2009) obtain similar results with children. In sum, examples of causal essentialism may show some graded membership; they may not. Given, however, that existing research has not directly compared causal and Platonic essentialism, it is difficult for us to make conjectures about the relative degree of graded membership judgments across the two cases.

Instead, we focus on a different way in which categories can fail to have sharp boundaries. This other way is to have what we will call *dual character*. In such cases, people actually associate a concept with two different criteria: a set of superficial features, and a hidden essence. Then they think that a given object is a member of the category if it fulfills both criteria, that it is not a member of the category if it fulfills neither, and that it has a more confusing status if it fulfills one but not the other. Note that this confusing status is different from the one normally associated with graded membership. It is not that there is a set of criteria that an object fulfills to an intermediate degree. Rather, it is that there are *two different criteria*. Thus, the very same object may be clearly a member of the category on one set of criteria but, at the same time, clearly not a member on the other.

Existing results suggest that natural kind concepts are seen as having dual character. For example, take the question as to whether a tomato is a fruit or a vegetable. A tomato has many of the superficial features we associate with vegetables, but botanical scientists tell us that it has the hidden properties associated with fruits. So what exactly is it? Here, participants do not simply go with the hidden essence and ignore the superficial features. Instead, Machery & Seppälä (2010) showed that the majority of participants agreed both with (2a) and with (2b).

(2a) In a sense, tomatoes are vegetables.

(2b) In a sense, tomatoes are not vegetables.

This pattern of judgment seems to suggest that people associate the concept VEGETABLE with two different criteria. According to one of these criteria, a tomato is a vegetable; according to the other, it is not. The result is a very distinctive pattern of ambivalence that appears to be uniquely associated with essentialized concepts.

Indeed, this pattern arises even for the very cases that motivated the traditional philosophical account. For example, consider an animal that has all of the observable features associated with tigers but that does not have the underlying biological properties associated with tigers. Is this animal a tiger? The traditional philosophical answer is that the animal is not a tiger (e.g., Kripke, 1972; Putnam, 1975). However, experimental studies on this case show that people's ordinary intuitions are more complex (Tobia, Newman & Knobe, 2017). People actually agree with both (3a) and (3b).

(3a). There's a sense in which the animal is a tiger.

(3b). Ultimately, if you think about what it really means to be a tiger, you'd have to say there's a sense in which the animal is not truly a tiger at all.

The dual character of people's criteria for membership in these categories also comes out when participants are simply given the relevant case and asked the straightforward question as to whether they agree that the animal is a tiger. When participants are asked to respond on a 1-7 scale, their responses are bimodally distributed, with some participants saying that they completely disagree, others saying that they completely agree, and the mean response falling at about the midpoint of the scale (Corcoran, 2017). Similar findings have been obtained for other causally essentialized concepts, such as WATER and GOLD (Corcoran, 2017; Tobia et al., 2017).

This pattern of intuitions seems to display in an especially explicit way what is most distinctive about essentialized concepts: The dual character criteria indicate that essentialized concepts are associated with two distinct representations, namely, a set of superficial features and with a hidden essence.

The key point now is that people's judgments about Platonic categories show precisely the same pattern. For example, in a recent study (Knobe et al., 2013), participants were asked to consider an individual who displays all of the superficial features associated with the concept SCIENTIST but who fails to embody the abstract value. Is this individual actually a scientist? Participants who received this case tended to agree with:

- (4) There is a sense in which she is clearly a scientist, but ultimately, when you think about what it really means to be a scientist, you would have to say that she is not a scientist at all.

Here again, when participants are simply asked whether they agree that the individual is a member of the category, their responses are bimodally distributed, with some saying that they completely disagree and others saying that they completely agree. Similar results arise for other Platonic concepts, including FRIEND, POEM, ROCK MUSIC and LOVE (Knobe et al., 2013).

In short, the messiness we find in cases of Platonic essentialism does not provide evidence that it is importantly different from causal essentialism. Just the opposite: the pattern of intuitions found for Platonic essentialism is strikingly similar to the pattern found for causal essentialism. Of course, difficult questions arise about precisely how to interpret the dual character intuitions observed in these cases, but regardless of how one interprets this pattern, the fact that it arises for both causal and Platonic essentialism suggests that these two kinds of judgments may reflect the same basic phenomenon.

2.2 *Generic language*

It has often been suggested that people's essentialist beliefs influence their use of a class of linguistic expressions called *generics* (Cohen, 2001; Gelman, 2003). To understand this claim, we need to distinguish between two different types of generic expressions. First, there are what are called 'bare plural generics':

(5a). Helium atoms have two protons.

(5b). Psychology professors don't know very much about quantum electrodynamics.

Bare plural generics allow people to make generalizations about members of a category. Thus, (5a) makes a generalization about helium atoms, while (5b) makes a generalization about psychology professors. A great deal of research has gone into understanding generics of this type (e.g., Leslie, 2008).

A second type of generic expression is what are called 'indefinite singular generics.' Strikingly, generics of the second kind cannot be used in certain cases where generics of the first kind would be perfectly acceptable (Greenberg, 2004; Leslie, Khemlani, Prasada, & Glucksberg, 2009; Prasada & Dillingham, 2006). For example,

(6a). A helium atom has two protons.

(6b). A psychology professor doesn't know very much about quantum electrodynamics.

Here we see that both (5a) and (6a) seems to be saying something true about helium atoms. By contrast, even though (5b) expressed a plausible generalization about psychology professors, (6b) seems to be saying something further that is not plausible. What then is the difference between these two kinds of generics?

One common view in the linguistics literature is that the answer has something to do with the way people think about *essences*. Within this broad approach, linguists have developed a number of more specific ways of working out the details, and research on this topic is ongoing (Cohen, 2001; Krifka, 2013; Lawler, 1973). Still, at a very broad level, the idea is that bare plural generics communicate generalizations about category members, whereas indefinite singular generics communicate claims about category essence. A sentence like (5b) says that it is a good generalization about psychology professors that they don't know much about quantum electrodynamics, while a sentence like (6b) says something different: that it is in some way follows from the very essence of being a psychology professor that they don't know much about quantum electrodynamics.

A question now arises as to whether causal and Platonic notions of essence show similar patterns with respect to generic language. Existing work in linguistics suggests that, in fact, they do (Cohen, 2001). For example, consider the sentences:

(7) A Christian helps those in need.

This sentence does not seem to be communicating something about a hidden causal factor. Instead, it seems to be communicating some kind of value judgment. Thus, if a priest sees that his parishioners are showing contempt for the poor, he could upbraid them by saying: 'What are you doing? Don't you know that a Christian helps those in need?' In short, the very same type of linguistic construction that can be used to communicate claims about hidden causal factors (in sentence (6a)) can also be used to communicate claims about deeper values (in sentence (7)).

If we assumed that causal essentialism and Platonic essentialism were represented in very different ways, we would face the difficult task of explaining why these two different representations both license the use of the same sort of linguistic construction. On the present

framework, however, the puzzle simply disappears. The key claim is that despite their apparent differences, (6a) and (7) express the same kind of claim, namely, a claim about essences. Sentence (6a) expresses a claim about a causal essence, while sentence (7) expresses a claim about a Platonic essence. Yet, on the present account, these two claims are fundamentally similar. It is therefore no surprise that they are expressed in natural language using the same type of construction.

Indeed, there may be cases in which it is not at all clear which particular form of essence a particular indefinite singular generic aims to describe. Consider:

(8) A mother bear protects her cubs.

It seems that (8) is naturally interpreted not as communicating a statistical fact about how bears typically behave, but rather something deeper about the essence of bears. But is it making a claim about causal essence or Platonic essence? On the present account, there need not be any clear answer to this question. People interpret the sentence using their capacity for general essentialism, and that they don't necessarily have to have any definite view about which form of essence it attributes.

3. Individual concepts and social categories

Having discussed some of the important similarities between causal and Platonic essentialism in categorization and generic language, we now discuss how the present framework can be applied to questions arising from ongoing research on individual concepts and social categories. In both cases, we will be concerned with the same type of question discussed in the previous section. Existing research has demonstrated certain effects of causal essentialism. We now ask whether those same effects arise for Platonic essentialism.

3.1 *Individual essences*

Thus far, we have been concerned with the attribution of essences to categories, but people also attribute essences to specific individuals (individual persons, individual nations, etc.). Within the existing literature, one common method for studying attributions of individual essence has been to look at intuitions about 'identity disruption' (Chen, Urminsky & Bartels, 2016; De Freitas et al., in press; Rips et al., 2001; Strohminger & Nichols, 2014). Consider an individual human being – say, Desiree Smith – and suppose that she has a variety of different properties. If she lost one of her non-essential properties (e.g., her specific hairstyle, or her taste for burritos), we would think that she was clearly still Desiree. However, she might also have some properties that we see as so essential to her that if they were lost, we would feel that the resulting person was not even truly Desiree anymore. Therefore, one way of determining which properties people regard as part of the individual's essence is to ask which properties are seen as 'identity disrupting.'

To the extent that people see an individual's identity as most disrupted when she loses causally central features, their judgments may be readily seen as reflecting causal essentialism. For example, suppose that Desiree has a variety of different beliefs, memories, emotions, traits, and so forth. Some of these mental states might be seen as causing many others, and therefore, as more causally central (Ahn, 1998). Just as one might predict, recent studies have found that people are more inclined to say that a human being's identity has been disrupted when her causally central mental states have been lost than when other mental states have been lost (Chen et al., 2016; Molouki & Bartels 2017). Such patterns provide evidence of causal essentialism.

However, existing data also suggest that people's judgments about identity disruption can be guided by beliefs in Platonic essences. That is, these judgments can be influenced by beliefs

about the deeper values that an individual embodies. For example, recent studies have found that people are the most likely to say that a human being's identity has been disrupted when she undergoes changes in her moral values compared to other changes in her mental states (Strohming & Nichols, 2014; Tobia, 2015). Such effects are observed in hypothetical cases, as well as among family members of individuals who are struggling with actual mentally-degenerative diseases (Strohming & Nichols, 2014). These results suggest that attributions of Platonic essence may have an impact on judgments of identity disruption.

One could argue that such patterns are the result of causal essentialism, and that people just regard moral traits as especially causally central. However, recent studies suggest that people's judgments do not conform to a pattern that is neatly predicted by this view. For example, suppose people believe that an individual's beliefs about controversial moral issues are especially causally central. They should then see the individual's identity as especially disrupted regardless of the direction in which these beliefs change. That is, the individual's identity should be seen as disrupted regardless of whether her beliefs about these issues become more liberal or more conservative.

Recent studies suggest, however, that in such cases, people's judgments show a very different pattern. Specifically, when people are told about an individual who changes her beliefs about a controversial political issue, such as abortion, liberal and conservative participants arrive at very different judgments about whether the change reflects the emergence of the person's "true self" (Newman, Bloom & Knobe, 2013; see also De Freitas, Sarkissian, Newman, Grossmann, De Brigard, Luco & Knobe, 2017). Specifically, if the target changes in a way that is morally aligned with conservatives, conservative participants are more likely than liberals to say the change reflects the emergence of the person's true self. Conversely, if the target changes in a way that is morally

aligned with liberals, liberal participants are more likely than conservatives to say the change reflects the emergence of the person's true self.

A similar effect arises when using identity disruption to examine attributions of individual essence. People make different judgments about whether an individual's identity is disrupted depending on whether this individual loses morally bad traits or morally good traits. Specifically, people do not see a human being's identity as disrupted when she loses morally bad traits; they only see disruption when she loses morally good traits (Tobia, 2015). Similar results obtain for the identities of non-human entities. For example, in one study, participants were told about a scouting club that included Christian prayer. Judgments of identity disruption indicated that conservatives regarded the Christian prayer as essential to the club, while liberals did not (De Freitas et al., 2017).

In other words, the data suggest that when reasoning about identity, people seem to take into account both causal and Platonic notions of essence. Judgments of causal centrality appear to have an impact on attributions of individual essence in exactly the way predicted by traditional accounts, but strikingly, judgments about the degree to which an entity embodies certain values appear to have that same impact. This provides evidence for the view that there is a single underlying phenomenon – general essentialism – that can be triggered both by causal judgments and by value judgments.

3.2 Social categories

In recent years, a number of researchers have become interested in how essentialism influences beliefs about social categories. This line of research can be traced back to the work of Gordon Allport (1954), who suggested that beliefs in category essence often occur in individuals with prejudiced attitudes. As Allport put it, “The ‘soul of the ‘Oriental,’ ‘Negro blood,’ Hitler’s

‘Aryanism,’ ‘the peculiar genius of America,’ ‘the logical Frenchman,’ ‘the passionate Latin’—all represent a belief in essence. A mysterious mana (for good or ill) resides in a group, all of its members partaking thereof’ (Allport 1954:169; see Haslam, Rothschild & Ernst, 2002).

This initial proposal led to a growing body of research within social psychology that has explored the ways in which essentialism influences stereotyping and prejudice. For example, participants exposed to biological explanations for gender differences were more likely to endorse a range of gender stereotypes (Brescoll & LaFrance, 2004); having students read an essay promoting genetic explanations caused them to report stronger in-group biases (Keller, 2005); and, participants who read that there was a biological basis for race were less interested in becoming friends with a student who was a different race from their own (Williams and Eberhardt, 2008). In short, a substantial amount of research in social psychology has demonstrated that the tendency to essentialize social categories is one of the stronger predictors of stereotyping (Dar-Nimrod & Heine, 2014).

Because people often think of a person’s genes as a causally-central determinant of many characteristics, the tendency to essentialize a given social category is often closely tied to a notion of causal essentialism. This is reflected in a number psychological instruments such as the Biological Basis (of Essentialism) Scale and the Belief in Genetic Determinism Scale, which both positively correlate with prejudice and negative racial stereotyping (Keller, 2005). Thus, research in social psychology has established an important conceptual link between essentialism in social categories and notions of biological determinism.

We suggest, however, that the tendency to essentialize social categories may not just be restricted to forms of causal essentialism. Here we have suggested essentialism is best understood as a general phenomenon that can emerge in forms that have nothing to do with biological thinking,

namely, Platonic essentialism. If this claim is true, we should predict that essentializing a category will lead to increased prejudice even when that essentialism does not involve seeing a trait as having a biological basis.

While this remains an open empirical question, existing results provide at least some tentative evidence for the claim. Perhaps the strongest evidence comes from research on anti-gay prejudice. Studies show that prejudice against homosexual males is *negatively* correlated with the belief that homosexuality is biologically based (Haslam et al., 2002). If one assumed that the only way to essentialize a social category was to see that category as biological, one would have to conclude that this is a case in which prejudice is actually associated with ‘anti-essentialism.’ By contrast, if one adopts a broader notion of essentialism, one can conclude that people who hold prejudiced attitudes against homosexuals are indeed likely to essentialize homosexuality, but that they simply do not understand this essence in purely biological terms. These individuals may instead think of the essence of homosexuality as constituted by a deeper moral (or immoral) value. While this interpretation is of course speculative, it presents a fascinating area for future research by suggesting that the key difference between anti-gay prejudice and, for example, racial prejudice is not solely in *whether* these social categories are essentialized, but rather in which *type* of essence is posited.

Relatedly, research by Rangel and Keller (2011) distinguishes between notions of biological determinism (stable category features that result from internal biological causes) and social determinism (stable category features that result from external factors, such as status, social background and upbringing). Interestingly, Rangel and Keller (2011) find that while these constructs seem to be independent, both appear to be related to key correlates of essentialist thinking, and in turn, stereotyping and prejudice. For example, exposing participants to salient

social-deterministic explanations for group differences enhances in-group favoritism. Thus, there is some empirical support for the notion that non-biological forms of essentialism may result in prejudice toward social groups in much the same manner as biological (i.e., causal) forms of essentialism.

Finally, Brescoll, Uhlmann & Newman (2013) examined the type of essentialism triggered by system justifying motives. Research in social psychology has shown that people have a fundamental motive to view their social system as just, fair, and good and that they engage in a number of strategies to rationalize the status quo (Jost & Banaji, 1994; Jost, Banaji, & Nosek, 2004). Often referred to as System Justification Theory, this view has been used to explain endorsement of hierarchical relationships, such as the lower socio-economic status of minority groups (Jost, Pelham, Sheldon, & Sullivan, 2003), and existing work suggests that system justification tends to lead to increased essentialism. The key question, however, is whether this effect is specific to one particular form of essentialism (e.g., biological essentialism) or whether it is more general. Brescoll et al. (2013) decoupled the type of explanation (biological vs. cultural) from whether those mechanisms were portrayed as incidental or essential. When participants experienced a system threat, they were more likely to endorse essentialist explanations for gender differences, regardless of whether those explanations were biological or cultural, suggesting an important conceptual link between system justification and beliefs in social category essences that are more general.

In sum, while there is still uncertainty regarding how Platonic notions of essence might inform people's understanding of social categories, there is at least good reason to suspect that causal and Platonic essentialism have strikingly similar downstream consequences for processes of stereotyping and prejudice.

4. *What gets essentialized and how?*

The framework we have been developing here posits two different distinctions among concepts. First, there is the distinction between causal and Platonic essentialism. Some concepts are seen as having causal essences (e.g., TIGER) while others are seen as having Platonic essences (e.g., CHRISTIAN). Second, and perhaps more importantly, there is the distinction between concepts that are seen as having some kind of essence and those that are not seen as having any sense at all. Thus, the concepts TIGER and CHRISTIAN might be seen as having very different kinds of essences, but they are both seen as having an essence, and in this respect, they are similar to each other and different from concepts that are seen as picking out a heterogeneous collection of entities that have no shared essence (e.g., FARM ANIMAL, CONNECTICUT RESIDENT).

In this section, we further explore both of these distinctions in an effort to address how people determine what form of essence a category has and whether that category has any essence at all.

4.1 *Causal vs. Platonic essentialism*

First, a question arises as to how people determine, among essentialized categories, whether a given category has a causal or Platonic essence. How do people determine, for example, that the concept TIGER should be associated with a causal essence, while the concept POEM should be associated with a Platonic essence?

Within existing work, there has already been research about the cues people use to attribute specifically causal essences. One such cue is self-generated movement (Gelman & Gottfried,

1996). As one example, Newman et al. (2008) found that when 14-month-old infants witness an entity move on its own, they expected the entity to have “causally potent” innards. Another cue is the domain to which the object belongs. Adults and children as young as 5-years-old form intuitions about causal mechanisms across various domains and seem to possess a general sense of the complexity of different causal systems (Danovitch & Keil, 2004; Kominsky, Zamm & Keil, 2017). A variety of studies show that this difference between domains impacts people's willingness to attribute causal essences, with people showing more willingness to attribute causal essences to natural kinds than to artifact kinds (Keil, 1989). This research has been helpful in its exploration of the cues used to determine whether a category has a specifically causal essence, and perhaps in future work, similar research will explore the question as to which cues lead to attribution of a Platonic essence.

The present framework does not directly answer this question, but it does make a contribution to a slightly different line of reasoning. Namely, on the hypothesis of general essentialism, there can be cases in which people simply *don't* carefully distinguish causal and Platonic essences. In cases of this type, people might attribute both causal and Platonic essences to the very same category, or they might have a definite sense that a category has an essence, without having any particular view about what sort of essence it has.

This phenomenon is particularly relevant in the case of social categories. For example, beliefs about gender categories tend to take on a causal form of essentialism—i.e., people believe there is some biological essence that is responsible for gender differences (LaFrance, Paluck, & Brescoll, 2004; Martin & Parker, 1995; Morton, Postmes, Haslam, & Hornsey, 2009). However, people may also hold beliefs about gender that seem more Platonic—like the belief that women should be warm and nurturing or that men should be dominant (Eagly & Karau, 2002; Heilman &

Eagly, 2008; Rudman, Moss-Racusin, Phelan, & Nauts, 2012). In the gender literature, different terminology is often used to refer to these different notions: biological gender (male vs. female) versus gender identity (masculinity vs. femininity) (Deaux & LaFrance, 1998; Deaux & Major, 1987).

Within the present framework, this duality can be captured by saying that people do not have a single, stable view about the kind of essence that gender categories have. Instead, they may be capable of thinking of these categories in two different ways: ascribing a causal essence, which conceives of gender difference as resulting from essentially different causes, while also ascribing a Platonic essence which conceives of gender differences as reflecting essentially different values or ideals. Thus, in some cases, people may apply both forms of essentialism to the very same category.

There are also cases in which people believe there is an essence to a category without having *any* view about what form of essence it might have. For example, after getting to know a new person, you might have a strong sense that her essence is her rebelliousness and irreverence. But what sort of essence is this? One possible answer is that there is actually a hidden causal factor within her mind that leads her to be rebellious and irreverent (causal essence); another would be that it is simply a matter of her various actions and mental states realizing a deeper value that involves rebelliousness and irreverence (Platonic essence). In actual practice, however, people might not have either of these two specific views in mind. They might have a strong sense that a particular human being has an essence without having any definite view about what sort of essence it is.

Of course, in saying all this, we do not mean to deny that people sometimes have quite definite views about the essence of a particular category. Existing research has shown that people

do see certain categories as having a specifically biological basis (see Gelman 2003; Keil, 1989), and nothing that we say here is meant to call those findings into question. Rather, the point is just that the present framework opens up a new possibility. Although people may in some cases have a specific understanding of a category's essence, it is in no way built into the structure of essentialism that people must always have a specific view of this type. On the contrary, the essence of essentialism may come out most clearly in cases where people have no specific view, but only a general sense that the category does have an essence.

4.2 *Essentialized vs. not essentialized*

The present account posits a notion of general essentialism that is simply a matter of seeing a category is having an essence and does not involve any specific commitment regarding the type of essence that category has. This claim leaves us with an important empirical question. How do people determine which categories have essences and which do not? Can we identify certain cues that people use quite generally to determine whether or not a category has an essence, independently of what type of essence it might turn out to have?

One input to essentialism that has received considerable attention in the literature is language and, in particular, noun labels. This view proposes that a very subtle distinction between different linguistic categories can yield different expectations about essence. For example, consider the difference between a noun phrase (e.g., she is a *carrot-eater*) and a verb phrase (e.g., she *eats carrots whenever she can*) (Gelman & Heyman, 1999). Despite the similarity of these phrases, research has shown that adults and even young children are more likely to show characteristics of essentialism when the target is referred to using a noun phrase than a verb phrase.

Similarly, Rhodes, Leslie & Tworek (2012) found that the use of generic language

regarding social categories (e.g., ‘Boys play with trucks’ vs. ‘This boy plays with trucks’) led to an increase in essentialist beliefs about those categories among adults and children. Interestingly, these effects were obtained for completely novel social categories and emerged in naturalistic parent-child interactions. Thus, each individual person may attend to subtle differences in the language used by other people as a way of distinguishing essentialized concepts from concepts that are not essentialized (Waxman, 1999).

The key question now is whether this effect is specific to causal essentialism or whether it is a broader effect that would arise for any form of essentialism. Although existing studies have not tackled this question directly, there is at least some reason to suspect that the effect is a broader one. To begin with, there is evidence to suggest that labels may play an important role in establishing essentialist beliefs about categories that are not necessarily biological in nature. For example, Hall and Moore (1997) found that children were more likely to make appearance/reality distinctions when an entity was referred to using a noun phrase (e.g., “This is a blue”), but not when an adjective was used (e.g., “This is a blue one”). In these cases, nouns seem to promote essentialist-like patterns of thought, but arguably the existence of an underlying *cause* is more ambiguous than in the case of categories like TIGER or INSECT.

It seems likely that this effect would also arise for Platonic essentialism. For example, consider the role of labels in setting up essentialist expectations for categories of music. To pick out a category of music, people usually use a noun phrase (e.g., 'Korean boy band,' 'Chicago blues') or a verb phrase (e.g., 'blends 70s soul with contemporary hip-hop'), but there are certain special cases in which we actually have a noun that picks out one particular type of music. Take the noun 'jazz.' This noun can be used to describe an incredible diversity of musical performances including song-based to free-form, tonal to atonal, individual performers to large ensembles, and instruments

of almost any type. Indeed, it seems incredibly difficult to construct a list of superficial features that encapsulate all of the different kinds of musical performances that can be referred to as ‘jazz.’ And yet, the fact that they are described by a single noun seems to suggest a deeper commonality among all of these diverse styles, such that it does not seem unreasonable to posit something like an essence of jazz music.

In short, there does not seem to be anything about the impact of noun labels that would make it be limited to causal relationships in particular. Thus, there is at least tentative reason to suspect that nouns, in cases where they promote essentialism, would lead to an increase in general essentialism rather than causal essentialism specifically. That said, there are of course lots of instances in which people use nouns and the category is not essentialized, suggesting that the tendency to essentialize categories is far more complex and draws on a wide-array of cues in the environment.

A second critical process seems to involve knowledge deference. For example, suppose people encounter several novel entities that are on the surface to be completely different. If a knowledgeable speaker referred to these entities with a common label, then people may be likely to infer that there was some unobserved essence that they had in common. However, if the speaker lacked relevant expertise, people might be far less likely to infer that they shared some deeper essential property. Indeed, a study very similar to this finds that in cases of ambiguous categorization, people defer to biological experts even if the decision involves an artifactualized natural kind (Proctor & Keil, 2006). In other words, people appear to spontaneously seek out information that may connect category labels to essences, and knowledge deference appears to play a critical role in that process.

Note that both of these processes are best understood in terms of general essentialism.

When people rely on the use of noun labels, or on knowledge deference, they are not relying on a cue that is specific to either causal or Platonic essentialism. They are relying on cues that are relevant, in a perfectly general way, to the question as to whether a category has an essence.

5. *Causes of essentialism: The inherence heuristic*

On the present framework, people show a very general tendency toward essentialist thinking that emerges even in cases that do not involve causal relationships. A question now arises about the cognitive processes that give rise to this tendency.

We propose that the tendency toward general essentialism might be explained by the *inherence heuristic* (Cimpian & Salomon, 2014)—a heuristic in which people tend to assume that observed patterns in the world are due “inherent features of things that instantiate those patterns.” For example, people may assume that ‘girls wear pink’ because there is something inherent about the color pink that makes it more suitable for girls than for boys; or, people may assume that there is something inherent about orange juice that makes it particularly well-suited for breakfast. The inherence heuristic is thought to potentially explain a number of phenomena, including forms of causal essentialism. Specifically, Cimpian & Salomon (2014) suggest that essentialism may arise out of a more general tendency to assume that associations in the world are due to inherent, rather than external factors. While this research has not looked directly at the causes of Platonic essentialism, we suggest that the research to date, as well as the theoretical proposal itself, naturally accommodate both causal and Platonic notions of essence.

According to this proposal, the initial belief that relations in the world are due to inherent, rather than external causes may eventually lead to full-fledged essentialist beliefs, as individuals elaborate on certain associations to posit an inherent, unobservable feature that explain the

regularities observed within a kind (Cimpian & Salomon, 2014). In other words, when people wonder why something is the way it is, they are likely to search for information that is salient and accessible to them. The inherence heuristic suggests that information about inherent features provides a shared basis for generating an in-the-moment explanation. In reasoning about such feature, people may naturally arrive at essentialist beliefs of the sort we have been advocating for here.

Existing studies provide compelling evidence for the claim that the inherence heuristic leads to causal essentialism. For example, the tendency to perceive relationships as due to inherent factors is significantly correlated with causal essentialist beliefs (that social groups are cohesive, have biological basis, are rooted in an underlying reality, etc.) (Salomon & Cimpian, 2014). Moreover, this relationship appears to exist independently from other variations in cognitive style such as ‘black and white thinking’ or participants’ own ideological beliefs (Salomon & Cimpian, 2014). In further studies, Salomon and Cimpian (2014) found that priming people toward inherent (versus anti-inherent) explanations subsequently led to greater endorsement of essentialist beliefs. Thus, there seems to be a growing body of evidence which identifies the inherence heuristic as an important antecedent of essentialism.

The key question for the present account is whether the inherence heuristic leads only to causal essentialism, or to general essentialism more broadly. Although existing empirical findings have focused specifically on causal essentialism (e.g., Salomon & Cimpian, 2014), the theories proposed to explain those findings seem to predict that the effect would emerge in a broader array of cases. Suppose people notice that rock music tends to involve themes of rebellion and transgression. Existing theoretical work suggests that the inherence heuristic should give them an initial inclination to think that this fact reflected something inherent to rock music (Cimpian &

Salomon, 2014). However, in this case, this initial inclination would not tend to be elaborated into a purely causal theory (e.g., in folk biology). Instead, people should tend to elaborate it by drawing a connection between rebellion and the values characteristic of rock music – precisely the sort of approach we have described as Platonic essentialism.

In short, although existing studies do not address the matter directly, there is reason to suspect that the impact of the inherence heuristic is not be limited to causal essentialism. Instead, it should lead to a greater tendency toward *general essentialism*, which could then take different forms in different cases (sometimes causal, sometimes Platonic). If this claim does turn out to be correct, the hypothesis would be that general essentialism is explained, at least in part, by the inherence heuristic.

In this section, we have been considering the hypothesis that the new phenomenon discussed in the present paper (general essentialism) can be explained in terms of a cognitive process explored within previous work (the inherence heuristic). We hasten to add, however, that these two ideas are fundamentally independent; either of them could be correct even if the other turns out to be mistaken. Thus, even if it turns out that we were wrong to think that causal and Platonic essentialism are manifestations of the same underlying phenomenon, it could easily still be the case that people make use of an inherence heuristic. For example: it could be that the inherence heuristic is at the root of causal essentialism but not of Platonic essentialism. Conversely, even if it turns out that we are wrong to think that the inherence heuristic explains at the root of general essentialism, it could still be the case that causal and Platonic essentialism are manifestations of the same underlying phenomenon. We would then have to look to some other cognitive process to explain people's tendency toward general essentialism.

6. Conclusion

Our aim has been to explore the relationship between different forms of essentialism. Although the evidence from existing research is certainly far from conclusive, we did find signs of a number of surprising similarities between causal and Platonic essentialism. Taken together, existing research thereby provides at least relatively strong evidence for the claim that these two phenomena are best understood as different manifestations of the same abstract structure, namely, general essentialism.

6.1 Implications for future research

We have discussed a number of the similarities that arise between causal and Platonic notions of essence in research domains related to identity and social categories. We suggest that the notion of general essentialism has the potential to inform further research in these areas.

In our discussion of individual essences, we noted the ways in which people seem to have conceptions of individuals that invoke notions of both causal and Platonic essentialism. We then suggested that these results provided some initial support for an account based on general essentialism. Future work could explore this issue further.

One helpful approach would be to conduct developmental studies. On the present hypothesis, people's understanding of individuals is based on an abstract notion of general essence that need not be specifically causal or Platonic. If this is indeed the case, it should in principle be possible for children to acquire the abstract notion of general essence *before* acquiring an understanding of any specific form that this essence could take. Existing developmental research provides strong evidence that children have a notion of individual essence (Hall, Waxman, Brédart, & Nicolay, 2003), and even infants appear to be able to associate traits with specific individuals

and then track those individuals over time (Hamlin, Wynn & Bloom 2007). A key question then is whether children acquire this more general understanding of individual essence before acquiring an understanding of either causal essence or Platonic essence. If the developmental trajectory does show this pattern, we would have powerful evidence that attributions of individual essence are based on a notion of general essentialism that can be grasped independently of any of its specific manifestations.

Turning to research on social categories, one helpful approach would be to explore individual differences in essentialist thinking. Existing research has led to the development of a number of different scales that measure individual differences in the degree to which participants apply causal essentialism to social categories (Haslam et al., 2000; No et al., 2008; Rhodes and Gelman, 2009). This research has demonstrated that different items measuring such causal essentialism are appropriately intercorrelated (Haslam et al., 2000; No et al., 2008; Rhodes and Gelman, 2009) and that the overall scale scores are correlated with other types of behaviors and judgments (e.g., Salomon & Cimpian, 2014). The present framework makes it possible to formulate an array of new questions regarding these individual differences. Specifically, questions arise about whether there are also individual differences in the tendency to apply Platonic essentialism and how individual differences in the two types of essentialism relate to each other.

One way to explore these questions would be to develop a single scale with items designed to measure both causal essentialism and Platonic essentialism about social categories. One possible outcome would be that such a scale would yield two entirely distinct factors (causal essentialism, Platonic essentialism). Individual differences in these two types of essentialism might then be uncorrelated and might predict quite different types of behaviors and judgments. A very different outcome, however, would be that both causal and Platonic items would load on a single factor of

"general essentialism." If this latter outcome does obtain, we would have evidence of individual differences in the overall degree to which people essentialize human categories, independently of the specific form this essentialism takes.

These are a few examples; the notion of general essentialism can also be used to generate new experimental research in numerous other areas, including everything from language use (e.g., generics) to applied psychology (e.g., stereotyping and prejudice).

6.2 An alternative history of essentialism

The study of essentialism in cognitive science first began as part of an attempt to explore people's ordinary understanding of natural kinds, especially biological kinds. Initially, there was an enormous amount of research, both empirical and theoretical, on essentialist intuitions in this specific domain. Then, only later, researchers began to consider the idea that essentialism might also be playing a role in other domains, such as in people's ordinary understanding of social categories, artworks, or the true self. As a result, it is easy to find oneself thinking that the effects one finds for biological kinds are in some sense the 'real' essentialism and that the effects observed in these other domains involve some degraded form of essentialism, a kind of 'essentialism lite.'

But now consider how the field might have looked if these discoveries had been made in the opposite order. Suppose that researchers had initially been concerned with how people understand social categories, artworks, and the true self. In exploring these aspects of people's psychology, they might have introduced the notion of essentialism. Then other researchers might have noticed that something strikingly analogous appeared to be arising in people's ordinary understanding of biological kinds such as tigers and raccoons. At that point, researchers might have concluded that the sort of thing that arises in the phenomena they had discovered first is real

essentialism, and that whatever we find for people's understanding of tigers and raccoons is merely 'essentialism lite.'

We have been developing an account that aims to capture what is correct in each of these possible views. On this account, neither of the two sorts of phenomena can be said to be more truly essentialist than the other. The essence of psychological essentialism, we have suggested, lies not in anything that is specific to one or another of these domains but in an abstract structure that can be found equally in both.

References

- Allport, G. (1954). *The nature of prejudice*. Reading, MA: Addison-Wesley.
- Ahn, W. (1998). Why are different features central for natural kinds and artifacts? The role of causal status in determining feature centrality. *Cognition*, *69*, 135-178.
- Ahn, W., Kalish, C., Gelman, S. A., Medin, D. L., Luhmann, C., Atran, S., Coley, J. D., Shafto, P. (2001). Why essences are essential in the psychology of concepts. *Cognition*, *82*, 59-69.
- Ahn, W., Flanagan, E., Marsh, J. K., & Sanislow, C. (2006). Beliefs about essences and the reality of mental disorders. *Psychological Science*, *17*, 759-766.
- Barsalou, L. W. (1985). Ideals, central tendency, and fluency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *11*, 629-649.
- Bastian, B., & Haslam, N. (2006). Psychological essentialism and stereotyping endorsement. *Journal of Experimental Social Psychology*, *42*, 228-235.
- Bear, A., & Knobe, J. (2016). Normality: Part descriptive, part prescriptive. *Cognition*.
- Bem, S. L. (1993). *The lenses of gender: Transforming the debate on sexual inequality*. New Haven, CT: Yale University Press.
- Bleier, R. (1984). *Science and gender: A critique of biology and its theories and its theories on women*. Elmsford, NY: Pergamon Press.
- Bloom, P. (2010). *How pleasure works: The new science of why we like what we like*. New York: Norton.
- Brescoll, V. L. & LaFrance, M. (2004). The correlates and consequences of newspaper reports of research on gender differences. *Psychological Science*, *15*, 515-521.

- Brescoll, V. L., Uhlmann, E. L., & Newman, G. E. (2013). The effects of system-justifying motivations on endorsement of essentialist explanations for gender differences. *Journal of Personality and Social Psychology, 105*, 891-908.
- Carey, S. (1995). On the origins of causal understanding. In D. Sperber, D. Premack, and A.J. Premack (eds.), *Causal Cognition*. Oxford: Clarendon Press, 268-308.
- Chen, S. Y., Urminsky, O., & Bartels, D. M. (2016). Beliefs about the causal structure of the self-concept determine which changes disrupt personal identity. *Psychological Science, 27*(10), 1398-1406.
- Cimpian, A., & Salomon, E. (2014). The inherence heuristic: An intuitive means of making sense of the world, and a potential precursor to psychological essentialism. *Behavioral and Brain Sciences, 37*(5), 461–480.
- Cohen, A. (2001). On the generic use of indefinite singulars. *Journal of Semantics, 18*, 183-209.
- Corcoran, P. (2017). The folk on Twin Earth. Unpublished manuscript. Yale University.
- Dar-Nimrod, I., & Heine, S. J. (2011). Genetic essentialism: On the deceptive determinism of DNA. *Psychological Bulletin, 137*, 800-818.
- Deaux, K., & LaFrance, M. (1998). Gender. In D. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *Handbook of Social Psychology (4th ed)* (pp. 347-387). New York: Random House.
- Deaux, K., & Major, B. (1987). Putting gender into context: An interactive model of gender-related behavior. *Psychological Review, 94*(3), 369-389.
- Del Pinal, G., & Reuter, K. (2016). Dual character concepts in social cognition: Commitments and the normative dimension of conceptual representation. *Cognitive Science*.
- De Freitas, J., Tobia, K., Newman, G. E., & Knobe, J. (in press). Normative judgments and individual essence. *Cognitive Science*.

- De Freitas, J., Sarkissian, H., Newman, G. E., Grossman, I., De Brigard, F., Luco, A., & Knobe, J. (in press). Consistent belief in a good true self in misanthropes and three interdependent cultures. *Cognitive Science*.
- Diesendruck, G., & Gelman, S. A. (1999). Domain differences in absolute judgments of category membership: Evidence for an essentialist account of categorization. *Psychonomic Bulletin and Review*, 6, 338-346.
- Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review*, 109(3), 573-598.
- Fausto-Sterling, A. (1985). *Myths of gender: Biological theories about women and men*. New York: Basic Books.
- Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. Oxford, UK: Oxford University Press.
- Gelman, S. A., & Davidson, N. S. (2013). Conceptual influences on category-based induction. *Cognitive Psychology*, 66(3), 327-353.
- Gelman, S. A., & Gottfried, G. M. (1996). Children's causal explanations of animate and inanimate motion. *Child Development*, 1970-1987.
- 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., & Hirschfeld, L. A. (1999). How biological is essentialism? In S. Atran & D. Medin (Eds.), *Folk biology*. Cambridge, MA: MIT Press.
- 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, 58*, 1532-1541.
- Gelman, S. A., & Rhodes, M. (2012). Two-thousand years of stasis. *Evolution challenges: Integrating research and practice in teaching and learning about evolution*, 1-26.
- Greenberg, Y. (2004). *Manifestations of genericity*. Routledge.
- Hall, D. G., and Moore, C. E. (1997). Red bluebirds and black greenflies: Children's understanding of the semantics of adjectives and count nouns. *Journal of Experimental Child Psychology, 67*, 236-267.
- Haslam, N., Rothschild, L., & Ernst, D. (2000). Essentialist beliefs about social categories. *British Journal of Social Psychology, 39*, 113-127.
- Haslam, N., Rothschild, L., & Ernst, D. (2002). Are essentialist beliefs associated with prejudice? *British Journal of Social Psychology, 41*, 87-100.
- Haslam, N., & Levy, S. (2006). Essentialist beliefs about homosexuality: Structure and implications for prejudice. *Personality and Social Psychology Bulletin, 32*, 471-485.
- Heilman, M. E., & Eagly, A. H. (2008). Gender stereotypes are alive, well, and busy producing workplace discrimination. *Industrial and Organizational Psychology, 1*(4), 393-398.
- Inagaki, K., & Hatano, G. (2002). *Young children's naive thinking about the biological world*. New York: Psychology Press.
- Jost, J. T., & Banaji, M. R. (1994). The role of stereotyping in system-justification and the production of false consciousness. *British Journal of Social Psychology, 33*, 1-27.
- Jost, J.T., Banaji, M.R., & Nosek, B.A. (2004). A decade of system justification theory: accumulated evidence of conscious and unconscious bolstering of the status quo. *Political Psychology, 25*, 881-919.

- Jost, J. T., Pelham, B. W., Sheldon, O., & Sullivan, B. N. (2003). Social inequality and the reduction of ideological dissonance on behalf of the system: Evidence of enhanced system justification among the disadvantaged. *European Journal of Social Psychology*, 33, 13-36.
- Kalish, C. W. (1995). Essentialism and graded membership in animal and artifact categories. *Memory & Cognition*, 23(3), 335-353.
- Keil, F.C. (1989). *Concepts, kinds, and cognitive development*. Cambridge, MA: MIT Press.
- Keller, J. (2005). In genes we trust: The biological component of psychological essentialism and its relationship to mechanisms of motivated social cognition. *Journal of Personality and Social Psychology*, 88, 686-702.
- Kim, N. & Ahn, W. (2002). Clinical psychologists' theory-based representations of mental disorders predict their diagnostic reasoning and memory. *Journal of Experimental Psychology: General*, 131, 451-476.
- Knobe, J., Prasada, S., & Newman, G. (2013). Dual character concepts and the normative dimension of conceptual representation. *Cognition*, 127, 242-257.
- Krifka, M. (2013). Definitional generics. *Generativity*, 372-389.
- Kripke, S. A. (1972). *Naming and necessity*. Springer Netherlands.
- LaFrance, M., Paluck, E. L., & Brescoll, V. L. (2004). Sex changes: A current perspective on the psychology of gender. In A. H. Eagly, A. E. Beall, & R. J. Sternberg (Eds.), *The Psychology of Gender (2nd ed.)* (pp. 328-344). New York, NY: Guilford Press.
- Lawler, J. M. (1973). *Studies in English generics* Doctoral dissertation, University of Michigan.
- Leslie, S. J. (2008). Generics: Cognition and acquisition. *Philosophical Review*, 117, 1-47.

- Leslie, S. J., Khemlani, S., Prasada, S., & Glucksberg, S. (2009). Conceptual and linguistic distinctions between singular and plural generics. *Proceedings of the 31st Annual Cognitive Science Society*. Amsterdam: Cognitive Science Society.
- Lynch, E. B., Coley, J. D., & Medin, D. L. (2000). Tall is typical: Central tendency, ideal dimensions, and graded category structure among tree experts and novices. *Memory & Cognition*, 28, 41-50.
- Machery, E., & Seppälä, S. (2010). Against hybrid theories of concepts. *Anthropology & Philosophy*, 10, 97-125.
- Mamdani, M. (2001). *When Victims Become Killers: Colonialism, Nativism, and the Genocide in Rwanda*. Princeton, NJ: Princeton University Press.
- Martin, C. L., & Parker, S. (1995). Folk theories about sex and race differences. *Personality and Social Psychology Bulletin*, 21, 45-57.
- Mayr, E. (1982). *The growth of biological thought: Diversity, evolution, and inheritance*. Cambridge, MA: Harvard University Press.
- Medin, D. L., and Ortony, A. (1989). Psychological essentialism. In S. Vosniadou and A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 179-195). Cambridge: Cambridge University Press.
- 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, 653-664.
- Newman, G.E., Herrmann, P., Wynn, K., & Keil, F.C. (2008). Biases towards internal features in infants' reasoning about objects. *Cognition*, 107, 420–432.

- Newman, G.E., & Keil, F.C. (2008). 'Where's the essence?': Developmental shifts in children's beliefs about the nature of essential features. *Child Development, 79*, 1344-1356.
- Newman, G. E., Bloom, P., & Knobe, J. (2013). Value judgments and the true self. *Personality and Social Psychology Bulletin, 40*, 203-216.
- Newman, G. E., De Freitas, J., & Knobe, J. (2015). Beliefs about the true self explain asymmetries based on moral judgment. *Cognitive Science, 39*, 96-125.
- No, S., Hong, Y. Y., Liao, H. Y., Lee, K., Wood, D., & Chao, M. M. (2008). Lay theory of race affects and moderates Asian Americans' responses toward American culture. *Journal of Personality and Social Psychology, 95*, 991.
- Prasada, S., & Dillingham, E. M. (2006). Principled and statistical connections in common sense conception. *Cognition, 99*, 73-112.
- Proctor, C. & Keil, F.C. (2006). Differences in deference: Essences and insights. In R. Sun & N. Miyake (Eds), *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, pp. 1974-1979, Mahwah, NJ: Lawrence Erlbaum.
- Putnam, H. (1975). The meaning of 'meaning.' *Minnesota Studies in the Philosophy of Science, 7*:131-193.
- Rangel, U., & Keller, J. (2011). Essentialism goes social: Belief in social determinism as a component of psychological essentialism. *Journal of Personality and Social Psychology, 100*, 1056-1078.
- Rehder, B., & Hastie, R. (2001). Causal knowledge and categories: the effects of causal beliefs on categorization, induction, and similarity. *Journal of Experimental Psychology: General, 130*(3), 323.

- Rhodes, M., Leslie, S.J., & Tworek, C. (2012). Cultural transmission of social essentialism. *Proceedings of the National Academy of Sciences, 109*, 13526-13531.
- Rhodes, M., & Gelman, S.A. (2009). Five-year-olds' beliefs about the discreteness of category boundaries for animals and artifacts. *Psychonomic Bulletin & Review, 16*, 920-924.
- 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*, 244-274.
- Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive psychology, 7*(4), 573-605.
- Rudman, L. A., Moss-Racusin, C. A., Phelan, J. E., & Nauts, S. (2012). Status incongruity and backlash effects: Defending the gender hierarchy motivates prejudice against female leaders. *Journal of Experimental Social Psychology, 48*(1), 165-179.
- Russett, C. E. (1989). *Sexual science*. Cambridge, MA: Harvard University Press.
- Salomon, E., & Cimpian, A. (2014). The inherence heuristic as a source of essentialist thought. *Personality and Social Psychology Bulletin, 40*(10), 1297–1315.
- Sidanius, J., & Pratto, F. (1999). *Social dominance: An intergroup theory of social hierarchy and oppression*. New York, NY: Cambridge University Press.
- Smith, E. E., & Medin, D. L. (1981). *Categories and concepts* (p. 89). Cambridge, MA: Harvard University Press.
- Sober, E. (1994). Evolution, population thinking, and essentialism. In E. Sober (Ed.), *Conceptual issues in evolutionary biology* (2nd ed., pp. 161 – 190). Cambridge, MA: MIT Press.
- Strohinger, N., Knobe, J., & Newman, G. (in press). The true self: A psychological concept distinct from the self. *Perspectives on Psychological Science*.

- Strohminger, N. and Nichols, S. (2014). The essential moral self. *Cognition*, 31, 159–171.
- Tobia, K. P. (2015). Personal identity and the Phineas Gage effect. *Analysis*, 396-405.
- Tobia, K. P., Newman, G.E., & Knobe, J. (2017). Water is and is not H₂O. Unpublished manuscript. Yale University. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2933100
- Waxman, S. R. (1999). The dubbing ceremony revisited: Object naming and categorization in infancy and early childhood. In D. L. Medin & S. Atran (Eds.), *Folkbiology* (pp. 233-284). Cambridge, MA: MIT Press/Bradford Books.
- 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, 1033-1047.
- Xu, F. & Rhemtulla, M. (2005) In defense of psychological essentialism. In B.G. Bara, L. Barsalou, and M. Bucciarelli (eds.), *Proceedings of the 27th Annual Conference of the Cognitive Science Society* (pp. 2377-2380). Mahwah, NJ: Erlbaum.