

# Constructivist Set-Theoretic Analysis: An Alternative to Essentialist Social Science

Philosophy of the Social Sciences  
2023, Vol. 53(4) 327–366  
© The Author(s) 2023  
Article reuse guidelines:  
[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)  
DOI: 10.1177/00483931231178100  
[journals.sagepub.com/home/pos](https://journals.sagepub.com/home/pos)



James Mahoney<sup>1</sup> 

## Abstract

Psychological essentialism is a cognitive bias through which human beings conceive the entities around them as having inner essences and basic natures. Social scientists routinely generate flawed inferences because their methods require the truth of psychological essentialism. This article develops set-theoretic analysis as a scientific-constructivist approach that overcomes the bias of psychological essentialism. With this approach, the “sets” of set-theoretic analysis are mental phenomena that establish boundaries and identify similarities and differences among entities whose natural kind composition is not known. The approach is illustrated through a consideration of research on intelligence, race, and poverty in the United States.

## Keywords

constructivism, essentialism, epistemology, ontology, set theory

---

Received 24 January 2023

<sup>1</sup>Departments of Sociology and Political Science, Northwestern University, Evanston, IL, USA

## Corresponding Author:

James Mahoney, Departments of Sociology and Political Science, Northwestern University, Scott Hall, 601 University Place, Evanston, IL 60208, USA.

Email: [james-mahoney@northwestern.edu](mailto:james-mahoney@northwestern.edu)

Of all the psychological biases of human beings, psychological essentialism may be the most important one. Psychological essentialism is the human disposition to believe that, and behave as if, the members of a category share underlying essences that endow them with a common identity and a predictable nature (Medin and Ortony 1989; Gelman 2003; Newman and Knobe 2019). Our essentialist disposition leads us to perceive categories as if they are coherent entities found in the world independently of our beliefs. For instance, we unconsciously assume that all scientists, all marriages, and all crimes ultimately share underlying properties (i.e., essences) that make them members of the categories *scientist*, *marriage*, and *crime*. We may not know the content of these essences, but we still proceed (i.e., we communicate, reason, and behave) under the assumption that such essences do exist. Our belief in these essences causes us to understand particular instances of social categories (e.g., particular scientists, marriages, or crimes) as objective entities that exist in the world.

A large literature in psychology and cognitive science has accumulated over the last 40 years in support of the proposition that human beings engage in essentialism for a wide range of categories, including both natural and social categories (for literature reviews, see Lakoff 1987; Gelman 2003; Newman and Knobe 2019).<sup>1</sup> As Gelman (2003, 6) writes, “Essentialism is a pervasive, persistent reasoning bias that affects human categorization in profound ways. It is deeply ingrained in our conceptual systems, emerging at a very young age across highly varied cultural contexts” (see also Rhodes and Mandalaywala 2017). Psychological essentialism is a *bias* because the members of a social category do not, in fact, share intrinsic properties that make them members of the category. Instead, our social categories depend on the operations and content of human minds—such as shared understandings and expectations—for their existence and efficacy. The pervasiveness of the essentialism bias conceals the extent to which it is a problem: if we all think and talk about categories from the standpoint of psychological essentialism, it is hard to see that this viewpoint is a distortion that needs to be corrected.

In this article, I focus on psychological essentialism as it applies to social scientists themselves rather than to the individuals and groups that social scientists study. My concern is with the effects of psychological essentialism for the research that social scientists produce. Psychological essentialism leads social researchers to inappropriately report that their social units of analysis (e.g., movements, countries) possess properties (e.g., cohesion, development) that exist

---

<sup>1</sup>For a dissenting view, see Strevens (2000); but see also the joint rebuttal of Strevens by eight scientists in Ahn et al. (2001).

in the world as mind-independent entities.<sup>2</sup> Mainstream quantitative methods (e.g., regression analysis and social experiments) assume and require that the members of social categories exhibit objective similarities that they do not possess. The basic assumptions of modern techniques of causal inference are violated because social categories do not possess internally or externally derived essences that make them homogeneous entities appropriate for causal analysis. The lack of true essences for social categories explains why the meanings of important social science concepts are always contested and why social science has struggled to accumulate stable findings about average causal effects.

I distinguish between psychological essentialism, which is a universal human bias, and essentialist social science, which is a mode of social research that assumes and requires the truth of essentialism. Most mainstream quantitative research, including most of the empirical research in the discipline of economics, is essentialist social science. I argue that this research ultimately fails to generate stable knowledge because it assumes the truth of essentialism. By contrast, constructivist scholars in sociology (and other disciplines) recognize and analyze the ways in which categories are historically constituted. They routinely call attention to the reification of social categories. However, even constructivist social scientists lack a means to report findings about relationships among social categories in a way that overcomes psychological essentialism. Simply by virtue of communicating through the perspective with which human beings perceive the world, constructivist researchers also reify their social categories, treating them as if they have an object-in-the-world status when in fact they do not.

Overcoming psychological essentialism, I argue, involves much more than a recognition that human beings create, maintain, and can potentially transform social categories. A real solution to psychological essentialism requires scholars to find a way to represent and analyze social categories as fundamentally mind-dependent phenomena. A non-essentialist approach must recognize that the mind groups together entities as members of a given social category despite the fact that these entities are heterogeneous in their natural and physical composition. The challenge facing constructivist social science is to formulate an approach that builds in a *mandatory* role for the mind in carving out the constructed joints of social reality. While it may be impossible to communicate in general without essentialism, it is possible for analysts to select certain categories and study them in a non-essentialist mode. My goal is to formulate a methodology that allows social scientists to analyze select categories in a non-essentialist way. In particular, I am interested in creating a methodology that allows scholars to represent and generalize about category

---

<sup>2</sup>Social researchers often use human beings (i.e., individuals) as their unit of analysis. Like other biological species, human beings approximate the features of a natural kind, whose existence does not depend on human beliefs and practices (see Dupré 1981).

relationships in a way that does not partake in the mind's default essentialism. In pursuing this quest, I follow in the footsteps of Max Weber (1906 [1949]), who in turn followed Rickert (1896 [1986]) in rejecting the idea that concepts reflect a coherent underlying reality.

I propose that set-theoretic analysis—when formulated as a constructivist approach—can meet the challenge of analyzing select categories in a way that does not partake in our default essentialism. Constructivist set-theoretic analysis offers the analytic tools needed to overcome psychological essentialism while remaining committed to the pursuit of scientific generalization. Set-theoretic analysis is a methodology in which researchers analyze all categories as sets in which other sets can have full membership, partial membership, or no membership (e.g., Ragin 2008; Schneider and Wagemann 2012; Oana, Schneider, and Thomann 2021). Set-theoretic researchers specifically study *set relations*, not variable or property relations. They ask questions about whether or the degree to which one set (or one combination of sets) coincides with, is a subset of, or is a superset of another set. Set-theoretic researchers make descriptive and causal generalizations about the relationships that exist among social categories that are analyzed as sets.

Scholars have recognized the distinctiveness of treating all categories as sets in which other sets can have membership, including degrees of partial membership. The utility of this distinctive approach for social research has been debated among scholars who advocate for or against qualitative comparative analysis (QCA) (see Thomann and Maggetti 2020 for a literature review). However, the scholars in this debate have not considered how understanding categories as sets located in the mind can ground a mode of constructivist research that does not partake in our default psychological essentialism. Therefore, the debate over set-theoretic analysis has not considered what I see as the main virtue of set-theoretic analysis: its ability to help social scientists overcome psychological essentialism. Unless the problem of psychological essentialism is recognized, it is hard to assess the real value of set-theoretic analysis for the social sciences.

To reconfigure set-theoretic analysis for constructivist research, I conceptualize the “sets” of set-theoretic analysis as *conceptual spaces* in the mind's representational system that human beings use to classify entities in the natural world (cf. Gärdenfors 2000, 2014). These conceptual spaces are linked to the meanings and understandings of categories for the individuals who use and understand the categories. Conceptual spaces are ontologically prior to the entities they categorize. Under this constructivist approach, each social category has a corresponding conceptual space, and an entity in the world is a member of a social category when the mind processes it as having

membership in the conceptual space for that category. An entity is a member of the category *scientist*, *marriage*, or *crime* insofar as the mind classifies that entity as having membership in the conceptual space for *scientist*, *marriage*, or *crime*.<sup>3</sup> Through conceptual spaces, human minds transform mostly incomprehensible natural substances and properties into the coherent social entities of human reality.

This article develops its arguments over four sections. The first section discusses psychological essentialism in everyday life and in the social sciences. A second section reformulates set-theoretic analysis as a constructivist approach that avoids essentialist biases and that captures the way in which the mind imposes boundaries onto the natural world. The third section explores how set-theoretical analysis can be used to combine a constructivist ontology with scientific epistemology. A final section illustrates constructivist set-theoretic analysis by comparing three studies focused on intelligence, race, and poverty in the United States: [Herrnstein and Murray 1994](#), [Fischer et al. 1996](#), and [Ragin and Fiss 2017](#).

## I. Psychological Essentialism

This section discusses psychological essentialism as a universal human bias that affects social science simply because social scientists are human beings and perceive the world through this bias. Psychological essentialism leads social scientists to analyze social categories as if their members share unknown essences that make them members of the category. The erroneous assumption that social categories are natural kinds leads social scientists to use inappropriate methods and to generate findings that are often unstable and invalid.

### *1.1. Psychological Essentialism as a Human Disposition*

*Psychological essentialism* refers to a cognitive disposition in which human beings implicitly believe that the members of a category share hidden properties that endow them with a common identity and predictable tendencies ([Medin 1989](#); [Medin and Ortony 1989](#); [Keil 1989](#); [Rothbart and Taylor 1992](#); [Sayer 1997](#); [Haslam 1998](#); [Haslam, Rothschild, and Ernst 2000](#); [Ahn et al. 2001](#); [Gelman 2003](#); [Newman and Knobe 2019](#); [Rose and Nichols 2019](#)). For instance, children assume that all tiger-entities are members of the category *tiger* because these entities possess common unobserved properties that give them a tiger appearance and a tiger nature. If confronted with a tiger-

---

<sup>3</sup>A conceptual space also exists for the category *entity* itself; see [Leslie et al. \(1998\)](#).

entity, children believe that the entity will behave like other tigers because of its possession of tiger-essences. Adults do the same.

While people believe that category members share a common essence, “essentialism does not entail that people know (consciously or unconsciously) what the essence is” (Gelman and Diesendruck 1999, 88; see also Medin and Ortony 1989; Gelman 2003; Newman and Knobe 2019; Rose and Nichols 2019). For many categories, the essence proves to be no more than a placeholder with unknown content. The placeholder structure invites inquiries to discover hidden essences and creative attempts to summarize the underlying similarities shared by all category members (Gelman and Roberts 2017).<sup>4</sup> Yet we always have trouble identifying the precise essences of social categories when asked to do so explicitly. We experience this difficulty because essences do not really exist for social categories. The lack of real essences explains why nearly all important concepts in the social sciences defy easy and consensual definition.

Psychological essentialism is functional for human beings in part because it helps to stabilize our experienced reality and creates solidity in the world (cf. Berger and Luckmann 1966; Searle 1995; Thomasson 2003; Rhodes and Mandalaywala 2017). Our disposition to see social categories as constituted by essential properties may be the secret to our ability to create and sustain social institutions. Because we are usually not aware that social institutions are dependent on our beliefs for their existence, we tacitly accept those institutions as basic facts about the world—we experience them as objective reality (Berger and Luckmann 1966). Psychological essentialism enables human beings to create complex civilizations in which one constructed institutional arrangement is built on top of another. If we were continuously aware that social reality depends on our collective understandings, we might not be able to sustain the matrix of interdependent social rules that constitute human societies.

Psychological essentialism also underpins our capacity to make useful generalizations about social categories and their relationships (see Gelman 2003, 27–43). This orientation provides a logical basis for inductive inference: all entities of the same kind have similar natures *because* they share essential properties. When we learn that a particular case is a member of a category, we infer that the case will function or behave like other members of the category. When we hear that an individual is a *scientist*, that an event is a *wedding*, or that an activity was a *crime*, we automatically begin to make inductive

---

<sup>4</sup>Some scholars argue that social categories are actually constituted by homeostatic property clusters of attributes in which no single attribute is necessary (Boyd 2010; Kornblith 1993). However, this mode of constituting categories is still built on the erroneous assumption that attributes of social categories exist in the world as mind-independent entities.

inferences based on our understanding of the tendencies of the categories. We cannot help it. Psychological essentialism may also explain our ability to use science to discover the real essences of *natural kinds*—i.e., the potentially unobservable but efficacious entities that exist in nature independently of human beings, such as the chemical elements and the elementary particles (Bloom 2000; Ellis 2001). Because we understand categories in terms of essences, we are drawn to investigate the hidden inner properties of natural entities in the physical world.

Psychologists have found that the *type of inference* that one can successfully draw from a category varies depending on the *type of essence* that the category members are imagined to possess (Gelman 2003, 48-52, 137-39; Haslam, Rothschild, and Ernst 2004; Haslam, Holland, and Karasawa 2013; Rhodes and Gelman 2009). With *innate essentialism*, individuals believe that the members of a certain group of living entities share an inherent, biological, and almost permanent internal structure that gives rise to common surface-level features and predictable behaviors. This belief supports generalizations about the physical features and behavioral tendencies of types of plants and animals. Constructivist researchers often rightly criticize the application of innate essentialism to racial, gender, and sexual orientation categories for human beings. Likewise, constructivist researchers in the field of medical sociology criticize the practice of placing human beings in categories for diseases and pathologies supposedly rooted entirely in physiology (e.g., *schizophrenic*, *depressed*, *HIV positive*) (e.g., Hacking 1995). For some scholars, social constructivism is a project mainly aimed at exposing the fallacies, dangers, and exploitative nature of innate essentialism when applied to categories for classifying human beings in the social world.

With *artifact essentialism*, by contrast, we understand inanimate objects to have enduring essences reflecting the intentions of designers (perhaps hypothetical designers) who endow the objects with certain functions (Keil 1989; Bloom 1998, 2000; Keleman and Carey 2007; Gelman 2013). For example, we implicitly believe that all hammers and all airplanes are created by designers in order to be and function as hammers and airplanes. Researchers in the field of science and technology studies routinely criticize the naturalization of artifacts and show how artifacts are socially constructed entities whose meaning and purposes shift over time (e.g., Hackett et al. 2008). While our artifact essentialism supports useful generalizations about artifacts on the basis of their functions (hammers can pound nails, airplanes

can fly), it does not support many additional inductive inferences. The reason is that artifact essentialism does not help us understand *why* and *how* artifacts function in the ways that they do. For example, artifact essentialism correctly informs us that airplanes can fly, but it does not tell us why and how airplanes can fly.<sup>5</sup>

Finally, but crucially, individuals engage in essentialism with social categories, including the majority of categories used in the social sciences (see Diesendruck et al. 2013; Haslam, Rothschild, and Ernst 2000, 2004; Haslam, Holland, and Karasawa 2013; Rhodes and Gelman 2009; Rhodes and Mandalaywala 2017; Tsukamoto et al. 2018). With *social essentialism*, people believe that the non-biological essences of an entity are historically acquired, perhaps even voluntarily chosen, via social relations and social positioning. Social essences are the non-inevitable and often relational attributes that define social categories as certain kinds of substantive entities (e.g., *movements*, *revolutions*, *scientists*) or as certain kinds of properties (e.g., *development*, *wealth*, *political orientation*). Social essences endow a category with the features of *entitativity*, including coherence, boundedness, unification, and meaningfulness (Campbell 1958). Our social essentialism allows us to treat social categories as meaningful, informative, and deeply seated entities in the world. We even feel comfortable counting the exact number of instances of social categories, such as the precise number of scientists at a laboratory, marriages at a church, or crimes in a city. We can point at an individual who “is” a scientist and confidently state, “Here *is* a scientist.”

Although we often recognize that social essences are malleable, we also believe that once a case has acquired the essences of a social category, the case exists empirically in the world with the category’s identity and its basic disposition. As a result, social categories support innumerable probabilistic generalizations and probabilistic predictions. For example, if we know that an individual is a scientist, we are able to make many educated guesses about other historical and social characteristics of that individual. Likewise, if we learn that a group of people are all married couples, we are able to predict other characteristics that these individuals are likely to share. Our knowledge of generalizations about social categories is vital to our ability to function normally within society. Social essentialism upholds these generalizations by attributing them to the historically acquired social properties that the category members possess.

---

<sup>5</sup>Constructivist researchers have found that scientific categories (e.g., *species*, *light*, *electron*, *gene*, *quark*) are also evolving entities whose meaning changes over time (Kuhn 1962 [1970]; Latour and Woolgar 1979 [1986]; Giere 2006). However, these entities do allow for many important generalizations beyond their designer functions, suggesting that they are partial or approximate natural kinds.



Despite its indispensable utility for everyday life and human civilization, psychological essentialism disguises the constructed nature of social reality by presenting it to us as naturally existing objects and properties (cf. James 1896 [1948]; Berger and Luckmann 1966). One way to make this point is to say that psychological essentialism leads us to unconsciously live within a matrix of self-fulfilling prophecies. In Merton's (1948) classic formulation, a self-fulfilling prophecy occurs when our initially false belief that some social phenomenon  $X$  exists causes  $X$  to exist when it otherwise would not have existed.<sup>6</sup> Merton's examples of self-fulfilling prophecies include the 1930s financial panic and Black people's acting as strikebreakers in the early-industrial United States. However, whereas Merton treats the final existence of social phenomenon  $X$  as an objective matter, the present point is that the existence of this phenomenon also depends on its own self-fulfilling prophecy rooted in psychological essentialism (cf. Pels 2002; Krishna 1971). Social phenomenon  $X$  itself (e.g., *the financial panic* or *the Black strikebreakers*) depends on our belief that the phenomenon exists over time; and we believe that it exists over time because our psychological essentialism leads us to understand it as an objective, coherent, meaningful, and self-sustaining entity in the world. Our psychological essentialism hides from us the fact that this outcome is an ongoing self-fulfilling prophecy.

Psychological essentialism is so entrenched in human reality and so necessary for individuals to function in society that we might be inclined to believe that it must be a valid way of perceiving the world. Yet the commonness and the utility of a psychological orientation do not establish its validity (Dennett 1987).<sup>7</sup> Understanding reality often depends on departing from our commonsense orientations, helpful as they otherwise may be. For example, our intuitive physics is necessary for everyday life, but it is misleading about the true nature of physical reality (McCloskey 1983). Natural scientists use intuitive physics for everyday purposes and in parts of their work, but they must depart from this orientation when studying targeted physical substances, properties, and processes. Similarly, our psychological essentialism is necessary for everyday life, but it hides from us the mind-dependent nature of social reality. Social researchers must engage in psychological essentialism most of the time, including for parts of their research, but they need to depart from this orientation for the analysis of select categories.

---

<sup>6</sup>In Merton's words, "The self-fulfilling prophecy is, in the beginning, a false definition of the situation evoking a new behavior which makes the originally false conception come true" (1948, 195).

<sup>7</sup>Indeed, our psychological orientations may have been shaped by natural selection to hide the truth, in order to promote adaptive behavior (Hoffman, Singh, and Prakash 2015).

## 1.2. Psychological Essentialism in the Social Sciences

In the social sciences, psychological essentialism manifests itself with the *property-possession assumption*: the implicit belief that social categories possess properties that endow them with an identity and dispositional tendencies. As social scientists, we embrace the property-possession assumption when we define and measure our units of analysis (e.g., organizations, cities, nation-states) on the basis of the essential properties that these units possess. We use the assumption when we compare our units of analysis using variables (e.g., size of membership, level of wealth, type of power) that we believe capture variation in a property possessed by the units. The property-possession assumption is at work when we test hypotheses to see if changes in one property (e.g., level of development) cause changes in another property (e.g., level of democracy). The property-possession assumption is so basic to social science as to appear obvious; the observation that social scientists study the properties possessed by their cases seems unproblematic and banal.

Social scientists reify their categories when they analyze them under the property-possession assumption. *Reification* occurs when individuals understand social formations as existing independently of human beliefs and practices when they actually depend on human beliefs and practices for their existence (Berger and Luckmann 1966, 89). Because of psychological essentialism, reification is the default mode of apprehending reality for human beings. The problem applies not only to a few select social categories, such as those related to race, gender, and mental illness. Rather the problem applies to most social science categories. We need a mode of constructivist research that allows researchers to overcome psychological essentialism for ordinary social categories such as scientist, marriage, crime, nurse, religion, war, and so on. If social scientists are going to analyze these everyday social categories, they must find ways to communicate about them without assuming that the categories exist in the world as natural kinds.

Reification has implications both for everyday life and for the products of social science. At an everyday level, reification keeps people from the empowerment and self-liberation associated with the profound realization that *all* social categories—not just categories such as *Black person* and *depressed person*—are mind-dependent constructions. All social categories are human creations that only seem to carve nature at its joints. Some may argue that recognition of the “in principle” transformability of social reality is a trivial consolation prize if we cannot transform society “in practice.” Yet, as Pitkin (1987, 287-88) points out, the alternative to this recognition is much worse: a failure to think through and theorize all of the conditions under which we might address our most urgent personal and collective problems.

For the social sciences, reification leads scholars to make problematic assumptions and generate flawed inferences. Insofar as social science aims to accumulate valid knowledge about descriptive and causal patterns, it will fail unless it finds a way to overcome essentialism for select categories. With causal analysis, the problem is that scholars erroneously assume that regularities among social categories exist because of the essential properties that the categories possess. Yet, if social categories do not possess essential properties, any associations among them cannot be traced back to or explained by such properties (cf. [Bourdieu 1989](#)). Patterns and regularities among social categories always depend on the beliefs that sustain these categories. Reification leads scholars to misconstrue the meaning of regularities, attributing their existence to hidden properties or relational positions that the categories do not actually possess.

Reification also leads scholars to study the social world with inappropriate methods that generate invalid findings. Notably, the statistical methods that social scientists use require that the instances of the categories under study are homogeneous entities. For example, the potential outcomes framework—which is the dominant approach to causal inference in economics, political science, and quantitative sociology—yields valid causal inferences only insofar as the stable unit treatment value assumption (SUTVA) is valid ([Rubin 1974](#); see also [Holland 1986](#); [Morgan and Winship 2007](#) [2015]). This assumption requires that a treatment ( $X = 1$ ) have the same form across all cases. Yet SUTVA does not obtain for social categories: a given change on a social category (e.g., a given change in level of education, crime, or democracy) is a heterogeneous occurrence across entities in the natural world. We regard these different occurrences as homogeneous only by virtue of our essentialist disposition, which tells us that the instances of a social category are homogeneous because they possess the same essences.

The bias of psychological essentialism helps to explain the lack of stable and convincing causal inferences in the social sciences. [Freedman \(1999, 2008\)](#) suggests that studies using statistical models of causality have yielded dozens of sound findings about causal effects. However, these dozens of successful inferences (e.g., inhaling tobacco fumes causes lung cancer) are from studies that focus on partial natural kinds, such as work in the field of epidemiology. It is far more difficult to identify similarly successful research findings that focus on social categories. This fact may underlie the skepticism that some natural scientists hold toward the social sciences. These problems of inference also make it hard to use social science findings as the basis for policy interventions and other efforts to manipulate the social world in desirable ways. If our inferences are not valid, they do not provide a strong foundation for acting on the world to change it.

Within the realist traditions of the social sciences, the *critical realist* school of philosophy goes the furthest in acknowledging the constructed aspects of

human reality (see the literature reviews in [Steinmetz 1998](#); [Gorski 2013](#)). Yet from a constructivist perspective, critical realists still reify social reality, despite their constructivist gestures. They still assume that social categories have effects (like those of natural kinds) because of their possession of real causal powers; in fact, they argue that we must treat social categories as causally efficacious entities because they really *are* causally efficacious entities. They do not acknowledge that the apparent objectivity of social categories is an illusion. Critical realists do not appreciate that social categories are dependent on human minds *at all levels* of explanation and analysis (see the critiques in [Fuhse 2022](#); [Reed 2008, 2011](#)). They do not recognize that regularities can exist between social categories without those categories' possessing any causal powers or efficacious mechanisms. While their acknowledgement that human reality is dependent on human minds is helpful, critical realists stop short of embracing a full-blown constructivist view, leaving themselves in a limbo between essentialism and constructivism.

With respect to alternatives to essentialism, scholars working from a relational perspective may have gone the furthest in proposing a viable option (see [Dewey and Bentley 1949](#); see also the literature discussion in [Emirbayer 1997](#)). Relational scholars raise serious concerns about essentialism through their criticism and rejection of what [Emirbayer \(1997\)](#) calls the substantialist perspective. The *substantialist perspective* holds that "it is substances of various kinds (things, beings, essences) that constitute the fundamental units of all inquiry." With a *relational perspective*, by contrast, "the very terms or units involved in a transaction derive their meaning, significance, and identity from the (changing) functional roles they play within that transaction" ([Emirbayer 1997](#), 282, 287). On this relational view, entities cannot be separated from the contexts within which they are situated. Social categories are literally constituted by their connections to other social entities. For example, the category *White person* or the category *bourgeoisie* has no substantive content independent of the relational characteristics that define it. This approach is highly appealing because it moves us decisively away from the erroneous idea that social categories possess *internally-generated essences* that constitute category identities.

However, relational analysts still assume that entities are defined by their possession of relational properties, and these properties are erroneously treated as mind-independent entities. For example, the relational properties of *bourgeoisie* might include the position of employer within an employer-employee relationship and the position of owner within an owner-user relationship vis-à-vis the means of production. Relational analysts treat these structural positions as mind-independent properties that individuals can possess. From a constructivist perspective, by contrast, the relational categories of *employer* and *owner of means of production* do not exist in the social world as mind-independent entities any more than social categories that are

fully defined by internally-generated properties. The problem with relational sociology is that the relational attributes that constitute social categories are reified (cf. Psillos 2006). Thus, although relational sociologists avoid internal essentialism by defining categories exclusively in terms of externally derived properties, they still lack a means to communicate about these externally derived properties as mind-dependent entities. They only push reification back one step: the relations and relational positions are now the reified entities.

A real solution to psychological essentialism requires scholars to develop an approach that does not ultimately assume mind-independent categories. Yet our essentialism is so basic to how we perceive, reason, and communicate that it is not apparent that we can truly transcend this viewpoint. Overcoming psychological essentialism is not only a matter of accepting that our linguistic categories function quite well despite their referential disconnection with natural kinds. Nor is it only a matter of accepting the human-constructed and mind-dependent nature of our categories, as constructivist researchers routinely do. Rather, to escape this insidious essentialism, we need an approach that allows us to analyze the social world as composed of something other than entities in possession of internally or externally-derived properties that do not depend ontologically on human minds. We need an approach that views human reality in a way that does not correspond to how we ordinarily experience it.

## 2. Foundations of Constructivist Set-Theoretic Analysis

A distinguishing feature of set-theoretic analysis is its treatment of all categories as sets. This focus on sets rather than variables produces major differences between set-theoretic analysis and variable-oriented analysis. Yet this focus does not in itself generate a constructivist approach that overcomes psychological essentialism. Rather, to create a *constructivist* set-theoretic approach, we need to conceive of sets as entities that are embedded in human minds. This section implements this constructivist move and argues that it provides a way to overcome psychological essentialism in the social sciences.

### 2.1. Categories as Conceptual Spaces

If we take seriously the argument that social categories depend on human minds for their existence, we need some model of the brain operations that underpin this mind-dependence.<sup>8</sup> Such a model stands to enrich our

---

<sup>8</sup>Some philosophers (e.g., Clark and Chalmers 1998) argue that the mind can be partially located outside of the brain and indeed the entire human body. For my purposes here, however, I follow the convention of treating minds and brains as coextensive entities, with the former existing virtually inside the latter.

understanding of what it means to say that an entity “is mind dependent” or that our categories “depend on shared understandings.” The model should be consistent with and draw from insights in cognitive and neuroscientific fields concerning how the minds works with respect to categories. At the same time, social science is not brain science, and reductionist explanations of categorization that emphasize neural and other lower-level processes are generally not helpful for most social science purposes.

I propose the use of a *conceptual space model* of human categorization to illuminate the mind-dependent character of social categories (Gärdenfors 2000, 2014; Warglien and Gärdenfors 2013; see also Fauconnier 1985 [1994]; Rosch 2011; Churchland 2012). The conceptual space model is a geometric understanding of how human categorization takes place. It uses the ideas of location, boundaries, points, and distance to summarize what occurs in the brain when entities from the natural world are deemed to be instances (or not instances, or partial instances) of a social category. The model is compatible with theories of the neural foundations of categorization, even though it emphasizes processes that occur at a higher level of aggregation (Balkenius and Gärdenfors 2016). The conceptual space model is built from the analysis of everyday concepts in ordinary human communication. A model derived from ordinary categories is appropriate because social scientists often study ordinary categories and because they function as their own ordinary language community. Social science categories are mind-dependent in the same way that nearly all human categories are mind-dependent. The most important difference is that social scientists have the opportunity to recognize and formally model this mind-dependence in their work. When compared to ordinary language users, social researchers can more easily and explicitly discuss how their minds are involved in the co-constitution of category meanings.

The conceptual space model asserts that the mind encompasses a representational hyperspace, and that categories correspond to *conceptual spaces* within this hyperspace (Gärdenfors 2000, 2014).<sup>9</sup> Conceptual spaces are *bounded locations* in the virtual space of the mind; they can be visualized as closed shapes, such as circles or ovals, although the mind’s hyperspace is in fact multidimensional. In the conceptual space model, each category has a corresponding conceptual space in the minds of those individuals who use the category. To know or to understand a category requires having a clearly bounded conceptual space for the category in one’s mind. If one lacks a

---

<sup>9</sup>In Gärdenfors’s (2000, 2014) seminal formulation, categories are intersections of convex regions (i.e., regions in which, for any two points, the region contains the full line that connects those two points) within conceptual space. Here I treat these convex regions as conceptual spaces themselves. Some scholars suggest that categories are better represented as star-shaped regions rather than convex regions.

conceptual space for a category, one cannot situate entities in the world in relationship to that category.

With the conceptual space model, categorization occurs when sensory input from the world activates a conceptual space, leading individuals to classify the input as an instance of the category (Gärdenfors 2000, 2014; see also Barsalou 1999, 2005, 2016).<sup>10</sup> More precisely, categorization occurs when the mind gathers and processes sensory input as an entity with membership within a conceptual space. With this membership, the conceptual space is activated, and the mind perceives the category as present in the world. For instance, the mind perceives an individual to be a scientist, or an event to be a marriage, when the mind situates sensory input as an entity that is a member of the conceptual space corresponding to *scientist* or *marriage*. The commonality shared by all scientists and by all marriages is that these entities trigger conceptual space activation in the minds of individuals with conceptual spaces for *scientist* and *marriage*. This activation causes the individuals to believe that certain entities in the world literally *are* scientists and marriages. If individuals share an understanding of *scientist*, the same sensory input activates their conceptual spaces for *scientist*, producing intersubjective agreement.

In the conceptual space model, *prototype points* exist at the center of conceptual spaces (Gärdenfors 2000, 2014; see also Rosch 1973, 1978, 1999). These points correspond to the best or most exemplary member of a category. The conceptual space model proposes that the mind situates sensory inputs from the natural world as entities in relation to prototype points as well as to their surrounding conceptual spaces. The *extent to which* an entity exemplifies a category depends on its proximity to the prototype point at the center of the conceptual space corresponding to that category. For some categories, the prototype point cannot be activated by sensory input from the actual world; for these categories, the prototypical case is a non-actual case.<sup>11</sup>

The conceptual space model offers a way of representing what it means for two or more individuals to *share an understanding* of a social category. When individuals share an understanding of a social category, a conceptual space for the category exists in each of their minds. These conceptual spaces stand in similar distance relations to other conceptual spaces for related categories and

---

<sup>10</sup>Conceptual space activation is the final phase of longer, complex processes of meaning-making that take place in the brain.

<sup>11</sup>Prototypes are essential for understanding many important aspects of categorization, including compound categories. For example, a goldfish may be the prototype of *pet fish*, and *pet fish* may be the intersection of *pet* and *fish*, but a goldfish is not the prototype of either *pet* or *fish* (cf. Smith and Osherson 1984). The prototypes of *pet* and *fish* (e.g., perhaps *dog* and *bass* in certain regions of the United States) are not located close to one another, and thus they both cannot be close to the prototype of *pet fish*.

background understandings of society. For example, if two or more individuals share an understanding of the category *marriage*, similar distance relations exist between their prototypes of *marriage* and other related categories, such as *two-person relationship*, *legal union*, *romantic relationship*, and *financial relationship*. Moreover, within the minds of these individuals, the *marriage* category is situated within similar background models of society (Lakoff 1987). For example, two individuals who share an understanding of the category *marriage* may place this category within a conceptual space standing for a society in which monogamous and legalized relationships are common. This similar positioning of the category *marriage* allows the individuals to communicate about marriage with a high level of agreement and understanding. Likewise, when presented with anomalous examples of marriage (e.g., Tarzan and Jane), they are likely to understand the anomalies in similar ways.

Our exposure to and engagement with the world carves out specific conceptual spaces in the representational hyperspace of our minds. From the beginning of life, experience causes our brains to create conceptual spaces that then allow us to perceive the stuff of the world as possessing various similarities and differences. Once conceptual spaces are created, we use them unconsciously and almost continuously in our everyday lives as we effortlessly process sensory information. Although our genetic constitution as human beings may predispose us to develop conceptual spaces in the first place, specific conceptual spaces cannot develop unless we have access to certain social environments (Barsalou 2016).<sup>12</sup> The vast majority of the conceptual spaces that exist in our minds depend on formal and informal learning. The specialized categories used in the social sciences are a case in point: social scientists must learn these categories through exposure to scholars and scholarly works. This exposure carves out conceptual spaces for social science categories and situates them in particular spatial relations vis-à-vis other categories.

Two or more entities in the world that trigger the activation of a particular conceptual space (e.g., *scientist*, *marriage*, *crime*) need not share the same natural kind composition. In fact, with social categories, the entities that activate a particular conceptual space are always heterogeneous in their natural kind composition. The data inputs that cause conceptual space activation do not have anything approximating a one-to-one correspondence (i.e., mathematical bijection) with specific natural kind properties in the world. On present knowledge, we can usually only understand the data and sensory correlates of a social category in terms of other social categories. For example, we must define a social category such as *scientist* in terms of its social

---

<sup>12</sup>For a contrasting view in which all categories are innate, see Fodor (2000).



correlates, such as *works in a laboratory*, *uses the scientific method*, and *is employed as a scientist*. These social correlates have no more referential connection with natural kinds than the category *scientist* itself. It is fruitless for social researchers to attempt to generalize about the natural kind properties that are shared by all scientists and that make them scientists as opposed to not-scientists. Instead, social researchers must define, understand, and communicate about *scientist* on the basis of this category's relationship to other social categories.

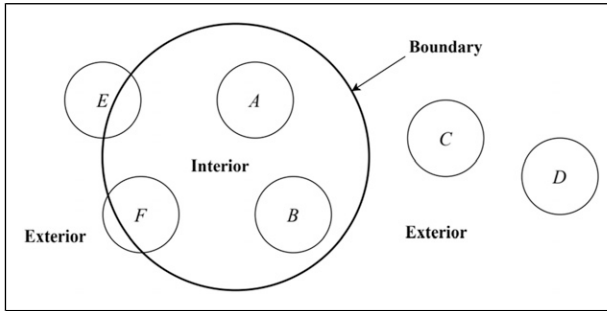
This conceptual space model provides set-theoretic analysis with a way of analyzing categories that does not partake in the mind's default essentialism. With this approach, the categories under analysis are assumed to correspond to conceptual spaces in the mind, rather than tracking natural boundaries in the external world. A given category has a one-to-one correspondence with a particular conceptual space, rather than with a particular set of natural kinds in the world. Heterogeneous natural entities can all be members of the same category if they activate the same conceptual space in the mind's representational system. The one thing that all instances of a given social category have in common is their activation of a conceptual space in the minds of those who share an understanding of the category. This activation can occur even if the natural kind composition of the entities is completely different.

## 2.2. Property Sets versus Spatial Sets

Researchers often understand the idea of a *set* as a way of discussing a group of entities that all share one or more properties (Tversky 1977). For example, they assume that the set of all scientists is the collection of all individuals who possess the properties of a scientist. I label these sets *property sets*, because they call attention to the property that the entities already possess; the shared property *makes* the entities members of the set.

An alternative definition of a set—what I label a *spatial set*—is a bounded location in space in which entities can have membership (Lakoff and Núñez 2000, 30–31, 43–45). As Figure 1 suggests, spatial sets have three parts: an interior, a boundary, and an exterior. The interior of a set includes all entities of a given kind that have membership in the set; the exterior includes all entities of a given kind that do not have membership; and the boundary partitions membership and non-membership. With spatial sets, the set is ontologically prior to its members. The boundaries of the set determine whether entities are members of the set; the properties of the entities do not determine the boundaries of the set. Membership boundaries can shift without any changes at all in the properties of the entities. Entities are similar or different *because of* their set membership.

The adoption of a particular understanding of *set* has important downstream consequences for how researchers interpret relationships among sets. To illustrate,

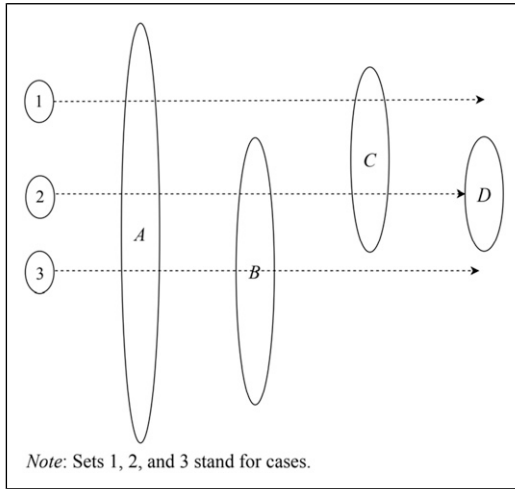


**Figure 1.** Illustration of a spatial set.

let us assume that all members of set *A* are also members of set *B*, but not all members of set *B* are members of set *A* (i.e., *A* is a proper subset of *B*). With property sets, one can hypothesize that the preexisting property shared by the members of set *A* explains why they are always members of set *B*. But one cannot hypothesize the same with spatial sets, because the members of set *A* may not share any preexisting properties. Instead, with spatial sets, one can hypothesize that the act of categorizing and understanding entities as members of set *A* explains why they are always members of set *B*. With spatial sets, the constitution of the set can cause heterogeneous entities to exhibit similarities and causal relations that they otherwise would not have.

Constructivist set-theoretic analysts treat all social categories as spatial sets. They reject the property-possession assumption and embrace the *spatial-set assumption*: the belief that social categories exist as spatial sets in human minds. Constructivist set-theoretic analysts elucidate the meaning of social categories by exploring their set-membership relations with other social categories, which are also treated as spatial sets. They formulate propositions by making statements about the set-membership relations that exist among social categories treated as spatial sets. These propositions are, or use, generalizations about spatial relations among sets. Constructivist set-theoretic analysts evaluate their propositions by gathering evidence that is analyzed under the spatial-set assumption; that is, the evidence is defined by its membership in one or more spatial sets. To practice constructivist set-theoretic analysis is to move into a world of spatial sets and spatial-set relations, leaving behind the experienced world of substances, properties, and property relations.

Set-theoretic analysts often understand and define the boundary line that demarks a spatial set in terms of other spatial sets. That is, the boundaries of two or more *source* sets establish the boundary for membership in a *target* set of interest. In [Figure 2](#), an entity achieves membership in the target set *D* only if it can pass through the logically prior source sets (we assume that the entity



**Figure 2.** A sequence of spatial sets.

moves horizontally across space from left to right). We may hypothesize that an entity has membership in *D* because of its prior membership in the more general source sets *A*, *B*, and *C*. If an entity lacks full membership in any of these source sets, the entity will lack full membership in *D*. But if an entity is a member of *A*, *B*, and *C*, its membership in these sets will always lead it to be a member of set *D*.

In [Figure 2](#), the boundary of set *D* could be *defined* as the intersection space of the more general sets *B* and *C*, in which case *B* and *C* would be the *constitutive sets* of *D*. The boundaries of *B* and *C* themselves could be defined by appealing to the still more general source sets of which they are members. In turn, these new boundaries could be defined by appealing to yet more sets. The boundary-establishment process does not bottom out with property sets or the material or physical features of membership units. Rather, with constructivist set-theoretic analysis, it is sets all the way down.

[Figure 2](#) also illustrates the relationship between category boundaries and category prototypes. In the figure, the prototypes of the categories are represented by the letters *A*, *B*, *C*, and *D*, which are located at the center of the sets. The prototypical member of set *D* (case 2) is a full member of sets *A*, *B*, and *C*, but it does not correspond with the prototypes of these source sets. In fact, the prototypical member of set *B*—case 3—falls outside of set *D*. Hence, while membership in set *B* is necessary for membership in set *D*, the prototypical member of set *B* is not even a member of set *D*.

This imagery, in which spatial sets constitute, are constituted by, and are spatially and temporally related to other spatial sets, captures the foundation of

constructivist set-theoretic analysis. Constructivist set-theoretic analysis avoids psychological essentialism by understanding social categories under the spatial-set assumption rather than the property-possession assumption. The members of social categories (e.g., all scientists, all marriages, all crimes) are similar by virtue of their membership in the same conceptual spaces, not by virtue of any mind-independent properties—whether internally or externally derived—that they possess.

The challenge for social researchers is to treat the categories under analysis as spatial sets that exist in the mind and that are ontologically prior to the entities they classify. Although social scientists cannot hope to avoid psychological essentialism in all of their scholarly communication, they can aim for the non-essentialist analysis of the main categories under study.

### 3. A New Constructivist Social Science

This section explores how constructivist set-theoretic analysis can be used to generate social-scientific knowledge. It focuses special attention on the identification of causal regularities within particular semantic communities. My concern throughout is with blending a constructivist ontology into a set-theoretic methodology. Constructivism enriches set-theoretic analysis by ridding it of essentialist foundations, thereby allowing for valid inference with social categories. At the same time, a set-theoretic approach enriches constructivism through the introduction of a well-developed scientific methodology for making descriptive and causal inferences. The overall result is a scientific-constructivist approach in which a constructivist ontology joins forces with a scientific epistemology.

#### 3.1. Constructivist Set-Theoretic Analysis

Like natural scientists, social scientists seek to generate epistemologically objective truths about the world through the explicit use of evidence and scientific methods ultimately grounded in logic. Yet, in the social sciences, the scope within which a proposition can be evaluated as true is relative to the semantic community in which that proposition bears a particular meaning. This dependence of truth on semantic context does not commit one to a radical epistemological relativism (see [Boyd 1990](#); [Sayer 2000](#)). It does not mean that the same proposition is true in one semantic community and false in another. Rather, the implication is that a proposition requires a particular semantic context in order to embody a certain meaning and *exist* as a certain kind of proposition. Outside of this context, the proposition carries a different meaning, and it is not the same proposition. Thus, the *truth* of a proposition does not shift from one community to the next. Instead, the *existence* of the proposition itself is relative to particular semantic communities.

Constructivist set-theoretic analysts seek to define, use, and code their categories in ways that are well understood by and meaningful to their readers. These readers are usually the principle semantic community of interest. The analyst seeks to construct shared conceptual spaces for categories of interest within this community. Outside of the community, category definitions and meanings may be different from those stipulated by the set-theoretic researcher. Hence, outside of the community of readers, the same propositions may not exist.

Constructivist set-theoretic researchers do not arbitrarily stipulate category definitions and meanings. Rather, to avoid the non-arbitrary coding of categories, researchers explicitly follow existing category usage in relevant communities. These existing communities may be specialized academics who already agree about the meaning of scholarly terms. They may also include non-academic groups who employ categories in clear and consistent ways within their communities.

To elucidate the meaning of categories within communities, constructivist set-theoretic analysts use a broadly *interpretive* approach. Interpretation is needed for calibrating the boundaries of categories and for coding whether specific cases are members, non-members, or partial members of categories (Ragin 2008; Schneider and Wagemann 2012; Oana, Schneider, and Thomann 2021). Interpretive methods are especially useful when social categories reference the intentions of actors (e.g., *vote*, *protest*, *discriminate*) and/or the values of actors (e.g., *democracy*, *development*, *terrorism*). One cannot understand the meaning of such categories without understanding the goals and values of the actors referenced by the categories.

Yet the ultimate goal of constructivist set-theoretic analysis is not primarily interpretive; constructivist set-theoretic analysts do not stop with the elucidation of category meanings and the drawing of category boundaries. Instead, they seek to generalize about regularities that exist among social categories. Their goal is to generate valid knowledge about category relationships—including causal relationships—in the social world.

With set-theoretic analysis, one does not study how changes in the level or magnitude of a given property in the natural world relate to changes in the level or magnitude of another property in the natural world. One does not formulate generalizations about the covariations that exist among properties in the natural world. Instead, one studies *patterns of spatial overlap* among social categories that are understood to exist as bounded regions in abstract mental space. One formulates generalizations about social category associations on the basis of their *spatial relationships*. One identifies and summarizes these spatial relationships using set-theoretic methods that focus on subset relations and degrees of set overlap. Crucially, the set-theoretic methods used to study patterns of spatial overlap yield findings that are different from and cannot be replicated with variable-oriented methods.

Three types of spatial-set relationships are at the core of constructivist set-theoretic analysis. The first relationship is a spatial pattern in which nearly all members of a target set of interest  $Y$  are members of a source set  $X$ . Here the source set  $X$  is an approximate superset of the target set  $Y$ . With this relationship, membership in  $X$  is nearly *necessary* for membership  $Y$ .<sup>13</sup> The second relationship is a spatial pattern in which nearly all simultaneous members of two or more source sets (e.g.,  $A$ ,  $B$ , and  $C$ ) are also members of a target set (e.g.,  $Y$ ). That is, the space created by the intersection  $ABC$  forms an approximate subset of  $Y$ . With this relationship, membership in the combination  $ABC$  is nearly *sufficient* for membership in  $Y$ . The third relationship is a spatial pattern in which set  $X$  and set  $Y$  share nearly identical members. Here the boundaries of the two categories are constructed in such a way that they almost perfectly overlap. That is, membership in  $X$  is nearly *necessary and sufficient* for membership  $Y$ . The set theory foundations of all of these necessity/sufficiency relationships are well established in both the literature on set-theoretic analysis (e.g., [Schneider and Wagemann 2012](#)) and the earlier literature on logic and set theory (e.g., [Stoll 1961](#)).

Scholars may seek to transform spatial sets into regular variables so that they can study them under the property-set assumption using mainstream statistical methods. However, the findings generated when spatial sets are transformed into variables do not parallel those generated with set-theoretic methods. Set-theoretic findings are fundamentally different from statistical findings. The literature on set-theoretic methods is replete with examples highlighting the differences between associations of necessity and/or sufficiency versus associations of statistical covariation ([Ragin 1987, 2008](#); [Schneider and Wagemann 2012](#); [Thiem, Baumgartner, and Bol 2016](#); [Whitaker et al. 2020](#)). The roots of these differences reside in part with the different mathematical principles associated with Boolean algebra versus linear algebra ([Thiem, Baumgartner, and Bol 2016](#)).

For example, a measure of set coincidence is fundamentally distinct from a measure of correlation. The *degree of set coincidence* between two social categories is measured as the number of cases that are members of both categories (i.e., set intersection) divided by the number of cases that are in either category (i.e., set union).<sup>14</sup> Two categories may have a high degree of set coincidence but not be correlated when analyzed as variables ([Ragin and Fiss 2017, 102-4](#)). Set coincidence and correlation are different metrics that measure different kinds of association under different assumptions about the nature of the entities being measured. If a researcher discovers that two spatial sets exhibit high set coincidence, it does not follow that two variables derived

<sup>13</sup>I focus on approximate relations of necessity and/or sufficiency (e.g., “nearly necessary”) because social scientists rarely discover exceptionless patterns.

<sup>14</sup>Set coincidence can be used with either dichotomous or continuous/fuzzy sets.

from those spatial sets will be highly correlated statistically. Just as statistics does not replicate or make formal logic redundant, variable-oriented methods do not encompass or reproduce set-theoretic methods.

Thus, variable-oriented scholars using standard statistical methods cannot replicate the findings of constructivist analysts using set-theoretic methods. The transformation of spatial sets into variables embodies a loss of information and can introduce systematic error. Scholars must resist the temptation of assuming that variable-oriented methods could yield results that parallel set-theoretic results or that could easily be cast as constructivist results.

### 3.2. A Regularity Theory of Causality

What is the constructivist set-theoretic approach to causality? It is not the dominant potential outcomes framework of the social sciences, which uses an interventionist/counterfactual theory of causality (e.g., [Rubin 1974](#); [Holland 1986](#); [Morgan and Winship 2007](#) [2015]). Constructivists reject the interventionist/counterfactual theory of causality because it requires homogeneity assumptions that are radically violated with social categories. For example, in [Rubin's \(1974\)](#) famous model, inferences depend on the validity of the stable unit treatment value assumption (SUTVA). This assumption requires that a treatment have the same form across all treated units. Yet, with social phenomena, the instances of treatment  $X$  (e.g., a type of public policy) are heterogeneous in their natural kind composition in ways that social scientists cannot even begin to model. The requirements needed to assess causality with interventionist theories of causality are not satisfied if the interventions (or the other variables or the units themselves) are not at least approximate natural kinds.

Instead of an interventionist theory, constructivist set-theoretic researchers use a *regularity theory of causality* (e.g., [Psillos 2002](#); [Baumgartner 2008, 2013](#)). In using this theory, these researchers agree to share a particular understanding of the social category *causality*. With a regularity theory, causality exists between social category  $X$  and social category  $Y$  if, on the basis of investigation, three conditions are believed to obtain: (1) *temporal succession* ( $X$  precedes  $Y$  in time), (2) *spatiotemporal contiguity* ( $X$  makes direct or indirect contact with  $Y$  in space and time), and (3) *constant conjunction* ( $X$  is part of the logically minimized solution set that is almost always constantly conjoined with  $Y$ ). With this approach, researchers treat a relationship among social categories as a *causal relationship* if they believe it is a member of the social category *relationship between categories that features temporal succession, spatiotemporal contiguity, and constant conjunction*.

When a regularity theory of causality is used in the social sciences, analysts often discover that no single condition is almost always followed by the outcome. Instead, analysts find that multiple categories are needed to explain

interesting outcomes. Accordingly, they focus on *combinations* of social categories that are approximately sufficient for an outcome category (Ragin 1987).<sup>15</sup> These combinations represent the intersection space that is shared by the categories. Analysts often find that multiple combinations are each approximately sufficient for the outcome (e.g.,  $ABC \vee DEF \rightarrow Y$ ). Each distinct combination acts like an alternative causal recipe leading to the same outcome. This overall concern with *sufficiency combinations* connects constructivist set-theoretic analysis to Qualitative Comparative Analysis (QCA) and other set-theoretic methodologies that use logical minimization techniques to remove redundancies and arrive at parsimonious solution sets (see Baumgartner 2008, 2013; Ragin 2008; Schneider and Wagemann 2012). These methodologies employ logical rules to weed out redundant factors that do not play an essential role in the explanation of the outcome. In principle, the final solution sets they generate consist of only conditions that are necessary parts of combinations of conditions that are sufficient for the outcome (Mackie 1965).

The concern of the regularity theory of causality with connecting  $X$  and  $Y$  in space and time adds a sequential and causal chain dimension to this approach. Analysts typically establish this connection through the analysis of a series of intervening categories—often called *mechanisms*—that stand between and link  $X$  and  $Y$  (Glennan 2009). The specification of mechanisms involves identifying a sequence of regularities among a series of temporally ordered social categories. Given this sequential component, analysts often see sufficiency combinations as particular *causal paths* leading to the outcome of interest. In some versions of QCA, analysts build temporality into their notation such that  $ABC \rightarrow Y$  is not the same causal path as  $BCA \rightarrow Y$ .

Unlike an interventionist theory of causality, a regularity theory of causality is fully compatible with a constructivist ontology. Indeed, the temporal and spatial aspects of a regularity theory of causality depend on collective understandings for their existence. For instance, with the first component concerning time order, analysts and their readers normally share a Newtonian understanding of absolute time. That is, the idea that  $X$  precedes  $Y$  in time assumes that  $X$  comes before  $Y$  in standard calendar time. While this non-relativistic approach to time might pose problems for scientific work on natural kinds, the same is not true for constructivist scholarship in the social sciences. Nearly all Western semantic communities—including social science communities—adopt a calendar understanding of time by default.

---

<sup>15</sup>Insofar as regularity theorists are interested in individual conditions, it is because certain individual conditions are nearly necessary for the outcome of interest (or occasionally nearly sufficient). That is, certain individual conditions almost always appear in the configurations of conditions that are constantly conjoined with the outcome.



Likewise, the idea that a physical connection exists between social category  $X$  and social category  $Y$  is not a mind-independent fact about nature. Rather, the social fact that  $X$  physically connects to  $Y$  depends on shared understandings of time and space within a community of language users. For instance, with the sequence  $X \rightarrow M_1 \rightarrow M_2 \rightarrow Y$ , the researcher must specify the time at which each category begins and ends on the basis of collective understandings within a particular semantic community. Likewise, the physical location of each category in the sequence must be specified according to how individuals understand the existence of the referents of the category within physical space. The fact that  $X$  is spatially adjacent to  $M_1$  depends on the meanings attributed to  $X$  and  $M_1$ , including their physical positions in the world. For example, consider the assertion that regular student attendance at lecture ( $X$ ) is a necessary causal condition for receiving a grade of A on the final exam ( $Y$ ). This statement assumes that a community of language users share an understanding of the category *lecture* as a space in the world at which students can be physically present and come into sensory contact with information relevant for the exam. Students physically carry the information retrieved at this setting in their computers, notebooks, and brains; the information then influences the answers that they physically record on the exam. These answers are the sensory input that the exam graders consider when assigning a letter grade to the exam. The general point is that constructivist set-theoretic researchers establish both the temporal location and the physical positioning of social categories on the basis of shared understandings of how those categories exist in time and space.

The move to a constructivist set-theoretic social science is a major departure from an essentialist variable-oriented social science. While the two approaches share a commitment to a scientific epistemology, they differ in their ontologies. Their different ontologies are associated with different kinds of research questions, different kinds of research methods, and different kinds of research findings. When deciding between the two approaches, the issue is not which kinds of questions, methods, and findings one prefers for the social sciences. Rather, the issue is which ontology is appropriate and necessary for the social sciences. In this paper, I have explored why a constructivist ontology is appropriate and necessary for the social sciences—if the social sciences seek to function as a science. Like it or not, the kinds of questions, methods, and findings associated with a constructivist ontology are the ones suited for the scientific study of social categories.

#### **4. Substantive Illustration: Intelligence, Race, and Poverty**

This section compares three studies of intelligence, race, and poverty: Herrnstein and Murray's *The Bell Curve* (1994), Fischer et al.'s *Inequality by*

*Design* (1996), and Ragin and Fiss's *Intersectional Inequality* (2017). I select these three studies because they represent three kinds of social science. Herrnstein and Murray explicitly adopt an innate essentialist approach, viewing attributes such as intelligence and race as genetically acquired properties of individuals. Fischer et al. reject this innate essentialism, but they still follow social essentialism and embrace the property-possession assumption. Finally, Ragin and Fiss use set-theoretic analysis in a way that largely avoids the property-possession assumption and is consistent with scientific-constructivist research.

#### 4.1. What Do Test Scores Measure?

All three studies use test scores from the Armed Forces Qualification Test (AFQT) as data in their analyses. Yet the way in which they understand the meaning of test scores differs greatly. Herrnstein and Murray treat the AFQT scores as a measure of intelligence, which they see as a property possessed to differing degrees by individuals. Their view of intelligence aligns with innate essentialism: invisible micro substances (i.e., genes) cause individuals to have varying cognitive abilities. While the property of intelligence itself cannot be directly observed, it does produce measurable effects, such as scores on the AFQT.

In contrast, the six authors of *Inequality by Design* (1996)—Claude S. Fischer, Michael Hout, Martín Sánchez Jankowski, Samuel R. Lucas, Ann Swidler, and Kim Voss—reject innate essentialism and embrace social essentialist assumptions. They sharply criticize the psychometric idea that intelligence is a single, fixed trait that exhibits a bell-curve distribution among a population of individuals. Instead, they assert that the property measured by AFQT scores is actually “how much instruction people encountered and absorbed” (1996, 62). Thus, whereas Herrnstein and Murray see AFQT scores as a measure of innate properties, Fischer et al. see them as a measure of socially acquired and non-permanent properties. However, both sets of authors understand the AFQT scores as representing real properties (i.e., intelligence or school knowledge) that do not depend on the beliefs of other people for their existence. Neither team questions the property-possession assumption.

Finally, in their book *Intersectional Inequality* (2017), Charles C. Ragin and Peer C. Fiss do not argue that AFQT scores are correlated with any particular individual attribute or trait, including the absorption of previous schooling. They choose to conceptualize AFQT scores not as a measure of a separate construct or human property, but rather as the construct *AFQT scores* itself. Ragin and Fiss construct two categories: *low-AFQT-score-person* and *high-AFQT-score-person*. With the authors' fuzzy-set analysis, respondents can have membership (100%), no membership (0%), or various degrees of

partial membership (between 0% and 100%) in each of these two categories as well as their negations.

Ragin and Fiss draw the membership boundaries for these categories by following the percentile divisions that the armed forces used in assessing whether individuals are suited for different kinds of military jobs. Because Ragin and Fiss are analyzing the categories *low AFQT score* and *high AFQT score*—as opposed to other categories such as *intelligence*, *absorption of school knowledge*, or *standardized test score*—they argue that it is appropriate to follow the cut-off points of the armed forces. Ragin and Fiss assume that two individuals with a given membership value (e.g., 75% membership in *low-AFQT-score-person*) may possess no similarity other than the fact that they received a similar score on the military qualification exam. If their similar membership value is related to membership in some other category (e.g., *in-poverty-person*), the authors cannot and do not attribute the relationship to knowledge or any property possessed by the respondents. Instead of adopting a property-possession approach to cognitive ability, Ragin and Fiss adopt a spatial-set approach to AFQT scores. *Low AFQT score* and *high AFQT score* are categories that the authors constructed using the military's distinctions as their basis for drawing non-arbitrary boundary lines and assigning heterogeneous respondents with unknown properties into membership groups.

#### 4.2. The Limitations of Essentialist Findings

Herrnstein and Murray seek to identify the effects of intelligence on various life outcomes. Using multivariate regression analysis, they control for a small number of potential confounders, such as family socioeconomic status, race, and education. The authors find that intelligence is an important cause of life outcomes net of other variables. In their explanation of poverty, they summarize their finding as follows: "If you have to choose, is it better to be born smart or rich? The answer is unequivocally 'smart'" (1994, 127). Likewise, they find that intelligence is a significant cause of high school completion, divorce, welfare dependency, parenting, crime, and civic participation. Most controversially, they find that race is an important cause of intelligence. Herrnstein and Murray understand these findings using the property-possession assumption: intelligence and race are properties that individuals physically possess, and the possession of these properties shapes individual decisions and behaviors in ways that show up in social statistics for large populations.

From a scientific-constructivist viewpoint, these findings are rooted in the false assumption that individuals possess a property of intelligence (or race). In fact, two individuals with the same test score (or the same race) may not possess any shared properties beyond those that constitute them as human beings. The invalidity of the property-possession assumption explains why

other authors have not reached similar conclusions in their research on intelligence, race, and poverty. For example, Fischer et al. (1996) find that formal schooling and adolescent environment are both *more* important than AFQT scores. Many other authors have also reached findings that contradict those presented in *The Bell Curve* (see Rushton 1997). Such inconsistent findings are what one might expect if variables such as test scores and race do not correspond to homogeneous properties possessed by individuals.

While Fischer et al. reject biological or innate essentialism, their findings remain rooted in the property-possession assumption. They still assume that individuals with similar test scores possess a similar underlying property related to educational absorption and mental self-management. Fischer et al. assume that this similar property is efficacious and can cause other properties among these individuals, such as poverty avoidance. Moreover, they assume that individuals who share other social experiences, such as exposure to certain kinds of parents and peers, possess additional efficacious properties. These properties lead them to behave in ways such that they possess still more objective properties, such as income and employment.

From a scientific-constructivist perspective, the essentialist assumptions of Fischer et al. lead to problematic research findings. The lack of homogeneity in variables yields findings that are fragile and sensitive to particular model specifications. Fischer et al.'s final statistical model has nearly 30 variables and is virtually impossible to interpret. As variables are removed, the coefficients change and with them conclusions about their effects. The Fischer et al. results are part of a larger series of regression studies on the causes of poverty in the United States whose findings are not widely regarded as true (Ma and Schapira 2017). The basic problem with the Fischer et al. study—like the Herrnstein and Murray study and many other quantitative studies of the causes of poverty—is its erroneous essentialist assumptions. These assumptions mask the violation of homogeneity requirements needed for valid causal inference. In turn, the violation of these homogeneity requirements casts serious doubt on the causal findings reported in Fischer et al.'s book and countless other studies of poverty in the United States.

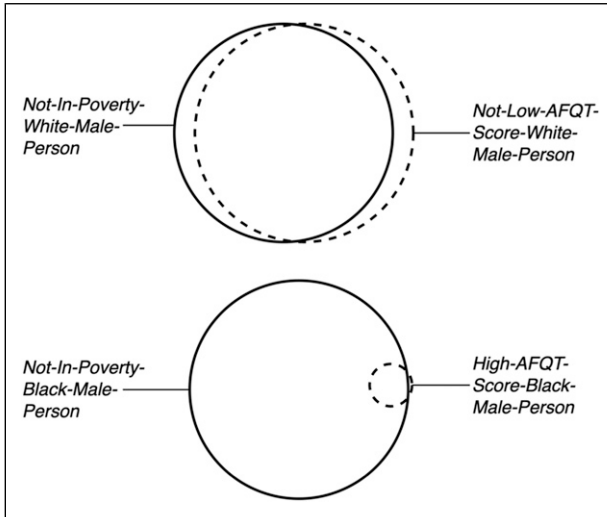
#### 4.3. Findings from Constructivist Set-Theoretic Analysis

Ragin and Fiss (2017) do not seek to discover the net effects of test scores, intelligence, educational background, or any other characteristic possessed by individuals. Instead, they are interested in the ways in which constructed categories overlap with one another and form approximate subset relations. They are particularly interested in identifying typical causal paths to the outcomes *in-poverty-person* and *not-in-poverty-person*; they understand a causal path as membership in a particular combination of categories that is almost always followed by membership in the outcome. The outcome

categories *in-poverty* and *not-in-poverty* are constructed and calibrated within the semantic context of “official” poverty thresholds. The authors calibrate these categories using cut-offs and distinctions created by the US Department of Health and Human Services and the Census Bureau (2017, 65-68). In terms of causal conditions, they focus not only on degree of membership in *low-AFQT-score-person* and *high-AFQT-score-person*, but also membership in the categories *Black person*, *White person*, *female*, *male*, and a few others related to social and economic class. They make no assumptions about the properties of the individuals who have been assigned membership in these sets. Thus, they do not assume any intrinsic differences between an individual who the Census Bureau assigns membership in *White female* and an individual to whom they assign membership in *Black male*. Ragin and Fiss assume only that these individuals differ with respect to their membership in these constructed categories; the ascribed categories make the people different.

Ragin and Fiss examine the degree to which any two categories overlap with one another using a set coincidence measure—which, again, is not the same thing as a correlation measure. They find that membership in the set *not-low-AFQT-score* and membership in *not-low-income-parents* have high set coincidence for individuals in the category *White person*, whereas membership in *not-high-AFQT-score* and membership in *not-high-income-parents* have high set coincidence for individuals in the category *Black person* (2017, 92-96). What this means is that if an individual is in the category *White person*, it is hard to separate the individual’s membership in *not-low-AFQT-score* from their membership in *not-low-income-parents*, because these categories overlap so closely. Likewise, if an individual is in the category *Black person*, it is hard to separate the individual’s membership in *not-high-AFQT-score* from their membership in *not-high-income-parents*. Ragin and Fiss show that these overlapping sets “should be considered jointly and not treated as separate or ‘independent’ . . . it is hazardous to try to separate the effects of test scores and parental income on poverty” (2017, 98-99). They also emphasize that this set coincidence is not visible in a correlational analysis of the properties of race, parental income, and test scores (2017, 96).

Ragin and Fiss discover various ways in which membership in the category *Black person* is disadvantageous compared to membership in *White person*. For example, individuals in the *White male* category usually have membership in *not-in-poverty* if they also have membership in *not-low-AFQT-score*. By contrast, individuals in the *Black male* category consistently have membership in *not-in-poverty* only if they also have membership in *high-AFQT-score*, as opposed to *not-low-AFQT-score*. This difference is important because membership in *not-low-AFQT-score* is far more common than membership in *high-AFQT-score*. The point is illustrated in Figure 3, which is derived from Ragin and Fiss (2017, 88). In the top part of the figure, membership in *not-low-AFQT-score* is reasonably close to being sufficient for membership in *not-*



**Figure 3.** Euler diagrams for test scores, race, and poverty. Note. *Not-In Poverty-White-Male-Person* is .09 contained within *Not-Low-AFQT-Score-White-Male-Person*. *Not Low-AFQT-Score-White-Male-Person* is .83 contained within *Not-In-Poverty-White-Male-Person*. *High-AFQT-Score-Black-Male-Person* is .97 contained within *Not-In-Poverty-Black-Male-Person*. *Not-In Poverty-Black-Male-Person* is .06 contained within *High-AFQT-Score-Black-Male-Person*. Source: Ragin and Fiss (2017, 88).

*in-poverty* among individuals in the *White male* category. The same is true for membership in *high-AFQT-score* among individuals in the *Black male* category, in the bottom part of the figure. Yet, whereas *not-low-AFQT-score* is a common approximately sufficient condition for poverty avoidance, *high-AFQT-score* is a rare approximately sufficient condition for poverty avoidance.

Ragin and Fiss (2017) use a regularity theory of causality in their inter-sectional analysis of antecedent categories that are conjoined to *not-in-poverty*. They focus on the outcome *not-in-poverty* rather than *in-poverty* because they discover that these outcomes have different antecedents, a point often not recognized or well understood in the existing literature (2017, 115–20). They explore how the causal combinations that lead to membership in *not-in-poverty* vary across individuals with membership in *Black female*, *White female*, *Black male*, and *White male*. They examine the extent to which distinct combinations of four particular factors are sufficient for membership in *not-in-poverty* for each of these four sets of individuals. The four causal factors are (1) *favorable-family-background*, (2) *not-low-AFQT-score*, (3) *educated-person*, and (4) *favorable-domestic-situation*. Ragin and Fiss are

interested in identifying which combinations of these four factors are consistently sufficient for *not-in-poverty* for each of the four categories of people.

The authors present many of their set-theoretic findings using Euler and Venn diagrams that visually illustrate patterns of spatial overlap and approximate subset/superset relations. These findings include the following:

1. Individuals with membership in *White person* have a greater number of causal pathways to *not-in-poverty* than individuals with membership in *Black person*.
2. Individuals with membership in *White person* are more likely to be members of the individual categories associated with these causal pathways than individuals with membership in *Black person*.
3. The category *educated-person* is a part of every causal combination leading to *not-in-poverty*. The condition *favorable-domestic-situation* appears in more than half of the causal combinations leading to *not-in-poverty*.
4. *Not-low-AFQT-score* always appears in tandem with *educated-person* in the causal pathways leading to *not-in-poverty*.
5. *Favorable-family-background* appears in most of the causal pathways leading to *not-in-poverty* for individuals with membership in *Black person*.
6. Individuals with membership in all four categories are highly likely to achieve membership in *Not-In-Poverty* Regardless of Whether they Are Members of *Black person*, *White person*, *Female Person*, or *Male Person*.
7. Individuals with membership in only one or two categories are far more likely to achieve membership in *not-in-poverty* if they are members of *White person* rather than *Black person*.
8. Individuals with membership in *Black female person* follow a pathway to *in-poverty* (as opposed to *not-in-poverty*) if they have membership in *not-low-AFQT-score*, *unfavorable-domestic-situation*, and either *not-educated* or *unfavorable-family-background*.

These *kinds* of findings are radically different from the correlational and net effects findings in the variable-oriented studies of [Hernstein and Murray \(1994\)](#) and [Fischer et al. \(1996\)](#). These set-theoretic findings are not intended to estimate the average effect of a change on a variable for poverty outcomes. Instead, they are intended to help readers understand regularities in the social world. They do so by summarizing the ways in which constructed categories overlap with other constructed categories. The findings are contingent on a semantic context in which particular categories mean certain specific things. However, the findings are not subject to essentialist biases. Ragin and Fiss escape essentialism insofar as their categories correspond with cognitions in

their minds and the minds of those readers who accept their category definitions and calibrations.

Finally, the Ragin and Fiss findings allow us to reflect on what it means for community members (including the researchers themselves) to be placing individuals within social categories. Their findings emphasize the profound consequences for inequality of sorting people into different racial categories. Ragin and Fiss expose some of the details of these consequences without insinuating that membership in categories for racial groups and test scores has anything to do with properties possessed by individuals. Likewise, while they include *high-AFQT-score-person* and *low-AFQT-score-person* in their analysis, they do not use or analyze categories such as *intelligent person* or *dumb person*. The idea of intelligence is not discussed in their work, beyond its usage in other scholars' studies. With constructivist set-theoretic analysis, social categories such as *intelligent person* are regarded as wholly inappropriate if they are intended to reflect coherent internal properties possessed by individuals. A constructivist set-theoretic analyst could study *intelligent person* and *dumb person*, but these categories would be understood as beliefs and understandings within a specific community or society, as opposed to representing any coherent and shared features possessed by individuals, whether internally or relationally derived.

## 5. Conclusion

When seen in appropriate historical context, the quest for an approach to the analysis of social categories that rejects essentialism but still embraces a scientific orientation and causal analysis is neither new nor radical. In important respects, Max Weber pursued this very quest in his early writings on causation and explanation at the beginning of the 20th century (Weber 1906 [1949]; see also Heidelberg 2010; Ringer 2002; Wagner and Zipprian 1986). Weber followed Rickert (1896 [1986]) in rejecting the idea that social categories are representations of a coherent underlying reality. Instead, Weber believed that categories impose meaning and structure onto an infinitely complex reality. His solution for analyzing causation with such categories centered on the case-based analysis of historical particulars. Not coincidentally, many of the set-theoretic tools discussed in this paper have their origins in Weber's methodology for analyzing causality in particular cases (Ragin and Zaret 1983).

Within this broadly Weberian tradition, constructivist set-theoretic analysis is designed to overcome the problem of psychological essentialism in social science research. To use this approach, analysts treat social categories as spatial sets rather than property sets. With spatial sets, the set itself constitutes similarities and differences among units rather than reflecting preexisting



similarities and differences, as with property sets. Heterogeneous entities that do not share preexisting properties can be members of the same set.

The use of the spatial-set assumption for social science is appropriate because the members of a social category do not share natural kind properties that make them members of the category. Instead, they are members of the social category because they all activate a particular conceptual space in the minds of those who share an understanding of the category. Conceptual spaces capture the beliefs and values that human beings use to classify heterogeneous natural entities as members of the same social category. Under this constructivist approach, conceptual spaces are what ultimately make entities similar. When scholars embrace this ontology, they overcome our default psychological-essentialist orientation. They no longer understand the members of a category to be in possession of shared properties that endow them with their common identity and common tendencies.

Because scholars (like other human beings) engage in psychological essentialism by default, constructivist set-theoretic analysis does not come easily or effortlessly. Scholars do not naturally discuss how social categories are members of other social categories; instead, they ordinarily assume that categories capture preexisting divisions among objective entities in the world. For instance, social researchers normally examine how individuals *have* certain genders, incomes, and marital statuses. But with constructivist set-theoretic analysis, they instead discuss how individuals are members of sets for certain genders (e.g., *female person*), incomes (*wealthy person*), and marital statuses (*married person*). Likewise, scholars normally consider how individual properties (e.g., gender, income, and marital status) causally affect other individual properties (e.g., religious orientation, depression, and education) net of everything else. But with constructivist set-theoretic analysis, they instead consider the set-membership relations between source categories (*female person*, *wealthy person*, *married person*) and target categories (*Catholic person*, *depressed person*, *highly educated person*). Set-theoretic methodology provides a full-blown apparatus rooted in logic for objectively assessing membership relations among categories that are dependent on subjective beliefs for their existence and membership boundaries.

The kind of findings that emerge from set-theoretic analysis are scientific generalizations about the regularities that exist among categories, including chains of categories that are linked across time. These regularities can be summarized using the concepts of necessity and sufficiency. They can be descriptive statements (e.g., membership in *free-press country* is almost always necessary for membership in *democratic country*), causal statements (membership in *White person*, *high-test-score person*, and *educated person* is consistently sufficient for subsequent membership in *not-in-poverty-person*), and/or normative statements (membership in *labor exploitation* is sufficient for membership in *morally wrong practice*). Set-theoretic analysts commonly

consider relationships in which membership in specific *combinations* of categories is approximately sufficient for membership in outcome categories. They also often consider whether membership in one particular category is approximately necessary for membership in another category. Constructivist set-theoretic analysts do not engage in the estimation of the net causal effects of individual variables, because they reject the property-possession assumption on which this kind of research depends.

Are we ready for a social science that views the world in a way that does not correspond to how we experience that world? I do not know the answer to this question. But if social scientists are prepared to make the shift away from psychological essentialism to scientific constructivism, set-theoretic methodology provides an invaluable foundation on which to build. Set-theoretic researchers have already developed tools and solutions for some of the important methodological issues facing social research that pursues the analysis of sets (Ragin 2008; Schneider and Wagemann 2012; Oana, Schneider, and Thomann 2021). But the challenge remains for social scientists to treat their categories as spatial sets, and to understand these spatial sets as linked to conceptual spaces in the mind. In principle, meeting this challenge is easy, because it requires only a commitment to a certain way of conceptualizing categories and sets. But in practice, it is very hard; this kind of conceptualization violates highly functional essentialist intuitions that have been shaped by millions of years of biological evolution and thousands of years of cultural evolution.

The good news is that a path forward to a non-essentialist social science does appear to exist, should researchers choose to follow it.

## Acknowledgements

For helpful comments on previous drafts, I thank Laura Acosta, Iza Ding, Jan Fuhse, Qin Huang, Isaac Reed, Rosa Runhardt, Carsten Schneider, Nancy Tropic, the graduate students in my “analytical tools” course, and the anonymous reviewers at *Philosophy of the Social Sciences*.

## Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

## ORCID iD

James Mahoney  <https://orcid.org/0000-0002-4913-4435>

## References

- Ahn, Woo-kyoung, Charles Kalish, Susan A. Gelman, Douglas L. Medin, Christian Luhmann, Scott Atran, John D. Coley, and Patrick Shafto. 2001. "Why Essences Are Essential in the Psychology of Concepts." *Cognition* 82 (1): 59-69.
- Balkenius, Christian, and Peter Gärdenfors. 2016. "Spaces in the Brain: From Neurons to Meanings." *Frontiers in Psychology* 7: 1820.
- Barsalou, Lawrence W. 1999. "Perceptual Symbol Systems" (with Comments). *Behavioral and Brain Sciences* 22 (4): 577-660.
- Barsalou, Lawrence W. 2005. "Abstraction as a Dynamic Construal in Perceptual Symbol Systems." In *Building Object Categories in Developmental Time*, edited by Lisa Gershkoff-Stowe, and David H. Rakison, 389-431. Mahwah, NJ: Lawrence Erlbaum.
- Barsalou, Lawrence W. 2016. "Situated Conceptualization: Theory and Application." In *Perceptual and Emotional Embodiment*, edited by Yann Coelle, and Martin H. Fisher, 11-37. London: Routledge.
- Baumgartner, Michael. 2008. "Regularity Theories Reassessed." *Philosophia* 36 (3): 327-54.
- Baumgartner, Michael. 2013. "A Regularity Theoretic Approach to Actual Causation." *Erkenntnis* 78 (Supplement 1): 85-109.
- Berger, Peter L., and Thomas Luckmann. 1966. *The Social Construction of Reality: A Treatise in the Sociology of Knowledge*. New York: Doubleday.
- Bloom, Paul. 1998. "Theories of Artifact Categorization." *Cognition* 66 (1): 87-93.
- Bloom, Paul. 2000. *How Children Learn the Meanings of Words*. Cambridge, MA: MIT Press.
- Bourdieu, Pierre. 1989. "Social Space and Symbolic Power." *Sociological Theory* 7 (1): 14-25.
- Boyd, Richard. Realism, Approximate Truth, and Philosophical Method. In: C. W. Savage (ed) *Scientific Theories*. Minneapolis: University of Minnesota Press, 1990, pp. 355-91.
- Boyd, Richard. 2010. "Homeostasis, Higher Taxa, and Monophyly." *Philosophy of Science* 77 (5): 686-701.
- Campbell, Donald T. 1958. "Common Fate, Similarity, and Other Indices of the Status of Aggregates of Persons as Social Entities." *Behavioral Science* 3 (1): 14-25.
- Churchland, Paul M. 2012. *Plato's Camera: How the Physical Brain Captures a Landscape of Abstract Universals*. Cambridge, MA: MIT Press.
- Clark, Andy, and David Chalmers. 1998. "The Extended Mind." *Analysis* 58 (1): 7-19.
- Dennett, Daniel C. 1987. *The Intentional Stance*. Cambridge, MA: MIT Press.
- Dewey, John, and Arthur F. Bentley. 1949. *Knowing and the Known*. Boston, MA: Beacon Press.
- Diesendruck, Gil, Rebecca Goldfein-Elbaz, Marjorie Rhodes, Susan Gelman, and Noam Neumark. 2013. "Cross-Cultural Differences in Children's Beliefs about the Objectivity of Social Categories." *Child Development* 84 (6): 1906-17.

- Dupré, John. 1981. "Natural Kinds and Biological Taxa." *Philosophical Review* 90 (1): 66-90.
- Ellis, Brian. 2001. *Scientific Essentialism*. Cambridge: Cambridge University Press.
- Emirbayer, Mustafa. 1997. "Manifesto for a Relational Sociology." *American Journal of Sociology* 103 (2): 281-317.
- Fauconnier, Gilles. 1985 (1994). *Mental Spaces: Aspects of Meaning Construction in Natural Language*. 2nd ed. Cambridge: Cambridge University Press.
- Fischer, Claude S., Michael Hout, Martín Sánchez Jankowski, Samuel R. Lucas, Ann Swidler, and Kim Voss. 1996. *Inequality by Design: Cracking the Bell Curve Myth*. Princeton, NJ: Princeton University Press.
- Fodor, Jerry. 2000. *The Mind Doesn't Work That Way: The Scope and Limits of Computational Psychology*. Cambridge, MA: MIT Press.
- Freedman, David. 1999. "From Association to Causation: Some Remarks on the History of Statistics." *Statistical Science* 14 (3): 243-58.
- Freedman, David. 2008. "On Types of Scientific Enquiry: The Role of Qualitative Reasoning." In *The Oxford Handbook of Political Methodology*, edited by Janet M. Box-Steffensmeier, Henry E. Brady, and David Collier, 300-18. Oxford: Oxford University Press.
- Fuhse, Jan A. 2022. "How Can Theories Represent Social Phenomena?" *Sociological Theory* 40 (2): 99-123.
- Gärdenfors, Peter. 2000. *Conceptual Spaces: The Geometry of Thought*. Cambridge, MA: MIT Press.
- Gärdenfors, Peter. 2014. *The Geometry of Meaning: Semantics Based on Conceptual Spaces*. Cambridge, MA: MIT Press.
- Gelman, Susan A. 2003. *The Essential Child: Origins of Essentialism in Everyday Thought*. Oxford: Oxford University Press.
- Gelman, Susan A. 2013. "Artifacts and Essentialism." *Review of Philosophy and Psychology* 4 (3): 449-63.
- Gelman, Susan A., and Gil Diesendruck. 1999. "A Reconsideration of Concepts: On the Compatibility of Psychological Essentialism and Context Sensitivity." In *Conceptual Development: Piaget's Legacy*, edited by Ellin Kofsky Scholnick, Katherine Nelson, Susan A. Gelman, and Patricia H. Miller, 79-102. Mahwah, NJ: Lawrence Erlbaum.
- Gelman, Susan A., and Steven O. Roberts. 2017. "How Language Shapes the Cultural Inheritance of Categories." *PNAS* 114 (30): 7900-07.
- Giere, Ronald N. 2006. *Scientific Perspectivism*. Chicago: University of Chicago Press.
- Glennan, Stuart. 2009. "Mechanisms." In *The Oxford Handbook of Causation*, edited by Helen Beebe, Christopher Hitchcock, and Peter Menzies, 315-25. Oxford: Oxford University Press.
- Gorski, Philip S. 2013. "What Is Critical Realism and Why Should You Care?" *Contemporary Sociology* 42 (5): 658-70.

- Hackett, Edward J., Olga Amsterdamska, Michael Lynch, and Judy Wajcman, eds. 2008. *The Handbook of Science and Technology Studies*. 3rd ed. Cambridge, MA: MIT Press.
- Hacking, Ian. 1995. "The Looping Effects of Human Kinds." In *Causal Cognition: A Multidisciplinary Approach*, edited by Dan Sperber, David Premack, and Ann James Premack, 351-83. Oxford: Clarendon Press.
- Haslam, Nick. 1998. "Natural Kinds, Human Kinds, and Essentialism." *Social Research* 65 (2): 291-314.
- Haslam, Nick, Elise Holland, and Minoru Karasawa. 2013. "Essentialism and Entitativity across Cultures." In *Culture and Group Processes*, edited by Masaki Yuki, and Marilynn Brewer, 17-37. Oxford: Oxford University Press.
- Haslam, Nick, Louis Rothschild, and Donald Ernst. 2000. "Essentialist Beliefs about Social Categories." *British Journal of Social Psychology* 39 (1): 113-27.
- Haslam, Nick, Louis Rothschild, and Donald Ernst. 2004. "Essentialism and Entitativity: Structures of Beliefs about the Ontology of Social Categories." In *The Psychology of Group Perception: Perceived Variability, Entitativity, and Essentialism*, edited by Vincent Yzerbyt, Charles M. Judd, and Olivier Corneille, 47-59. New York: Psychology Press.
- Heidelberger, Michael. 2010. "From Mill via von Kries to Max Weber: Causality, Explanation, and Understanding." In *Historical Perspectives on Erklären and Verstehen*, edited by Uljana Feest, 241-65. Dordrecht, the Netherlands: Springer.
- Herrnstein, Richard J., and Charles Murray. 1994. *The Bell Curve: Intelligence and Class Structure in American Life*. New York: Simon and Schuster.
- Hoffman, Donald D., Manish Singh, and Chetan Prakash. 2015. "The Interface Theory of Perception." *Psychonomic Bulletin and Review* 22 (6): 1480-1506.
- Holland, Paul W. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81: 945-60.
- James, William. 1896 (1948). "The Will to Believe." In *Essays in Pragmatism*, edited by Alburey Castell, 88-109. New York: Hafner.
- Keil, Frank C. 1989. *Concepts, Kinds, and Cognitive Development*. Cambridge, MA: MIT Press.
- Keleman, Deborah, and Susan Carey. 2007. "The Essence of Artifacts: Developing a Design Stance." In *Creations of the Mind: Theories of Artifacts and Their Representation*, edited by Eric Margolis, and Stephen Laurence, 212-30. Oxford: Oxford University Press.
- Kornblith, Hilary. 1993. *Inductive Inference and Its Natural Ground: An Essay in Naturalistic Epistemology*. Cambridge, MA: MIT Press.
- Krishna, Daya. 1971. "'The Self-Fulfilling Prophecy' and the Nature of Society." *American Sociological Review* 36 (6): 1104-7.
- Kuhn, Thomas S. 1962 (1970). *The Structure of Scientific Revolutions*. 2nd ed. Chicago: University of Chicago Press.
- Lakoff, George. 1987. *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago: University of Chicago Press.

- Lakoff, George, and Rafael E. Núñez. 2000. *Where Mathematics Comes From: How the Embodied Mind Brings Mathematics into Being*. New York: Basic Books.
- Latour, Bruno, and Steve Woolgar. 1979 (1986). *Laboratory Life: The Construction of Scientific Facts*. 2nd ed. Princeton, NJ: Princeton University Press.
- Leslie, Alan M., Fei Xu, Patrice D. Tremoulet, and Brian J. Scholl. 1998. "Indexing and the Object Concept: Developing 'What' and 'Where' Systems." *Trends in Cognitive Science* 2 (1): 10-18.
- Ma, Christine, and Michael Schapira. 2017. *An Analysis of Richard J. Herrnstein and Charles Murray's The Bell Curve: Intelligence and Class Structure in American Life*. London: Routledge.
- Mackie, John L. 1965. "Causes and Conditions." *American Philosophical Quarterly* 2 (4): 245-64.
- McCloskey, Michael. 1983. "Intuitive Physics." *Scientific American* 248 (4): 122-31.
- Medin, Douglas. 1989. "Concepts and Conceptual Structure." *American Psychologist* 44 (12): 1469-81.
- Medin, Douglas, and Andrew Ortony. 1989. "Psychological Essentialism." In *Similarity and Analogical Reasoning*, edited by Stella Vosniadou, and Andrew Ortony, 183-96. Cambridge: Cambridge University Press.
- Merton, Robert K. 1948. "The Self-Fulfilling Prophecy." *Antioch Review* 8 (2): 193-210.
- Morgan, Stephen L., and Christopher Winship. 2007 (2015). *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. 2nd ed. Cambridge: Cambridge University Press.
- Newman, George E., and Joshua Knobe. 2019. "The Essence of Essentialism." *Mind and Language* 34 (5): 585-605.
- Oana, Ioana-Elena, Carsten Q. Schneider, and Eva Thomann. 2021. *Qualitative Comparative Analysis Using R: A Beginner's Guide*. Cambridge: Cambridge University Press.
- Pels, Dick. 2002. "Everyday Essentialism: Social Inertia and the 'Münchhausen Effect.'" *Theory, Culture, and Society* 19 (5-6): 69-89.
- Pitkin, Hanna Fenichel. 1987. "Rethinking Reification." *Theory and Society* 16 (2): 263-93.
- Psillos, Stathis. 2002. *Causation and Explanation*. Montreal, Canada: McGill-Queen's University Press.
- Psillos, Stathis. 2006. "The Structure, the Whole Structure, and Nothing but the Structure?" *Philosophy of Science* 73 (5): 560-70.
- Ragin, Charles C. 1987. *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*. Berkeley: University of California Press.
- Ragin, Charles C. 2008. *Redesigning Social Inquiry: Fuzzy Sets and Beyond*. Chicago: University of Chicago Press.
- Ragin, Charles C., and Peer C. Fiss. 2017. *Intersectional Inequality: Race, Class, Test Scores, and Poverty*. Chicago: University of Chicago Press.

- Ragin, Charles C., and David Zaret. 1983. "Theory and Method in Comparative Research: Two Strategies." *Social Forces* 61 (3): 731-54.
- Reed, Isaac. 2008. "Justifying Sociological Knowledge: From Realism to Interpretation." *Sociological Theory* 26 (2): 101-29.
- Reed, Isaac. 2011. *Interpretation and Social Knowledge: On the Use of Theory in the Human Sciences*. Chicago: University of Chicago Press.
- Rhodes, Marjorie, and Susan A. Gelman. 2009. "A Developmental Examination of the Conceptual Structure of Animal, Artifact, and Human Social Categories across Two Cultural Contexts." *Cognitive Psychology* 59 (3): 244-74.
- Rhodes, Marjorie, and Tara M. Mandalaywala. 2017. "The Development and Developmental Consequences of Social Essentialism." *WIREs Cognitive Science* 8 (4): e1437.
- Rickert, Heinrich. 1896 (1986). *The Limits of Concept Formation in Natural Science: A Logical Introduction to the Historical Sciences*, edited and translated by Guy Oakes. Cambridge: Cambridge University Press.
- Ringer, Fritz. 2002. "Max Weber on Causal Analysis, Interpretation, and Comparison." *History and Theory* 41 (2): 163-78.
- Rosch, Eleanor. 1973. "Natural Categories." *Cognitive Psychology* 4 (3): 328-50.
- Rosch, Eleanor. 1978. "Principles of Categorization." In *Cognition and Categorization*, edited by Rosch, Eleanor, and Lloyd, Barbara B., 27-48. Hillsdale, NJ: Lawrence Erlbaum.
- Rosch, Eleanor. 1999. "Reclaiming Concepts." *Journal of Consciousness Studies* 6 (11-12): 61-77.
- Rosch, Eleanor. 2011. "'Slow Lettuce': Categories, Concepts, Fuzzy Sets, and Logical Deduction." In *Concepts and Fuzzy Logic*, edited by Radim Belohlavek, and George J. Klir, 89-120. Cambridge: MIT Press.
- Rose, David, and Shaun Nichols. 2019. "Teleological Essentialism." *Cognitive Science* 43 (4): e12725.
- Rothbart, Myron, and Marjorie Taylor. 1992. "Category Labels and Social Reality: Do We View Social Categories as Natural Kinds?" In *Language, Interaction and Social Cognition*, edited by Gün R. Semin, and Klaus Fiedler, 11-36. London: Sage.
- Rubin, Donald B. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66 (5): 688-701.
- Rushton, J. Philippe. 1997. "Review Essay." *Society* 34 (3): 78-82.
- Sayer, Andrew. 1997. "Essentialism, Social Constructionism, and Beyond." *Sociological Review* 45 (3): 453-87.
- Sayer, Andrew. 2000. *Realism and Social Science*. London: Sage.
- Schneider, Carsten Q., and Claudius Wagemann. 2012. *Set-Theoretic Methods for the Social Sciences: A Guide to Qualitative Comparative Analysis*. Cambridge: Cambridge University Press.
- Searle, John R. 1995. *The Construction of Social Reality*. New York: Free Press.

- Smith, Edward E., and Daniel N. Osherson. 1984. "Conceptual Combination with Prototype Concepts." *Cognitive Science* 8 (4): 337-61.
- Steinmetz, George. 1998. "Critical Realism and Historical Sociology: A Review Article." *Comparative Studies in Society and History* 40 (1): 170-86.
- Stoll, Robert R. 1961. *Set Theory and Logic*. New York: Dover.
- Strevens, Michael. 2000. "The Essentialist Aspect of Naïve Theories." *Cognition* 74 (2): 149-75.
- Thiem, Alrik, Michael Baumgartner, and Damien Bol. 2016. "Still Lost in Translation! A Correction of Three Misunderstandings between Configurational Comparativists and Regression Analysts." *Comparative Political Studies* 49 (6): 742-74.
- Thomann, Eva, and Martino Maggetti. 2020. "Designing Research with Qualitative Comparative Analysis (QCA): Approaches, Challenges, and Tools." *Sociological Methods and Research* 49 (2): 356-86.
- Thomasson, Amie L. 2003. "Realism and Human Kinds." *Philosophy and Phenomenological Research* 67 (3): 580-609.
- Tsukamoto, Saori, Yoshihisa Kashima, Nick Haslam, Elise Holland, and Minoru Karasawa. 2018. "Entitativity Assumptions of Individuals and Groups across Cultures." In *The Psychological and Cultural Foundations of East Asian Cognition: Contradiction, Change, and Holism*, edited by Julie Spencer-Rodgers, and Kaiping Peng, 335-51. Oxford: Oxford University Press.
- Tversky, Amos. 1977. "Features of Similarity." *Psychological Review* 84 (4): 327-52.
- Wagner, Gerhard, and Heinz Zipprian. 1986. "The Problem of Reference in Max Weber's Theory of Causal Explanation." *Human Studies* 9 (1): 21-42.
- Warglien, Massimo, and Peter Gärdenfors. 2013. "Semantics, Conceptual Spaces, and the Meeting of Minds." *Synthese* 190 (12): 2165-93.
- Weber, Max. 1906 (1949). "Objective Possibility and Adequate Causation in Historical Explanation." In *The Methodology of the Social Sciences*, edited and translated by Edward A. Shils, and Henry A. Finch, 164-188. Glencoe, IL: Free Press.
- Whitaker, Rebecca Garr, Nina Sperber, Michael Baumgartner, Alrik Thiem, Deborah Cragun, Laura Damschroder, Edward J. Miech, Alecia Slade, and Sarah Birken. 2020. "Coincidence Analysis: A New Method for Causal Inference in Implementation Science." *Implementation Science* 15: 108.

## Author Biography

**James Mahoney** is the Gordon Fulcher Professor of Decision-Making in the Departments of Sociology and Political Science at Northwestern University. His most recent book is *The Logic of Social Science* (Princeton University Press, 2021). He is currently writing a book on political revolutions that uses a constructivist set-theoretic approach.