



University of Pennsylvania
ScholarlyCommons


Publicly Accessible Penn Dissertations

2020

The Social Network Dynamics Of Category Formation

Douglas Richard Guilbeault
University of Pennsylvania

Follow this and additional works at: <https://repository.upenn.edu/edissertations>

 Part of the [Cognitive Psychology Commons](#), [Social and Cultural Anthropology Commons](#), and the [Sociology Commons](#)

Recommended Citation

Guilbeault, Douglas Richard, "The Social Network Dynamics Of Category Formation" (2020). *Publicly Accessible Penn Dissertations*. 4233.
<https://repository.upenn.edu/edissertations/4233>

This paper is posted at ScholarlyCommons. <https://repository.upenn.edu/edissertations/4233>
For more information, please contact repository@pobox.upenn.edu.

The Social Network Dynamics Of Category Formation

Abstract

Category systems are remarkably consistent across societies. Stable partitions for concepts relating to flora, geometry, emotion, color, and kinship have been repeatedly discovered across diverse cultures. Canonical theories in cognitive science argue that this form of convergence across independent populations, referred to as 'cross-cultural convergence', is evidence of innate human categories that exist independently of social interaction. However, a number of studies have shown that even individuals from the same population can vary substantially in how they categorize novel and ambiguous phenomena. Contrary to findings on cross-cultural convergence, this individual variation in categorization processes suggests that independent populations should evolve highly divergent category systems (as is often predicted by theories of social constructivism). These puzzling findings raise new questions about the origins of cross-cultural convergence. In this dissertation, I develop a new mathematical approach to cultural processes of category formation, which shows that whether or not independent populations create similar category systems is a function of population size. Specifically, my model shows that small populations frequently diverge in their category systems, whereas in large populations, a subset of categories consistently reach critical mass and spread, leading to convergent cultural trajectories. I test and confirm this prediction in a large-scale online social network experiment where I study how small and large social networks construct original category systems for a continuum of novel and ambiguous stimuli. I conclude by discussing the implications of these results for networked crowdsourcing, which harnesses coordination in communication networks to enhance content management and generation across a wide range of domains, including content moderation over social media and scientific classification in citizen science.

Degree Type

Dissertation

Degree Name

Doctor of Philosophy (PhD)

Graduate Group

Communication

First Advisor

Damon M. Centola

Keywords

creativity, cultural evolution, culture, online experiment, semantics, social networks

Subject Categories

Cognitive Psychology | Social and Cultural Anthropology | Sociology

THE SOCIAL NETWORK DYNAMICS OF CATEGORY FORMATION

Douglas Richard Guilbeault

A DISSERTATION in

Communication

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2020

Supervisor of Dissertation



Damon Centola

Professor of Communication, Sociology & Engineering

Graduate Group Chairperson



Marwan M. Kraidy, Professor of Communication and the Anthony Shadid Chair in Global
Media, Politics and Culture

Dissertation Committee

Kathleen Hall Jamieson, Elizabeth Ware Packard Professor of Communication

Michael Delli Carpini, Oscar H. Gandy Professor of Communication & Democracy

ACKNOWLEDGMENT

As a result of an extremely lucky series of events, I managed to find myself in the fortunate position of pursuing a dissertation whose subject matter I find endlessly fascinating and awe-inducing. To reach this place of privilege requires a level of social support and encouragement that far exceeds the allowable length of such an acknowledgement note. Suffice to say that, to the extent that I have achieved or ever will achieve something genuinely beautiful, it will be a mere testament to the beauty that is demonstrated to me day to day by the special people to whom I have found myself bound.

I am particularly thankful to the students, staff, and faculty at the University of Pennsylvania. Together they have facilitated the ideal Ph.D. experience. Special thanks goes to my committee members who have been extremely helpful, both in terms of philosophical discussions and professional advice. Kathleen Hall Jamieson has provided unfailingly astute guidance in terms of which opportunities and ideas to explore. Her commitment to fighting for the public value of science will continue to inspire me. Similarly, Michael Delli Carpini demonstrated true leadership as the Dean of Annenberg throughout my degree, and I will always look to his clarity of thought as an enduring example to emulate. Lastly, my deepest appreciation extends to my advisor, Damon Centola. When I entered my Ph.D., I was quite fond of the Wittgensteinian maxim that “Whereof one cannot speak, thereof one must be silent.” I am immensely grateful to have received the mentorship of someone who not only deeply understands this maxim, but who also has the creativity and vision to defy it! I feared that I may always remain silent on the scientific study of meaning and collective creativity to which my heart most powerfully compelled me. In working with Damon, he achieved what I believe to be the pinnacle of mentorship. Through intense commitment and Socratic training, he helped me to say something clear and personally meaningful about that which I thought could not be clearly spoken of, and in doing so, he has helped me to build a self-understanding that will forever deepen and enrich my personal

experience of this mysterious world. This dissertation is the longest and most authentic expression to date of my effort to speak of that which I thought impossible to speak of.

ABSTRACT

THE SOCIAL NETWORK DYNAMICS OF CATEGORY FORMATION

Douglas Richard Guilbeault

Damon Centola

Category systems are remarkably consistent across societies. Stable partitions for concepts relating to flora, geometry, emotion, color, and kinship have been repeatedly discovered across diverse cultures. Canonical theories in cognitive science argue that this form of convergence across independent populations, referred to as ‘cross-cultural convergence’, is evidence of innate human categories that exist independently of social interaction. However, a number of studies have shown that even individuals from the same population can vary substantially in how they categorize novel and ambiguous phenomena. Contrary to findings on cross-cultural convergence, this individual variation in categorization processes suggests that independent populations should evolve highly divergent category systems (as is often predicted by theories of social constructivism). These puzzling findings raise new questions about the origins of cross-cultural convergence. In this dissertation, I develop a new mathematical approach to cultural processes of category formation, which shows that whether or not independent populations create similar category systems is a function of population size. Specifically, my model shows that small populations frequently diverge in their category systems, whereas in large populations, a subset of categories consistently reach critical mass and spread, leading to convergent cultural trajectories. I test and confirm this prediction in a large-scale online social network experiment where I study how small and large social networks construct original category systems for a continuum of novel and ambiguous stimuli. I conclude by discussing the implications of these results for networked crowdsourcing, which harnesses coordination in communication networks to

enhance content management and generation across a wide range of domains, including content moderation over social media and scientific classification in citizen science.

TABLE OF CONTENTS

ACKNOWLEDGMENT	II
ABSTRACT	IV
LIST OF ILLUSTRATIONS	VII
CHAPTER 1: THE PUZZLE OF CROSS-CULTURAL CATEGORY CONVERGENCE	1
1.1. Nativism and the Problem of Creative Interpretation	3
1.2. Social Constructivism and the Limits of Path Dependence	9
1.3. A Paradox in the Effects of Network Size on Category Formation	16
CHAPTER 2: A MATHEMATICAL MODEL OF COLLECTIVE CATEGORY FORMATION	22
2.1. Extending the Categories Model	28
CHAPTER 3: THE EMERGENCE OF CROSS-CULTURAL CATEGORY CONVERGENCE IN AN ONLINE SOCIAL NETWORK EXPERIMENT	42
3.1. Experimental Design	43
3.2. Methods of Analysis	47
3.3. Results	49
3.4. Discussion	59
CHAPTER 4: CONCLUSION.....	62
BIBLIOGRAPHY	68

LIST OF ILLUSTRATIONS

Figure 1. A sample of stimuli used in Rosch et al.'s seminal (1976) experiments.

Figure 2. A sample of stimuli used in Shepard & Cermak's (1973) study.

Figure 3. A schematic display of the coordination logic of the categories model.

Figure 4. The main results of the original categories model.

Figure 5. A schematic of how different probability distributions are defined over the set of possible labels from which agents sample the new labels they introduce.

Figure 6. Results of 50 simulations of the categories model comparing the effects of population size on bias and cross-cultural convergence.

Figure 7. Using the hypergeometric distribution to model the effect of population size on the likelihood of labels reaching critical mass.

Figure 8. Screenshots of "The Grouping Game" interface

Figure 9. Using the Zipfean distribution to model the initial frequency of labels.

Figure 10. Comparing the level of convergence in category systems that emerged in small ($N=2$) and large ($N=50$) populations.

Figure 11. Convergence in the vocabularies that emerged in populations of different sizes.

Figure 12. Larger populations amplify the spread of initially frequent labels.

Figure 13. Time series showing the adoption of confederates' rare label ("sumo") by noncommitted subjects (i.e. experimental subjects).

CHAPTER 1: THE PUZZLE OF CROSS-CULTURAL CATEGORY CONVERGENCE

Is the way we categorize the world determined by cognitive universals, or socially constructed? This question, among the oldest in Western philosophy, underlies core lines of inquiry in social science, from how people use categories¹ to assign value to cultural products –e.g. companies (Hannan, Pólos, and Carroll 2007) and art (DiMaggio 1987) – to how marketing campaigns frame beliefs and attitudes in consumer and political contexts (Goffman and Berger 1986; Jamieson 1996). Today, the debate on this question is split between two camps. The first, *nativism*, holds that people independently categorize the world in highly similar ways as a result of innate and universal cognitive processes (Fodor 1998; Gelman 2005; Laurence and Margolis 2002; Medin and Atran 2004; Medina et al. 2011; Pinker 1994, 2003), which are said to account for widespread similarities in the category systems that have emerged among independent populations around the world – referred to here as *cross-cultural category convergence* (Brown 1984; Kauffman 1993; Lant and Mezias 1992; Malt 1995). The second, *social constructivism*, holds that people can vary wildly in how they categorize the world, and that communication amplifies this variation, leading to highly divergent category systems (Berger and Luckmann 1967; Burr 2003; David 2007; DiMaggio 1987, 1997; Searle 1995; Shaw 2015; Smith 2010). For this reason, social constructivism is widely held to be incompatible with cross-cultural category convergence, because of its implication that

¹ I follow the work Rosch et al. (1976) in defining a category as “a number of objects which are considered equivalent” by virtue of being “designated by a name” (or more generally, by a label). I prefer this definition because of its simplicity and the opportunities it affords for cross-disciplinary synthesis, since this definition has furnished some of the more influential formal work on the theory of categories in sociology (Hannan et al, 2007) and anthropology (Brown 1984). Crucial to this definition is the idea that a category is not simply the word itself, nor the set of objects in the world, but rather the mapping between the label and the objects, where this mapping signifies that a set of objects are perceived as members of a common “kind” with a shared essence.

communication inherently leads to path dependence (Malt 1995; Pinker 2003; Pinker and Morey 2014). My central thesis is that social processes provide a more compelling explanation of how cross-cultural category convergence emerges than nativism; however, I argue that social processes give rise to cross-cultural category convergence in deterministic ways that are largely incompatible with canonical views of social constructivism, indicating the need for a new theoretical approach. In what follows, I use mathematical modeling and online social network experiments to argue that communication in social networks can generate either divergence or convergence in category systems across independent populations, as a function of social network size.

I begin this dissertation by reviewing the longstanding debate about the nature of category formation between nativism and social constructivism. In reviewing this debate, I reveal an unresolved paradox concerning the effects of social network structure on the emergence of cross-cultural category convergence. In particular, I find a number of studies proposing that increasing network size should increase cultural variation in category systems, while a number of other studies indicate that increasing social network size should increase cross-cultural category similarities. To address this paradox, I transition into chapter 2, where I develop a novel formal model of category formation in social networks which identifies the conditions under which coordination in social networks can lead to the divergence of category systems (contrary to the nativist position), and the conditions under which it can lead to the convergence of category systems (contrary to both the nativist position and the constructivist position). These findings offer a new interpretation of past observational data on cross-cultural similarities in category systems. I suggest that rather than cross-cultural convergence providing evidence of innate, universal cognitive categories, instead it may indicate that

communication in large social networks filters cognitive and lexical diversity in such a way that promotes the development of similar category systems across diverse populations. Chapter 3 focuses on an experimental validation of this model, and the concluding chapter details the implications of this perspective, and the future studies and analyses I will pursue to expand its logic.

1.1. Nativism and the Problem of Creative Interpretation

The theory of cross-cultural category convergence derives from a large body of cross-cultural linguistic data indicating striking similarities in the size, structure, and content of category systems that have emerged among distinct cultures (Brown, C. 1979, 1984; Brown, D. 2004; Brown and Witkowski 1981; Goddard 2008; Youn et al. 2016). The subfield of cognitive anthropology has focused on documenting examples of cross-cultural category convergence as evidence for a panhuman structure of the mind. Most representative of this effort is Brown's (2004) widely cited "List of Classification Universals", which details a long list of semantic domains for which separate cultures have been consistently found to produce highly similar vocabularies and semantic partitions (i.e. organizations of the domain into subtypes and relations). Examples include geometry (Burris 1979), flora (Brown 1986), fauna (Malt 1995), body parts (Brown and Witkowski 1981), music (Trehub 2015), weather and geology (Youn et al. 2015), and more controversially, emotions (Jackson et al. 2019; Pinker 2003), gender (Brown 2004), kinship (Kemp and Regier 2012; White 2012 [1963]), and race (Gil-White 2001).

The predominant explanation for why cross-cultural category convergence is observed for such a wide array of semantic domains is summarized by Brown in his claim that: “Whatever is constant through all human societies must be due to something that goes with people wherever they go – i.e. their psychobiology” (Brown 2004: 50). While researchers vary in the specific cognitive mechanisms they propose to account for how the mind gives rise to regularities in category formation at the individual-level, the broader claim that similarities in cognitive structure are responsible for cross-cultural category similarities is widely maintained. For instance, a recent study opens with the question – “How universal is human conceptual structure?” – and responds, based on observational, cross-linguistic data showing similarities in vocabulary systems, that: “Semantic clustering structure is independent of culture and environment in many domains” (Youn et al. 2016: 1768).

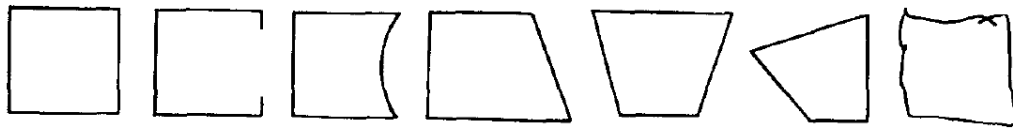


Figure 1. A sample array of stimuli used in Rosch et al.'s seminal (1976) experiments.

The idea that psychobiology can provide a bottom-up account of categorization was solidified as canon in cognitive science by the work of Eleanor Rosch (C B Mervis and Rosch 1981; Rosch 1973, 1975, 2002; Rosch et al. 1976; Rosch and Mervis 1975). Rosch aimed to critique the central view in developmental psychology at the time that, to

a child, the physical and social world is perceived as fundamentally continuous, and that the child imposes categorical structure onto the world only once they are taught names². Rosch argued, to the contrary, that people are born hardwired to perceive the world in terms of certain categories, based on universal similarities in the resolution of sensory perception, as well as similarities in which features people detect and group together. To test whether individual-level cognition drives similarities in category formation, Rosch conducted experiments where separate individuals were given a finite set of stimuli— e.g. an array of shapes, clothing articles, or furniture – and were asked to group them based on a label (e.g. “circle” for the array of shapes), (Fig. 1). She found a high degree of similarity in which images subjects grouped together. This result was seen as evidence for her claim that “human categorization should not be considered the arbitrary product of historical accident or whim, but rather the result of universal psychological principles” (2002: 329).

Importantly, Rosch maintained that while she only showed subjects a small set of discrete stimuli, her results generalized to the case of how people build categories for continuous, novel domains of stimuli; as she writes herself, “Most categories partition domains whose stimuli are not discrete but composed of continuous variation” (1973:

² Rosch quotes the following statement by the anthropologist Edmund Leach as epitomizing this broader view: “The physical and social environment of a young child is perceived as a continuum. It does not contain any intrinsically separate ‘things.’ The child, in due course, is taught to impose upon this environment a kind of discriminating grid which serves to distinguish the world as being composed of a large number of separate things, each labeled with a name” (Leach 1989, p. 34). The idea that the world is inherently continuous to the infant’s mind has firm roots in the history of psychology, reaching at least as far back as William James’ (2000 [1890]) “The Principles of Psychology”, where he poetically describes the world first experienced by the child as highly continuous in this often cited passage: “The baby, assailed by eyes, ears, nose, skin, and entrails at once, feels it all as one great blooming, buzzing confusion; and to the very end of life, our location of all things in one space is due to the fact that the original extents or bigness of all the sensations which came to our notice at once, coalesced together into one and the same space.”

329). Consider, for example, the continuities that frequently characterize domains of cultural products. What meaningfully differentiates a ‘smartphone’ from a ‘tablet’, and a ‘tablet’ from a ‘laptop’? As smartphones get larger, size is no longer a reliable dimension for distinction. Similarly, technologies in each category now frequently possess both touchscreen and call-making functionality. Yet, despite the numerous complexities of such cultural domains, Rosch argued that her theory of cognitive universals could still account for category formation in continuous, multi-dimensional spaces. For this reason, her theory gained high levels of uptake among cognitive anthropologists seeking bottom-up explanations of how distinct societies managed to develop highly similar category systems for domains that would have been highly continuous prior to categorization, such as the domains of geometry or flora.



Figure 2. An example of an amorphous shape (far left) and a sample of how different subjects categorized this shape (to the right of the base image) in Shepard & Cermak's (1973) study.

A key limitation of Rosch's experiments, however, is that she asked subjects to group images from familiar domains (e.g. shapes, clothing, and furniture) using familiar, already-existing labels (e.g. "circle" or "shirt"). As a result, the subjects in her

experiments (all undergraduate students from the same university) may have categorized the stimuli in the same way because they already learned to identify the same objects with common labels through their cultural experience. In other words, processes of social learning still provide a plausible account of her results.

By contrast, a number of studies have shown that even individuals from the same population can vary substantially in how they form categories. In experiments where subjects are shown continuous domains of novel stimuli they have never seen before, people vary considerably in how they label and group together stimuli (Brennan and Clark 1996; Clark and Wilkes-Gibbs 1986; Ranjan and Srinivasan 2010; Shepard and Cermak 1973). For example, Shepard & Cermak (1973) presented subjects with a grid of amorphous shapes (fig. 2), and asked them to group these shapes together and label their groupings. They found significant variation both in which shapes subjects grouped together, as well as in the labels they used for their groupings. These findings have since been shown to be mediated by a number of factors, including personal experience (Spalding and Gregory 1996), expertise (Medin et al. 1997), and cognitive style in feature selection (Medin, Wattenmaker, and Hampson 1987). Altogether, these studies indicate that when people are confronted with a semantic domain for which they have not had prior experience or cultural learning, individuals can vary wildly in how they categorize the space.

The fact that individuals can vary wildly when categorizing novel and continuous domains poses a serious puzzle for the nativist position, especially when considering the widespread view in cognitive science that all domains of worldly objects (physical and social) are highly continuous when first encountered (see footnote 2). This intuition applies equally to the semantic domains of objects for which separate social groups are

said to have achieved cross-cultural category convergence. One experimental study of the Hadza (a hunter-gatherer tribe in Northern Tanzania) concluded the surprising result that individuals from the same tribe can vary in how they categorize colors (Lindsey et al. 2016) – one of the perceptual domains that is most frequently associated with universal and innate human categories (Baronchelli et al. 2010; Guilbeault et al. 2020; Regier, Kay, and Khetarpal 2007). Another relevant domain is geometry, which is associated with consistent naming patterns across over 40 languages (Burriss 1979); and yet, geometry is the epitome of a continuous space that exhibits completely continuous variation along all dimensions (shape, depth, curve, etc.). The puzzle for the nativist perspective, then, is: How can separate social groups arrive at highly similar category systems (i.e. *cross-cultural category convergence*) for continuous domains, if individuals both within and between populations vary substantially in how they perceive and categorize novel and continuous domains?

Contrary to nativism, a longstanding view in social science argues that the purpose of categories is not to accurately describe the world, but rather to communicate with others for the purpose of coordinating perceptions and actions (Berger and Luckmann 1967; Swidler 1986; Wittgenstein 1973). This view entered social science largely by importing the work of Wittgenstein and like-minded pragmatist philosophers (e.g. Dewey and Rorty) who argued that “meaning is use”, implying that the meaning of a category is determined by how it is used in social contexts to communicate with others. This alternative theory of categories contributed to the paradigm of social constructivism, a theory in which categories are viewed as arbitrary in nature and defined relative to the norms, practices, and social structures of a given culture (Berger and Luckmann 1967; Burr 2003; Searle 1995; Smith 2010). In a bold (and yet popular) specification of this

view, communication is said to inherently lead to path-dependent (i.e. divergent and unpredictable) category systems, which only deepens the puzzle of cross-cultural category convergence (Berger and Luckmann 1967, 1967; Blumer 1986; David 2007; DiMaggio 1987; Rawlings and Childress 2019; Shaw 2015). Indeed, social constructivism is commonly viewed as the only reasonable alternative to the theory of cognitive universals; a number of cognitive scientists who challenge the theory of lexical universals similarly provide social constructivist arguments which maintain that communication inherently leads to cultural variation in meaning systems (Clark and Wilkes-Gibbs 1986; Fay et al. 2010b, 2018; Fay and Ellison 2013). However, a more recent version of constructivism (“formalism”) seeks to empirically study the formation of category systems in communication networks, which has paved the way toward to a new solution to the problem of cross-cultural convergence, due in large part to the development of novel methodologies in computational sociology.

1.2. Social Constructivism and the Limits of Path Dependence

The capacity for individuals to vary in how they categorize the world is a core feature of an alternative view of category formation that emphasizes both individual and collective level variation – namely, social constructivism (Berger and Luckmann 1967; Blumer 1986). Constructivism goes the extra step by inferring that, as a result of individual variation, communication in social groups can lead to the adoption of highly idiosyncratic category systems that set a culture on radically unique, path dependent trajectory. This theory has been deemed as ‘foundational’ to social science broadly construed (David

2007), and its canonical role is perhaps most succinctly captured in Smith's (2010) introduction to social construction:

One of the amazing things about human persons is the ability to engage beliefs and ideas in order to creatively form patterns of actions, interactions, and collective social environments. Unlike other animals, a great deal of human social existence is not directly determined by genetic codes or instinctual species behaviors. Instead, human persons are free to use their manifold capacities for representation, belief formation, language, memory, creativity, identity development, and so on variously to shape the meanings and structures of their social existence together. The result is the immense variety, richness, and complexity of human cultures and subcultural meaning systems evident in history and the world today. (119)

In contrast to nativism, constructivism emphasizes divergence in the category systems between social groups. One of the more colorful examples comes from an ancient Chinese encyclopedia in which the animal kingdom is divided into highly idiosyncratic categories, including (a) those that belong to the Emperor, (b) those drawn with a very fine camel's hair brush, (c) those that have just broken a flower vase, (d) those that resemble flies from a distance, and last but not least, (e) mermaids (Borges 1973: 108). Such variation, according to constructivism, is not the exception, but the norm. Countless studies have detailed examples of how communication in social groups can lead to highly idiosyncratic categorizations across a wide range of domains, including art

(Becker 1984; DiMaggio 1987), music (Cerulo 1995; Peterson 1999), technology (Pinch and Trocco 1998), law (Gordon 1984; Gorwa and Guilbeault 2018), business (Hannan et al. 2007), politics (Krippendorff 2005), sexuality (Foucault 1990), race (Allport 1954), fashion (Obukhova, Zuckerman, and Zhang 2014), and even domains of scientific inquiry, from the classification of disease (Bowker and Star 2000; Foucault 1988) to the fundamentals of physics, chemistry, and biology (Collins 1998; Kuhn 1996; Latour 1988; Shwed and Bearman 2010).

Due to its focus on individual variation and path dependence, the theory of social construction has been widely held to assume that communication in social groups can only lead to path dependence, and as a result, this theory is said to be incompatible with data supporting cross-cultural category convergence (Malt 1995; Pinker 2003). Indeed, a particularly strong (and popular) version of constructivism – also known as ‘relationalism’ (Erikson 2013) – goes so far as to argue for an ‘anti-categorical imperative’ in sociological research, where the social world is construed as irreducibly continuous and constantly changing through the interdependence of ideas and actors, such that it is impossible for regularities in category formation to emerge and for these regularities to be examined scientifically (Emirbayer 1997: 298). This style of thinking has also gained expression in the work of sociologists Latour and Woolgar (1986), in their argument that every feature of experience is known only through the application of an arbitrary category, whose meaning is determined solely by the idiosyncrasies of an individual’s perspective, framed within a broader idiosyncratic culture. This strong version of social constructivism deepens the puzzle of cross-cultural category convergence. If individuals vary radically in how they categorize novel continuous domains, and if communication in social groups only serves to amplify this variation, leading to highly divergent cultural

trajectories, then how are the known patterns of cross-cultural category convergence at all possible? Thankfully, this radical view of social constructivism, while popular, is not the only the shop in town.

An alternative approach to sociological research – formalism – provides a theory of social construction that encourages formal inquiry into social dynamics and provides key intuitions that help resolve the paradox of cross-cultural category convergence. Smith (2010) and Erickson (2013) characterize formalism as the view – tracing back to Durkheim (1912) and Simmel (1964) – that processes of social construction are shaped by social structures (e.g. institutions, norms, and social networks) in causally coherent ways. As such, the chief accomplishment of formalism is that it approaches processes of social construction as something that can be modeled and measured scientifically. In his essay “Measuring Meaning Structures,” Mohr (1998) reviews a range of methodologies in sociology that have been developed for the purpose of measuring the effects of social structure on category formation. Many of these use individual-level experiments, similar to Rosch’s paradigm, where members of different societies are tasked with sorting objects into groups, for the purpose of understanding how social status and other cultural factors can influence category schemes (e.g. Shweder and Bourne 1982). Mohr further reviews a number of observational studies that challenge the arbitrariness of meaning formation by illustrating clear structural factors that underlie category formation. Tilly (1997), for instance, contends that the arguments made by people in the British Parliament (from 1758 to 1834) was a direct result of their membership in one of 64 social categories (e.g. farmer, constables, militia, and practitioners). Writing in the 1990s, Mohr concludes his essay with laudatory comments on a promising new method for measuring meaning structures in formal sociology – network analysis. Since the

writing of this article, network science has inspired a paradigm shift in formalist social science, with direct implications for the problem of cross-cultural category convergence (Centola 2010, 2011, 2015, 2018).

Building on foundational theories of social networks from Simmel and Blau, early developments in sociological network analysis sought to unpack how the structure of communication networks causally shaped the structure of the category systems that people develop within these networks. As a canonical example, DiMaggio's "Classification in Art" outlined a number of formal hypotheses for how the topological structure of social networks directly shaped the classifications of art that social networks produced. For instance, in proposition A-2, DiMaggio argued that "*The greater the range of social networks, the greater the level of genre differentiation.*" In other words, this proposition (along with its corollaries) was developed to provide a structural social network account for why we observe differentiation of artistic classification systems, where the diversification of social networks here is proposed as the underlying mechanism.

While still maintaining the view that classification systems are path dependent in social systems, network analysis became increasingly amenable to theories concerning the interaction of cognitive constraints and communication dynamics. In DiMaggio's (1997) essay "Culture and Cognition", he argues that much of category formation depends on the use of analogies, where people draw from past experience and cultural associations to categorize new cultural products (e.g. art forms or institutions) entering

the cultural market³. As DiMaggio emphasizes, the widespread use of analogies in category formation suggest that much of categorization in social networks need not be arbitrary in nature, because these novel categories bare systematic similarities to pre-existing domains. Nevertheless, DiMaggio still maintains that which categories form can be path dependent, because of how “networks are crucial environments for the activation of schemata, logics, and frames” (282). In other words, DiMaggio proposes that communication in social networks mediate which analogies form, where communication among people drive which analogies are used in the formation of new categories.

DiMaggio’s “Culture and Cognition” marked a critical move in sociology because, by relaxing the assumption of arbitrariness, DiMaggio sparked a broader discussion of how cognitive processes (e.g. analogy making) can constrain social construction in ways that are not path dependent. A recent extension of DiMaggio’s network approach has been Hannan et al.’s (2007) *The Logics of Organization Theory*, which argues that universal cognitive constraints are needed to explain a core phenomenon in sociology – the “illegitimacy discount” – where people prefer to invest in and consume cultural products (e.g. technologies) that they can readily categorize by analogy to familiar products (Zuckerman 1999, 2004, 2012). Hannan et al.’s work starts by assuming that category systems exist, and then it examines the social implications of how organizations categorize audiences, and how they disseminate categories to audiences through marketing. Yet, they recognize that the question of how category systems emerge through network interactions is of pressing importance. Indeed, they conclude

³ For a nice example, see Suarez and Stine’s (2015) discussion of the initial categorization of the “snowboard”, which had initially (and unsuccessfully) been named the “snurfer”, in direct analogy to surfing.

their book with a note on the need for such network analysis in future work: “we have refrained from developing a formal account of audience structure, because we think we need a better empirical foundation. Along these lines, we are interested in how audience structure might affect consensus formation in category emergence” (302).

In recent years, there have been major advancements in the use of network analysis to study the emergence of categories (and more broadly, conventions) in social networks, using formal simulations, online experiments, and observational analysis. Much of this work has been developed in computational sociology and evolutionary linguistics. As I review these developments in the following section, I uncover an unresolved paradox for how the structure of communication networks impacts the level of cross-cultural convergence in the category systems that grow in separate social groups. I find a number of studies proposing that increasing network size should increase variation among emergent category systems, while a number of other studies report the opposite intuition, that increasing social network size should increase the similarity of emergent category systems. To address this paradox, I develop a novel formal model of category formation in networks which identifies the conditions under which coordination in social networks can lead to path dependency (contrary to the nativist position), and the conditions under which it can lead to cross-cultural convergence (contrary to both the nativist position and the constructivist position). I detail this model in chapter 2.

1.3. A Paradox in the Effects of Network Size on Category Formation

The core intuition behind social constructivism is that increasing the diversity of possible categories in a population increases the number of possible trajectories in category formation, leading to path dependent and cross-group variation (Smith 2010; Burr 2015). Following this intuition, a number of formal models have been proposed which claim that, if agents have no preferences for which of a set of categories (or, in some models, cultural traits) to adopt, then peer influence in coordination networks leads to path dependent (DellaPosta, Shi, and Macy 2015; Flache and Macy 2006, 2011; Shaw 2015; see van de Rijt 2019 for further review). These models are consistent with numerous experiments in cognitive science illustrating that separate dyads arrive at strikingly different descriptions of the same stimuli when engaging in coordination tasks involving linguistic reference (Brennan and Clark 1996; Clark and Wilkes-Gibbs 1986; Fay et al. 2010; Galantucci and Garrod 2011). A recent study conducted a visual communication task in social networks of 8 people, and found that increasing network size increased variation among the visual categories produced by separate networks (Fay et al. 2018). The same intuition has been reiterated in large-scale analyses of language change, where increasing population size is said to increase the rate at which new words enter a language, and thereby increase variation among different languages (Bowerman 2010; Bromham et al. 2015; Keller 2005). Indeed, a recent issue of the *Philosophical Transactions of the Royal Society* is dedicated to the topic of how social network complexity begets communicative complexity in both animals and people (Freeberg, Ord, and Dunbar 2012). Common across all these explorations is the claim that increasing social network size increases variation in the category systems developed in

different social groups, as a result of increasing the diversity of potential options competing in the social system.

This view stands in direct contradiction to a number of studies reporting the opposite effects of population size on cultural evolution. Following a fundamental intuition from complex systems – that large populations can exhibit qualitatively different dynamics (Anderson 1972) – a number of formal models have found that communication in large populations can cause sharp transitions in collective dynamics that radically compress a diversity of competing labels into a small, finite vocabulary (Baronchelli et al. 2006, 2010; Gong et al. 2012; Puglisi, Baronchelli, and Loreto 2008; Steels and Belpaeme 2005). A recent online experiment tested this theory in a real-time communication game (“the Name Game”), where players coordinated to establish a name for a person’s face in networks of varying topologies (Centola and Baronchelli 2015). While the naming game does not capture a process of category formation, because it does not involve the use of labels to group multiple stimuli together, it presents the simplest case for how network interactions can consistently compress a diverse space of competing elements. This study found that in clustered networks where peers had a small number of local connections, diversity among names in the population was preserved; however, they found that in the condition where the size of subjects’ peer neighborhoods was increased to form a fully-connected network, separate populations reliably converged on a single globally adopted name.

While the above evidence suggests that increasing network size can cause separate populations to reliably compress a diverse space of competing options, it does not speak directly to similarities in the content of category systems that separate populations converge on. The formal models simulating category formation in social

networks assume, by design, that the labels proposed by agents are inherently arbitrary and bare no direct relation to the simulated stimuli, such that no label could be more likely to be used than others as a product of analogy making or common cultural background. The trivial result is that while separate populations can reliably converge on finite vocabularies, there is no process to guide their convergence dynamics to a common attractor state, leading to random path dependent. By consequence, it was not possible for these original models to produce cross-cultural category convergence, indicating that a key factor was missing.

In more detail, another way in which large populations can qualitatively differ in their dynamics is in the predictability of the convergence state of a system. A canonical example is the Fisher-Wright model of genetic drift, which examines how a population of genes will evolve when a subset of a population has been randomly segmented and forced to evolve on its own (Gould 2002; Kauffman 1993); the model finds that if certain alleles are more represented in the population than others, then as the size of the segmented subpopulation increases, the predictability in the evolutionary trajectory of the subpopulation also increases. But if the subpopulation is small, it is more likely that its chance combination of alleles will not include those that were initially most representative, leading to path dependent evolutionary trajectories based on the infrequent allele combinations that happened to be distributed in the subpopulation. In this model, alleles are assumed to be differentiated by fitness solely in terms of their baseline frequency, but its results are consistent with a definition of allele fitness based on adaptive properties (Kauffman 1993).

The general logic of this evolutionary model has been widely applied in the study of language (Atkinson et al. 2008; Mufwene 2001; Newberry et al. 2017; Pagel et al.

2019). The “iterative learning” paradigm, which examines language change as a result of transmission between generations, has found that communication overtime can amplify weak biases for particular grammatical structures in a population, leading separate populations to arrive at similar grammars (Kirby, Cornish, and Smith 2008; Kirby et al. 2008; Kirby, Dowman, and Griffiths 2007). Observational data suggests these results concerning grammar may apply to the formation of additional linguistic features, including phonetics and even vocabulary (Newberry et al. 2017; Pagel, Atkinson, and Meade 2007; Silvey, Kirby, and Smith 2015). Several studies have found that frequent communication in large populations can incentivize the global adoption of simple terms that are easy to use and learn, in the place of complex and polysemous vocabularies (Kemp and Regier 2012; Regier, Carstensen, and Kemp 2016; Regier et al. 2007; Xu and Regier 2014). Indeed, a recent observational study found a strong correlation between the popularity of a label and the likelihood of this label gaining future adoption in a population, hinting at the possibility that population dynamics amplified their likelihood of spreading (Pagel et al. 2019). Even in the anthropological data on cross-cultural category convergence, population size is a significant positive predictor of similarities in the size, structure, and content of category systems (Brown 1984; Fay and Ellison 2013; Fay, Garrod, and Roberts 2008; Nettle 2012; Witkowski and Burris 1981). Further consistent with the Fisher-Wright model, it has been found that terminologies used in small populations are subject to higher levels of idiosyncrasy and variation (Boone 1949; Bower 2010; Nettle 1999, 2012; Pagel et al. 2007). Altogether, these studies point to the prediction that communication in small social groups can, contrary to the nativist position, lead to highly divergent, path dependent category systems for novel continuous domains, whereas large social groups can generate predictable evolutionary trajectories, leading to similar category systems in terms of both size and content.

This hypothesis, if true, has ramifications for any domain of inquiry that uses categories –and potentially all of them. Until only recently, it has been nearly impossible to test this hypothesis experimentally. The experimental study of category formation in cognitive science remains largely focused on individual-level tasks. The select few studies that examine social dynamics in category formation have focused on dyads for observing communication effects (Brennan and Clark 1996; Clark and Wilkes-Gibbs 1986; Galantucci and Garrod 2011). But for sociologists and a number of evolutionary linguists, studying communication effects requires studying macro-level dynamics in populations much larger than dyads.

Scarce few experiments have attempted to scale category tasks beyond dyads. One study, discussed above, examined a visual communication task in networks of 8 people, but found that communication lead to path dependence; meanwhile, a network of 8 people is not sufficiently large to observe the expected effects of large population sizes on categorization (Fay et al. 2018). A related experiment tested whether the size of a communication network affected how people coordinate through a virtual maze game, and found that in larger social networks ($N=10$), separate groups were much more likely to adopt similar verbal strategies for winning the game (Garrod and Doherty 1994). An experimental extension of the naming game tested whether a committed minority of confederates using the same name in fairly large social networks (e.g. $N > 20$ people) could trigger global adoption, consistent with the claim that communication networks amplify labels that are more representative in a population at baseline (Centola et al. 2018). This finding was supported in both their model and experiment. At present, this result remains merely suggestive in the context of cross-cultural category convergence,

since the Name Game does not capture a category formation process, which involves using labels to group together stimuli from a common continuum.

In what follows, I report the results of a novel formal model designed specifically to test the effects of population size on the emergence of cross-cultural category convergence. The model is an extension of a recent formal model designed to answer a question that is highly relevant to the puzzle of cross-cultural category convergence, and intellectually precedes it: i.e. how it is possible for a given population to establish a finite vocabulary for a novel continuous domain (e.g. colour) when there may be an infinite number of possible partitions and an infinite number of possible labels (Baronchelli et al. 2006, 2010; Puglisi et al. 2008)? Nativists sought to resolve this question by suggesting that individuals, at baseline, simply do not vary enough to create an infinitely large space of competing categories. This model challenges these results by illustrating that communication in large populations alone is sufficient to trigger sharp transitions from a vast diversity of competing elements in the population, to the population holding a finite vocabulary. By tweaking a key assumption of the model – namely, the arbitrariness of labels – I show that this model is consistent with the hypothesized effects of population size on cross-cultural category convergence, where small populations are more likely to vary substantially in the final category systems they adopt, while increasing diversity in large populations increases similarity in the category systems that emerge across separate social groups. In subsequent chapters, I report the results of large-scale online experiments that validate these predictions, and I develop a plan for additional applied extensions to illustrate how these predictions have major practical import to a domain of central importance in communication studies today – i.e. content moderation over social media.

CHAPTER 2: A MATHEMATICAL MODEL OF COLLECTIVE CATEGORY FORMATION

The formal model on which this analysis is based is referred to as the “Categories Game”. It is a more complex extension of the “Name Game”, in which agents interact sequentially in dyads drawn each round from their network neighborhood in a larger connected graph. The interaction logic of the Name Game represents a simple repeated coordination game, where each agent attempts to use a name for a single referent (e.g. a person’s face), and they are incentivized in each dyadic interaction to use the same name as their partner for that round. Then, in subsequent rounds, their next partner is sampled from their network neighborhood.

The Categories Game involves several key extensions to the Name Game. First, in the Name Game, each pair of agents each round offer a name simultaneously and evaluate whether it agrees, such that the roles of agents are equivalent. However, the Categories Game recreates a paradigmatic scenario in language interaction (first specified by Wittgenstein⁴), where a speaker uses a word to direct the actions of a hearer, and if the hearer demonstrates the correct action, coordination is successful. For this reason, for each round in the categories model, one agent is randomly assigned to

⁴ Wittgenstein (1965) lays out simple scenario of linguistic interaction that this game is designed to capture, while also reflecting on the definition of meaning it seeks to demonstrate in the “The Blue Book”: “Suppose I give to an Englishman the ostensive definition “this is what the Germans call ‘Buch’”. Then in the great majority of cases at any rate, the English word “book” will come into the Englishman’s mind. We may say he has interpreted “Buch” to mean “book”. The case will be different if e.g. we point to a thing which he has never seen before and say: “This is a banjo”. Possibly the word “guitar” will then come into his mind, possibly no word at all but the image of a similar instrument, possibly nothing at all. Supposing then that I give him the order “now pick a banjo from amongst these things.” If he picks what we call a ‘banjo’ we might say ‘he has given the word ‘banjo’ the correct interpretation”; if he picks some other other instrument – “he has interpreted ‘banjo’ to mean ‘string instrument’” (1965: 2).

be speaker, and the other is randomly assigned to be hearer. Both the speaker and the hearer are presented with an array of objects; the speaker attempts to get the hearer to select one of the objects by sending a label. The hearer checks whether it has a meaning for the label stored in its memory, and if so, it selects the object associated with the label. Following Wittgenstein's logic, if the hearer selects the correct object, we can say the speaker and hearer have the same meaning for the label used (i.e. the same mapping between the word and the group of objects to which it properly refers to in the world).

Another way that the categories model extends the Name Game is by introducing the interaction dynamics of using a label to group objects together from a continuum; whereas in the Name Game, agents negotiate to name a single fixed referent. The categories model involves a population of N individuals (or players), committed in the categorization of a single continuous perceptual channel, each stimulus being represented as a real-valued number ranging in the interval $[0, 1]$, where the range $[0, 1]$ creates a fully continuous novel dimension. Here, we identify categorization as a partition of the interval $[0, 1)$ in discrete subintervals, from now onwards denoted as "perceptual categories". This numeric continuum is standardly used in the literature as a close analogy for the color spectrum, which is highly continuous; but various studies discuss at length how this model of a continuum is indeed intended to generalize beyond the color spectrum, and even into multidimensional perceptual channels (e.g. the set of objects used to contain liquid, where there is no natural discontinuity between cups and glasses). The idea is that this model, even while reducing the phenomenon to the case of a one-dimensional continuum, unveils a mechanism that can be easily extended to any kind of space, once it has been provided with a topology.

For the negotiation dynamics in the model, each individual agent is initialized with a dynamic inventory of form-meaning associations linking their perceptual categories (the partitions of the continuum) to labels, where the “meaning” of the label is how it maps onto a subset of perceptual categories in the continuum. Perceptual categories and the labels associated to them co-evolve dynamically through a sequence of elementary communication interactions, simply referred to as games. All players are initialized with only the trivial perceptual category $[0,1]$, with no name associated to it. At each time step, a pair of individuals (one playing as speaker and the other as hearer) is selected and presented with a new “scene”: i.e., a set of $M \geq 2$ objects (stimuli), denoted as $O_i \in [0, 1)$ with $i \in [1, M]$.

The speaker discriminates the scene, if necessary adding new category boundaries to isolate the topic; then she names one object and the hearer tries to guess it. The label to name the object is chosen by the speaker among those associated to the category containing the object, with a preference for the one that has been successfully used in the most recent game involving that category. A game round is successful if the hearer makes the correct guess. Based on the game’s outcomes, individuals may update their category boundaries and the inventory of the associated words: in a successful game, both players erase competing words in the category containing the topic, keeping only the word used in that game; in failed games, the speaker points out the topic and the hearer proceeds to discriminate it, if necessary, and then adds the spoken label to its inventory for that category. A detailed overview of these dynamics is present in fig. 1 (taken from the original modeling paper).

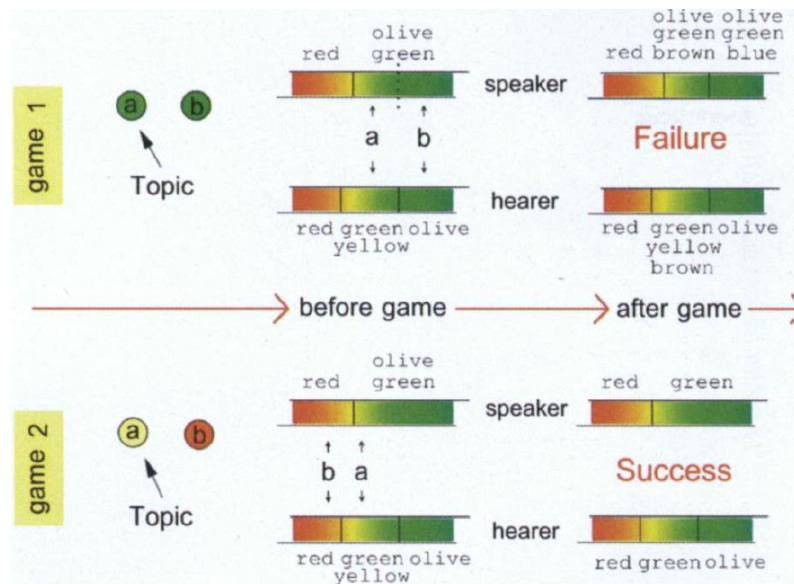


Figure 3. A schematic display of the coordination logic of the categories model (copied from Puglisi et al. 2008). The following paragraph walks through the details of this logic verbally.

The behavioral rules for discrimination and communication are displayed in fig. 3, which present a case of both failed coordination (game 1) and successful coordination (game 2). In each round of the game, two players are selected to interact, conditional on them sharing an edge within a social network. One player is randomly designated as speaker, the other as hearer. A set of objects are presented to both players. The speaker selects the topic. In game 1, the speaker has to discriminate the chosen topic (“a”) by creating a new boundary in his rightmost perceptual category at the position $(a + b)/2$. The two new categories inherit the words inventory of the parent perceptual category (here, the words “green” and “olive”) along with a different brand new word each (“brown” and “blue”). Then, the speaker browses the list of words associated to the

perceptual category containing the topic. There are two possibilities: if a previous successful communication has occurred with this perceptual category, the last winning word is chosen; otherwise, the last created word is selected. In the present example, the speaker chooses the word “brown” and transmits it to the hearer. The outcome of the game is a failure because the hearer does not have the word “brown” in her inventory. The speaker unveils the topic (e.g. by pointing at it), and the hearer adds the new word to the word inventory of the corresponding perceptual category. In game 2, the speaker chooses the topic “a,” finds the topic already discriminated, and refers to it using the label “green” (which, e.g., may be the winning word in the last successful communication concerning that category). The hearer knows this word and refers to the correctly to the object that represents the speaker’s intended topic. In the case of a success, both the speaker and the hearer eliminate all competing words for the perceptual category containing the topic, leaving “green” only (on the assumption that this is now the accepted word for this percept). In general, when ambiguities are present (e.g. the hearer finds the label associated to more than one category containing a possible object), these are resolved by making an unbiased random choice.

An additional parameter controls the perceptual resolution power of the individuals limits their ability to distinguish objects/stimuli that are too close to each other in the perceptual space: to take this into account, the model defines a threshold d_{min} inversely proportional to their resolution power. In a given scene, the M stimuli are chosen to be at a distance larger than this threshold: i.e., $|o_i - o_j| > d_{min}$. The results vary by d_{min} , though not qualitatively, and altering the size of M does not alter the qualitative outcomes of the model, but only the amount of time it takes for convergence.

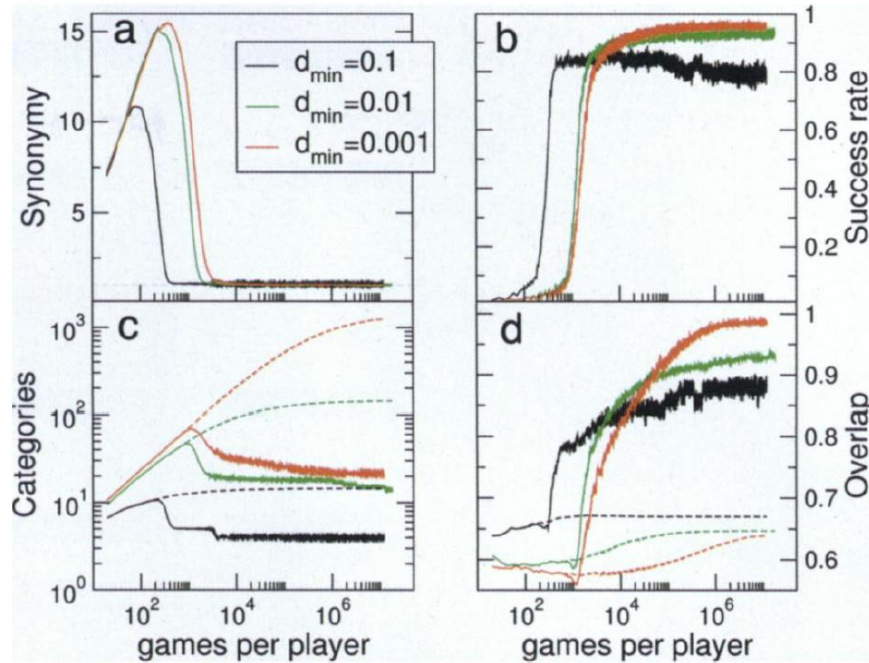


Figure 4. The main results of the original categories model (Puglisi et al. 2008). Results are from simulations with 100 agents and different values of d_{min} . (a) Synonymy, i.e. average number of words per perceptual category. (b) Success rate measured as the fraction of successful rounds, indicating a sharp transition. (c) average number of perceptual (dashed lines) and linguistic (solid lines) categories per individual, showing an initial increase, and then a sharp decrease in the number of categories per individual, indicating compression effects. (d) Averaged overlap (i.e. alignment) among speakers within each network, for both perceptual and linguistic categories, showing sharp increases in the adoption of a shared verbal vocabulary.

The main results of the original categories model are displayed in fig. 4 (copied from the original paper by Puglisi et al. 2008 with the authors permission). Panel A shows that the average number of words in the population for each perceptual category (i.e. the average linguistic ambiguity) sharply decreases overtime as a function of repeated communicative interactions. Panel B shows a similar sharp transition in terms of success rate, measured at the fraction of successful rounds, where after each agent has played roughly 100 games, the system goes from nearly no successful interactions to almost perfect success of coordination (akin to the experimental results reported for the Name Game model; Centola & Baronchelli, 2015). Panel C shows the averaged number of categories (both perceptual and linguistic) per individual, showing a sharp decrease in the number of categories for both individual, indicating the effects of network coordination in compressing the diverse space of many competing labels. Lastly, Panel D shows the averaged overlap (i.e. alignment) among speakers within each network (i.e. not between separate networks), for both perceptual and linguistic categories, showing a rapid increase in the adoption of a shared verbal vocabulary.

2.1. Extending the Categories Model

A limiting assumption in the original model is that the labels proposed by agents are, by design, arbitrary, since they are defined as random strings. As a result, each new label proposed by an agent in the model bares no prior relation to the simulated continuum or to the labels already in use in the population. The result is that, when agents introduce new labels, it is not possible for some labels to be introduced at greater frequency than others, as a product of analogy-making or common cultural background. The trivial

consequence is that while separate populations in the model show similar compression effects, by converging on finite vocabularies that are shared among agents in the same network, there is no process to guide their convergence dynamics to a common attractor state, leading to totally random path dependence. As such, the model is trivially consistent with the strong social constructivism account, in showing that the category systems generated by separate realizations of the model are incapable of bearing systematic similarities, unless purely by chance.

In later iterations, researchers began to consider how the categories model can speak to the emergence of universal patterns of category structure in separate populations (Baronchelli et al. 2010). Because the original model was framed with respect to the classical case of colour terms, with an arbitrarily continuous wave spectrum, the question of universality they focused on concerned similarities in the structure of color vocabularies across populations. While a straightforward extension, the topic of universal similarities in color vocabularies has two idiosyncrasies: (1) the patterns of universality often discussed concern the rate at which new color terms are added to a population, where a key predictor of the distinctions made by a color vocabulary is the size of the vocabulary, and (2) color is uniquely physiologically constrained by a highly modular neural architecture that is conserved across all humans. For these reasons, the underlying drivers for universality in color terms are distinct from those that are expected to underlie emergent universality in domains such as emotions, fauna, or kinship.

Yet, due to its focus on color categories, later iterations of the categories model approached the problem of cross-cultural convergence in terms of how many words a color vocabulary has, and when new terms are added to these vocabularies (Baronchelli

et al. 2010). Further fixated on the technicalities of color perception, this extension of the model endeavored to account for these universal patterns by building in fixed physiological biases into agents, so that they perceive the color spectrum in terms of a rugged landscape, where some regions are easier to categorize and more likely to be grouped together than others. The result is broadly relevant to the study of cross-cultural convergence, but it has to be taken with a grain a salt to due to the focus on color. In particular, they find that the color vocabularies which emerged in separate realizations of the model became increasingly similar when agents shared strong and similar physiological biases. In other words, the model forces convergence by assuming a nativist viewpoint, with category similarities emerging as a result of shared individual neurobiology. While this result is broadly relevant to the claim that networks serve to amplify populations biases, the kind of ‘bias’ used in this model does not generalize to the kinds of biases expected to operate across the vast range of domains with continuous stimuli. E.g. similarities in how plants are categorized cannot be accounted for by shared physiological biases for how to perceive plants, because the existence of an innate neurobiological module for perceiving plants is far from plausible⁵. Furthermore, these extensions do not directly engage the effects of population size,

⁵ Though, to be noted, a branch of nativism rightly referred to as ‘radical concept nativism’ does maintain that all categories, including categories for complex objects like “plants” and human-made objects like “door-knobs” are innate (Laurence and Margolis 2002; Margolis and Laurence 2013). The logic of this viewpoint rests on a subtle argument that denies the ability to learn these concepts from experience. This view was more strongly made by Fodor (1998), who argues (to overly simplify) that to experience any category at all, including complex categories, we need to first have a mental representation that precedes the experience and directs our attention and processing of stimuli relevant to the concept. Recent extensions of this view argue that children are preloaded with massive libraries of categories (and potentially all categories) when they are born. It is beyond the scope of this dissertation to review this literature (Huang and Snedeker 2009; Medina et al. 2011).

because these hardwired physiological biases are strong enough to eventually lead to the same convergence state across networks of any size, in the limit of infinite rounds.

The question of cross-cultural convergence requires a different operationalization of population bias than hardwired, neurobiological universals, due to the evidence that individuals vary substantially in how they perceive novel continuum, with no clear evidence of underlying, cognitive universals. This is due, in part, to the fact that the continua of interest in this study (for which cross-cultural convergence has been observed) cannot plausibly be accounted for by physiological universals. It is worth noting, though, that the critique of universal physiological constraints can also be marshalled against the theory of cross-cultural convergence in the case of color perception. Ethnographers asked various members of the Hadza (an indigenous ethnic group in north-central Tanzania) to group together and label tiles from a color spectrum, and they found striking variation and disagreement among the Hadza, even though prior linguistic data suggested a universally held color system across the tribes (Lindsey et al. 2016).

Instead, the constraints of central interest in the present examination of cross-cultural convergence concern the probability of certain labels being introduced into the population, as a function of analogy-making and commonalities in the similarities agents perceive (which need not be a function of nativist universals, but which could easily be a product of shared cultural conditioning). Consistent with the Fisher-Wright model and recent experimental results on critical mass dynamics in the Name Game, the hypothesis is that if certain labels are more likely to be introduced than others by independent individuals, then increasing population size will amplify the spread of these more popular labels, leading to greater cross-cultural convergence across separate large

populations; connectedly, the hypothesis is that, similar to the effects of sampling in genetic drift, small population sizes are more likely to get trapped in communicating with idiosyncratic labels that get locally reinforced and adopted, setting these small groups on path dependent trajectories, thereby giving rise to lower overall levels of cross-cultural convergence.

To test this hypothesis, I extended the categories model by constraining the set of possible labels agents can introduce into the game, and by altering the uniformity of the probability distribution governing the likelihood that certain labels from this constrained set are introduced. These parameter updates are intended to create the dynamics described in DiMaggio's discussion of analogy-making, and also in the broader literature on the illegitimacy discount, where it is expected that when individuals are confronted a genuinely novel domain of stimuli, they will propose categories that are drawn in reference to objects and domains they are already familiar with, such that the set of labels they are drawing from is not functionally infinite. More generally, this captures the dynamics of category formation when an agent, and their social network, is situated in a culture with already existing category systems. We will see in later discussions of the experimental results that not only that this assumption is key for studying people (where subjects already possess a multitude of categories), but that it is also empirically valid with respect to how people naturally prefer to form categories for novel continua – i.e. by analogy to existing domains, rather than using novel and arbitrary names.

In my extension, I add a global set of possible labels L defined as a sequence of real values on the range $[0, |L|]$ from which all agents in a network draw from when introducing new labels when discriminating ambiguous objects (i.e. when agents do not

have an existing category for an object in the scene, and they must differentiate it from other objects by creating a label; for speakers, the label they create is then sent to the hearer to engage in coordination, at which point it may be adopted and internalized by the hearer in the case of a success). Following prior empirical data concerning the likelihood of terms being used in a population, I define a probability distribution over L , denoted as L_B where B refers to population bias. In my simulations I use the Zipf distribution (Adamic and Huberman 2002), which drives from Zipf's effort to determine the frequency at which n^{th} most common word in a language is used. The Zipfian distribution is defined by Zipf's law, which states that the size of the r^{th} largest occurrence of the event (in this case, the frequency of a word) is inversely proportional to its rank: $y \sim r^{-b}$, with $-b$ close to unity.

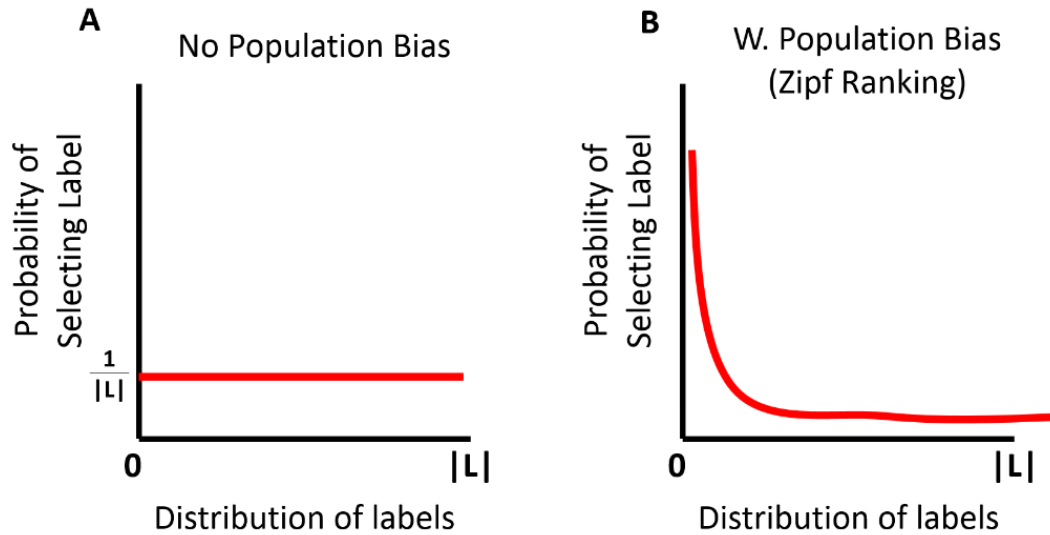


Figure 5. A schematic of how different probability distributions are defined over the set of possible labels (L) from which agents sample the new labels they introduce. (A) the case of no bias that leads to radical path

dependence, where the probability function defined over L is uniform; (B) the case of Zipfian bias, where a small subset of labels are exponentially more likely to be sampled when agents introduce new labels into the population.

These formal additions allow me to control the amount of sampling bias among individuals in the population when introducing labels – in other words, the likelihood that a certain number of labels are likely to be introduced by separate individuals, independently of each other. Fig. 4A is a general qualitative depiction case where there is no sampling bias giving rise to unequal population densities of labels, such that each label is equally likely to be introduced as others. This is the case of the original model that leads to path dependence regardless of population size (to be shown below). Fig. 4B is a general qualitative depiction of the case where the likelihood of labels being introduced is constrained by a Zipfian distribution; in this case, a very small subset is exponentially more likely to be introduced than others (for example, by 20% of the population), whereas the vast majority are highly idiosyncratic and unlikely to be introduced by independent agents in the network, creating the diversity of competing elements of key interest to the social constructivist account.

By defining L , I test the hypothesis that (1) in the case of no bias defined over L , network size will have no effect on the similarity of category systems that emerge in separate populations, and (2) that in the case with minimal population bias, increasing network size will amplify the spread and adoption of the same labels across separate populations, thereby increasing cross-cultural convergence. Furthermore, these simulations allow me to identify nonlinearities in the effects of population size, where

after a certain threshold, cross-cultural convergence increases in a nonlinear fashion. As an initial test, I measure cross-cultural convergence in terms of the average pairwise Jaccard distance between the vocabularies that emerged in separate populations of the same size. When computing the Jaccard distance between two trials, this measure refers to the number of unique words that occur in both trials, divided by the sum total of all unique words that occurred in either trial. These equally hold in terms of how the vocabularies in separate populations grouped elements from the continuum together (i.e. in terms of the overlap of partitions). All networks are initialized as fully-connected networks. Later chapters in this dissertation will examine the effects of different network topologies on convergence dynamics in category systems.

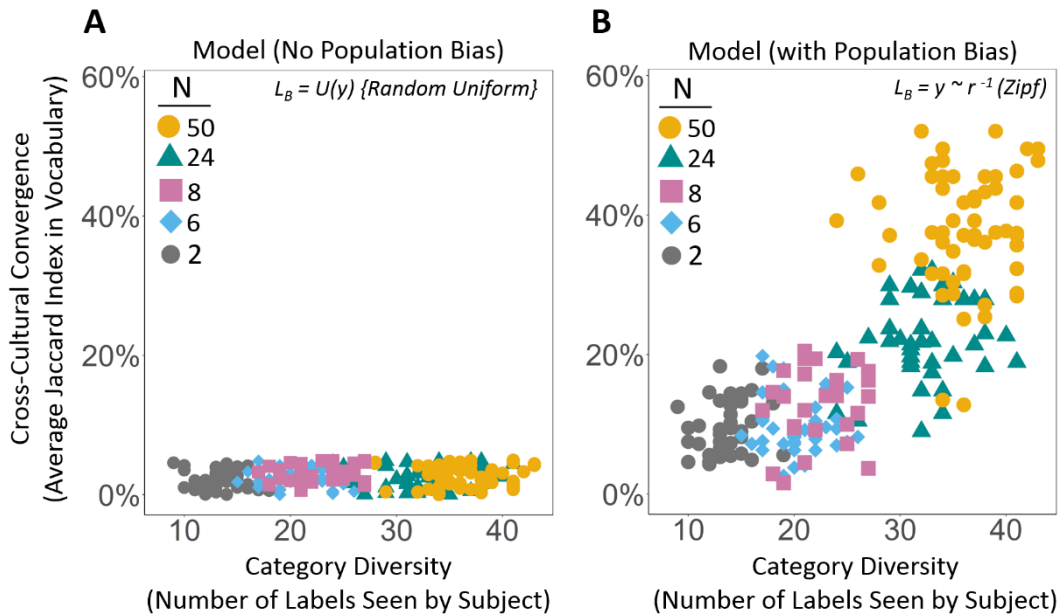


Figure 6. Results of 50 simulations (100 rounds; $d_{min} = 0.01$; $|L| = 6000$) comparing the effects of population size on bias and cross-cultural convergence. Each data point represents a single network in each condition, where the

horizontal axis indicates the level of category diversity in a population (i.e. the average number of unique labels that agents encountered in that population). (A) Cross-cultural convergence in vocabulary after 100 rounds in networks of varying population size where the model is initialized with L_B (the level of population sampling bias) is random and uniform, indicating no bias. (B) Cross-cultural convergence in vocabulary after 100 rounds in networks of varying population sizes where the model is initialized with L_B defined by the standard Zipfian distribution (i.e. where $y \sim r^{-b}$, and b approximates unity).

The modeling results provide strong support for the hypothesis that when L is defined with no population bias (fig. 5A), there is no cross-cultural convergence above chance in any network sizes, with no effect of network size on path dependence. By contrast, the model shows that when L is defined by minimal population bias using the standard Zipfian ($y \sim r^{-b}$), population size has a direct effect on the amount of cross-cultural convergence between separate networks of the same population size. Specifically, the model reveals the counterintuitive result that by increasing population size, and thereby increasing the diversity of competing elements, the predictability of the convergence state increases, rather than decreases, contrary to the popular story of social constructivism, where social interaction is assumed to inherently lead to path dependence. In this case, the number of elements in L is 6000. These results are robust to a wide range of sizes for L , where the lower bound for sufficient variation is approximately 50. We can also show that these results are robust to a range of values for d_{min} and M .

Here I posit that the key mechanism captured by the above agent-based model concerns the effect that population size has on the ability for labels to reach ‘critical mass’ and thereby spread (Centola et al. 2018). This intuition can be clearly articulated through a surprising connection to the birthday paradox in probability theory. First, following prior work, a label is said to reach a sufficiently large ‘critical mass’ when a large enough minority of subjects exist the population who are committed to spreading the label (Centola et al. 2018). Prior work on the spread of arbitrary linguistic conventions (i.e. ‘names’ in the name game) has shown that when a single name is introduced by an average of 25% of subjects in a network, it reaches a ‘tipping’ point, after which it rapidly spreads and gains widespread adoption in the overall population (though they also show that the critical mass threshold can vary according to key parameters like memory length and social resistance).

Now, the question of whether a particular label is likely to reach a critical mass within a population of size n is strikingly analogous to the infamous birthday paradox (Borja and Haigh 2007): the counterintuitive result in probability theory where the extremely unlikely event that two randomly sampled individuals share the same birthday becomes vastly more likely as the size of the sample population increases. Once a population reaches at least 23 people, it becomes more likely than chance for two people in this population to share a birthday. Once populations surpass 50 people (the size of the largest population in our experiment), it becomes almost mathematically guaranteed that two people in the population will share a birthday.

By analogy, the problem of whether a label reaches critical mass within a population is a question of the probability that a certain proportion of people in the population introduce the same label (i.e., “have the same birthday”). But the problem of

critical mass involves two subtle differences that require formalization. First, the question of whether a label reaches critical mass in a population is not about the *specific* proportion that introduces the label, but about whether the proportion of the population introducing the label reaches the *minimum requirement* to trigger a tipping point. Secondly, the distribution of label frequencies is not uniformly distributed: some labels are more likely to be introduced than others. Recent work shows that the birthday paradox holds with nonuniform distributions, though its formalization rarely incorporates this subtlety (Borja and Haigh 2007; Munford 1977). Both of these additional features of critical mass can be readily modeled using the hypergeometric distribution.

The hypergeometric distribution is derived by computing the probability of selecting k successes in a sample of size n from a total population of size N , where the total frequency of successes in the population is given by K . To determine the likelihood of obtaining *at least* k successes in a sample of size n , we sum the probabilities of obtaining $k, k + 1, \dots, k + (n - k)$ successes using the following formula.

$$\sum_k^n \frac{\binom{K}{k} \binom{N - K}{n - k}}{\binom{N}{n}} \quad (1)$$

A “success”, for our purposes, refers to selecting an individual from the population who introduces label x . The total number of possible successes in N (i.e. K) corresponds to the number of individuals who will introduce label x . Within this model, we can distinguish rare from common labels by altering the probability of an individual introducing x at baseline; in other words, by modifying K as a function of the probability $P(x)$ of an individual introducing label x . For instance, say the overall test population is

1480 subjects (the size of our experimental population, detailed in Chapter 3). If a given label is introduced by 32% of people on average, then assuming $K = N \cdot P(x)$, we arrive at $K = 296$ for this label. By contrast, if a rare label were introduced by 0.5% of the population, K would then equal 7.

The final component of this model incorporates the critical mass size of interest, denoted by cm . Following prior work, we assume here that $cm = 25\%$. To incorporate this element into equation (1), we constrain the model so that for each n , we set k as the $\min(k)$ such that $k/n \geq cm$.

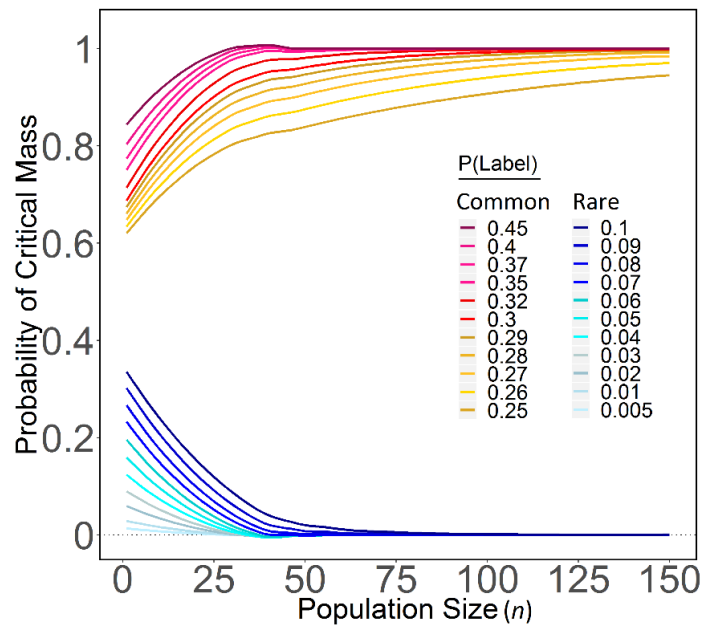


Figure 7. Using the hypergeometric distribution to model the effect of population size on the likelihood of labels reaching critical mass. Horizontal axis displays the size of a population sample (n) from the total population size N , which is set to 1460 to emulate the size of the test population in our main experiment. The vertical axis displays the

probability of a label reaching critical mass (25%). The colors indicate the probability of an individual being drawn who uses a given label. Rare labels are highlighted by cooler colors, and common labels are highlighted by warmer colors. Fig. 3B uses shaded regions to represent a smoothed display of the above distributions, distinguishing between common and rare labels accordingly.

Drawing from analytic properties of the hypergeometric distribution, we show that population size directly affects the likelihood of common labels reaching critical mass. Fig. 7 displays these results while assuming $cm = 25\%$, following prior research (Centola et al. 2018). We find that common labels with $0.25 \leq P(x) \leq 0.45$ are more likely to reach critical mass in larger populations ($n > 20$), where the probability they will reach critical mass approaches unity when $n > 50$. These results indicate that common labels are much more likely to reach critical mass and spread in larger populations, as a result of the properties of the hypergeometric distribution. These results are robust to a range of cm values. We thus arrive at the following prediction: assuming that populations possess a similar bias landscape in the likelihood of certain categories being introduced, we would expect that increasing population size drastically increases the likelihood that labels associated with greater bias (i.e. common labels) reach critical mass and spread, leading to consistent and replicable trajectories in the cultural construction of categories.

These results provide a formal, mathematical theory for the empirical hypothesis that cross-cultural convergence is the result of coordination and diffusion dynamics in social networks. My theoretical model predicts that as population size increases –

thereby increasing the diversity of category options in the population – the similarity of the convergence state of separate populations also counterintuitively increases. In the chapter to follow, I outline the design and results of a large-scale online network experiment on real-time category formation which successful confirms this hypothesis with high fidelity.

CHAPTER 3: THE EMERGENCE OF CROSS-CULTURAL CATEGORY CONVERGENCE IN AN ONLINE SOCIAL NETWORK EXPERIMENT

There are three main technical challenges that needed to be overcome in order to be able to test the hypothesis. First, we need an experimental environment that supports social interaction in large social networks. Prior experiments involving communication and categorization have been largely restricted to tasks done by individuals or dyads (Galantucci and Garrod 2011). Second, we need to construct a genuinely continuous and novel domain of stimuli such that people actually have to engage in category formation. Prior experiments on category formation have relied almost exclusively on small sets of discrete stimuli, or otherwise on drawing tasks where subjects are asked to depict tokens of already existing categories (e.g. “soccer”), thus failing to capture the process of categorizing large continuous sets of novel stimuli in natural language. This is especially challenging because the continuum constructed needs to be genuinely novel. Unlike modeling environments where agents can be constructed as blank slates, we cannot run an experiment with people categorizing a color spectrum, because people already have robust categories for colors. Lastly, we need an experiment that enables people in networks of varying sizes to collaboratively categorize this genuinely novel continuum, in real-time, and for long periods of time. We address each of these issues through the design of an online web platform that allows us to experimentally control the size of people’s social networks as they collaboratively labeled a novel continuum of arbitrary in the context of a communication game built to resemble the logic of the categories model.

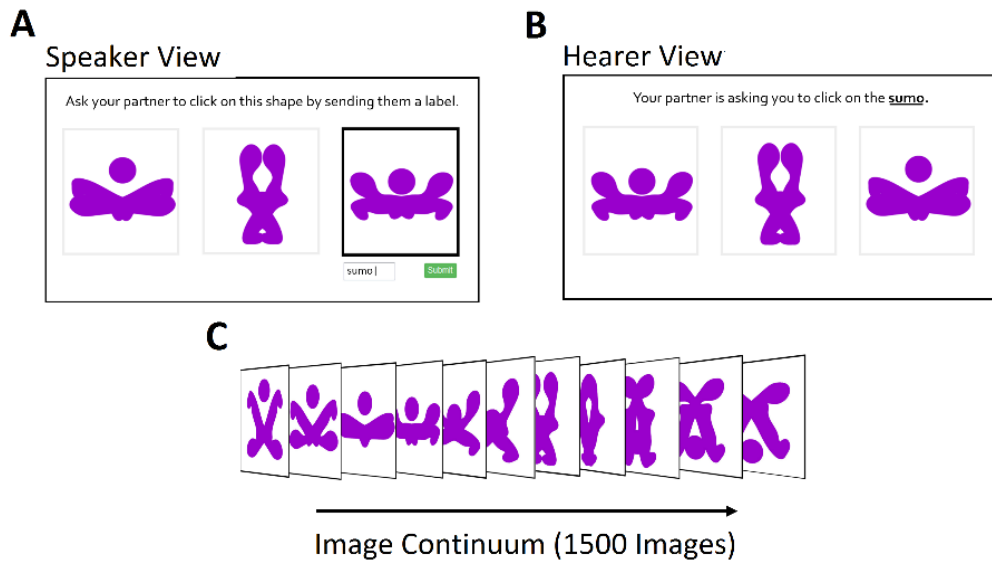


Figure 8. Screenshots of “The Grouping Game” interface from (A) the view of a speaker and (B) the view of a hearer on a given round. (C) A sample of the continuum of novel shapes used as stimuli in the experiment.

3.1. Experimental Design

1480 subjects we recruited from Amazon’s Mechanical Turk to participate in the experiment. Subjects registered to play a paid online language game called “The Grouping Game” that involved grouping abstract shapes from a novel continuum. Upon arrival to the game, subjects were randomized into either the “dyad” condition, where they collaboratively categorized the continuum with the same partner, or into one of four “social network” conditions where they categorized shapes in a fully-connected network of either 6, 8, 24, or 50 people. Each trial in each condition consisted of unique individuals. We collected 80 independent dyads and 15 independent social networks for

each network size. We also built a separate web platform where individuals independently categorized a sample of the continuum without any social interaction, and the results were highly consistent with our main findings reported below. The data was collected between July 1st and August 30th, 2018.

The image continuum was constructed using *Adobe Animate 2018* (Fig. 8). To create a continuous space of images, we began by defining 3 separate amorphous shapes on a blank white background, creating a composite image akin to Rorschach ink blots. Then, we used *Animate's* motion path functionality to create an arbitrarily large continuum of images formed by the intermediate combination of shape orientations generated by gradual rotation and motion. With this method, we were able to create a continuum at an arbitrary resolution, demarcating the amount of rotation between each frame. After pretesting, we decided to deploy a continuum consisting of 1500 images, based on participant feedback indicating that this resolution permitted an intuitive and manageable amount of variety to categorize in the timeframe of the experiment.

The gameplay was designed to recreate the Wittgenstein language game used in formal models simulating cultural processes of category formation (Baronchelli et al. 2010; Puglisi et al. 2008) (Fig. 8). In each round, subjects were randomly paired with another subject in their network. In the dyads, they were always paired with the same person. For each pair in each round, one subject was randomly chosen to be the speaker and the other hearer. Each round, the speaker was presented with a random selection of 3 images from the continuum of 1500 images (fig. 8), abiding by a minimal distance constraint of 75 images (thus approximating a d_{min} of 0.05, in terms of the logic of our formal model; see Chapter 2). The continuum was held constant across conditions.

The image selection algorithm was designed so that subjects were never shown the same shape from the continuum twice, unless (1) our system could not find a set of images that neither speaker nor hearer had seen before, or (2) the available images satisfying (1) violated the distance constraint. This design ensured that subjects were forced to categorize images from round to round by generalizing across the features observed in prior rounds. For the same label to be applied on different rounds, subjects had to group images based on perceived similarities across rounds, thus requiring a process of generalization and category formation. In each condition, subjects interacted for 100 rounds. For each round, both the speaker and the hearer were given 30 seconds to respond. Each game lasted approximately one hour regardless of condition, where in every experimental condition, each subject played 100 rounds.

To initiate categorization, one image was randomly highlighted, and the speaker was asked to enter any label into a free text-entry window that would allow their partner to click on the highlighted image (fig. 8). The order of the images displayed varied for the speaker and the hearer, so they could not coordinate on the basis of superficial strategies relating to position. The hearer received the speaker's label and was asked to click on the shape expected to be the label's referent. If the hearer clicked the correct shape, both players received 10 cents. If they failed, 1 cent was subtracted from each of their earnings, and the hearer was shown the shape originally intended as the referent to enable social learning. Following Centola & Baronchelli (2015), subjects were incentivized monetarily to emulate the positive feedback associated with successful coordination. Once the round was complete, each player returned to a waiting page where they were asked to wait while the system paired them with their next partner. Regardless of condition, subjects received the same instructions and messages on each

page, and subjects were given no information about their partners or the size of the network they were in. For this reason, any differences in the category systems that emerged across conditions can be attributed to the effects of network size.

All 1480 subjects from Amazon's Mechanical Turk were required to live in the U.S. with English as their first language. There were no differences in the distribution of demographic traits across conditions, in terms of gender ($p = 0.56$), ethnicity ($p = 0.42$), and age ($p = 0.67$), (Kruskal-Wallis H Test). We imposed these constraints on recruitment to arrange the optimal conditions to observe individual-level uniformities in how subjects labeled the continuum, as predicted by the nativist position. Thus, any variations in how subjects categorized the stimuli across conditions can be attributed to the size of the network they are in, and not to variations in the subject population between conditions.

Subjects were contacted via email with an invite to come play a paid online game at a scheduled time. One hour prior to the game, subjects were sent an email with instructions on how to play. They were told that (1) "In each round, you will be asked to label one of three shapes"; that (2) "Based on your label, your partner will have to click on the shape they think you're referring to"; that (3) "If your partner clicks on the correct shape, you both receive money"; and that (4) "Sometimes you will be the player who will have to click on the shape." Shortly before game times, subjects were emailed with a link to the experimental platform. Upon arrival, subjects were brought to a waiting page, where the above instructions were displayed for them again. When enough subjects arrived to the platform to fill all conditions, subjects were randomized to each condition, and a trial was initiated.

We excluded trials in the dyads where at least one subject stopped playing before the game was over, because under these conditions, the remaining subject had no one else to coordinate with. This occurred in 5 of the 80 trials in the dyad condition, resulting in 75 usable trials. Overall attrition rates were very low across all conditions, with an average of 2% of participants failing to successfully finish the game. There was no significant difference in the distribution of attrition across conditions (Kruskal-Wallis H Test, $p = 0.72$).

3.2. Methods of Analysis

To identify the labels that were adopted in each trial, we selected the top 5 labels used most frequently to successfully coordinate in a single trial of the experiment. All of the results reported below are robust to a wide range of vocabulary sizes, ranging from the top 1 to 10 most successful labels. Then, we identified the images that a label was used to refer to along the continuum, as an indication of which images it grouped together. This allowed us to compare conditions in terms of the boundaries their labels drew between image sets along the continuum, as an indication of the similarities among images lexicalized by each population. Following prior work (Baronchelli et al. 2010; Hannan et al. 2007), a category in this study refers to a mapping between a label and a set of referent images from the continuum.

These techniques provided two direct measures of cross-cultural convergence. First, given that our subjects were given a free-text entry window, enabling them to enter any type of linguistic form of their choice, we measured convergence in terms of the

similarity of the vocabularies that were adopted by the end of the experiment. We computed vocabulary overlap using the average pairwise Jaccard distance between all trials in a condition, where Jaccard distance is the number of words that occur in both trials, divided by the total number of unique words across both trials.

Secondly, to calculate similarities across trials in terms of which images were grouped together, we use Baronchelli et al.'s (2010) measure of centroid overlap. The centroid of a category is the median image in the image range that each label refers to within the continuum. The distance between centroids is the absolute number of images along the continuum between the centroids of two categories from different trials. For each category c_i in population X , Baronchelli et al.'s (2010) measure calculates the minimum centroid distance between c_i and all categories in population Y . It then takes the average across all minimum centroid distances between the two populations as a measure of overall centroid alignment.

To measure population bias in the sampling of images, for each label, we computed the proportion of subjects in each trial that introduced each label at any point in the experiment, prior to exposure to this label from another players. Then, as a window into the hypothesized effect of population size, we examined the rank correlation between the proportion of subjects who independently introduced each label, and the proportion of adopters in each trial that adopted each label. To illustrate the logic of this analysis, consider that in a dyad, the largest (and only) proportion of subjects who can independently introduce a label is 50%. As a result, even if some labels are more likely to be introduced than others in a larger population, in the context of a dyad, this bias in sampling does not allow some labels to be meaningfully distinguished from others – they are all introduced by no less and no more than half the population. However, as network

size increases, the extent to which sampling bias differentiates labels also increases. If label_i is introduced by 20% of the population in a network of 50, while most labels are independently introduced by only 1 person, this means that label_i is introduced by 10 people. If the correlation between the proportion of label originators and the proportion of label adopters increases with population size, this indicates that population size serves to amplify the spread of labels with greater population density at baseline. We show how this process is crucial to understanding the observed results concerning cross-cultural convergence.

3.3. Results

I begin this analysis by establishing that the distribution of labels introduced into the experiment follows a Zipfian curve to striking precision. Consistent with Zipf's law (Adamic and Huberman 2002), a small number of labels like "crab" and "bunny" were common, meaning they were more likely to arise independently, whereas the vast majority of labels were rare and introduced by only a few individuals. I then confirm that, indeed, increasing population size drastically increases the likelihood that common labels reach critical mass and thereby spread, giving rise to consistent trajectories in the cultural formation of categories. As a result, I show that in small network sizes ($N < 10$), social interaction lead to highly path-dependent category systems, both in terms of vocabulary and continuum partitions; however, I show that in large networks ($N > 20$), communication substantially increased cross-cultural similarities in category formation across totally separate and replicated populations, both in terms of the specific words they adopted, as well as how they used these words to group together stimuli. A novel

hypothesis follows from these results which I then directly test. A popular nativist intuition holds that categories gain popularity due to their intrinsic cognitive appeal. However, my theory and results suggest that the success of a category is largely a function of whether it is associated with a sufficiently large critical mass to trigger its diffusion. I evaluate this prediction in a robustness experiment, where I test whether artificially inflating the baseline popularity of uncommon labels can trigger cross-cultural convergence on these labels rather than on more cognitively appealing ones. I find across six replicated trials that a committed minority of confederate subjects (37% of a network of 24 people) could trigger the adoption of uncommon labels in the place of more cognitively appealing ones. Implications for cultural evolution and communication engineering are discussed in the concluding chapter.

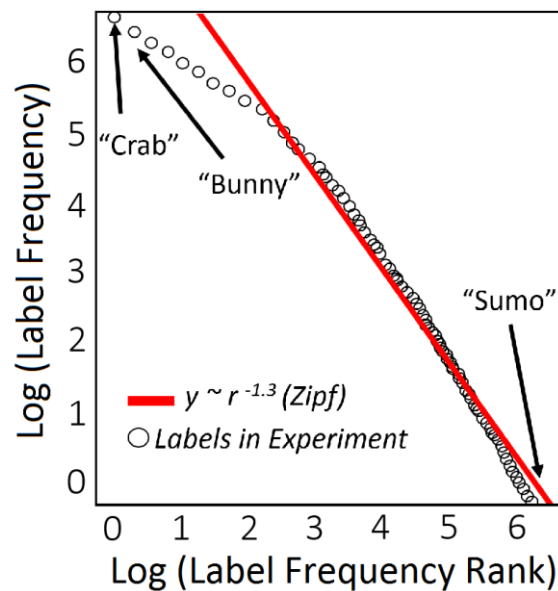


Figure 9. Using the Zipfean distribution to model the initial frequency of labels (including data from all conditions; $N=2$, $N=6$, $N=8$, $N=24$, and

$N=50$), where initial frequency refers to the number of individuals who introduced a label without any prior exposure to the label in the task. Vertical axis displays the log of each label's initial frequency. Horizontal axis displays the log of each label's frequency rank. The data represent 80 dyads and 15 social networks of each size.

As predicted, I find that the likelihood of labels being independently introduced into a given condition was characterized by a Zipfian distribution with the standard curve ($b = 1.3$), (Fig. 9). The majority of labels were “rare” and were originated by less than 1% of the population, on average, whereas a small subset (less than 1%) of labels were “common” and were originated by an average of 25% of the population. In supplementary analyses, I show that the same labels emerged as popular when separately analyzing the data from each condition (Kruskal-Wallis H test, $P=0.65$), suggesting that the initial population bias represented in each condition was indistinguishable. Supplementary analyses further show that common labels were associated with regions of the continuum that were easier to label in general, suggesting that label popularity corresponded to regions of the continuum that were less ambiguous. In total, a surprising diversity of over 5000 unique labels were attempted in the experiment, across all conditions.

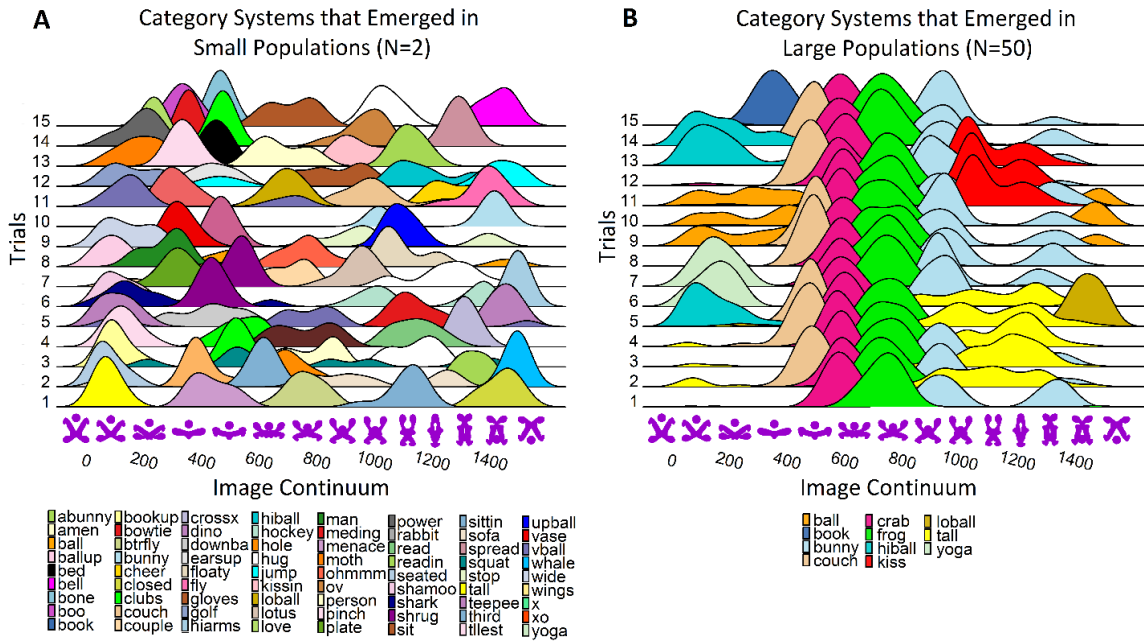


Figure 10. Comparing the level of convergence in category systems that emerged in (A) small ($N=2$) and (B) large ($N=50$) populations. Each row displays the category system constructed by a single unique population in each condition after 100 rounds of interaction. The horizontal axis displays the image continuum of shapes, consisting of 1500 slices. Density distributions display the frequency of successful coordination for each label, as well as the region of the continuum to which each label referred. Each color indicates a unique label. Similarity in the category systems across populations indicates convergence.

Figure 10 displays the category systems that emerged in each population. As predicted, fig. 10A shows that small populations ($N=2$) produced highly divergent category systems. Only 5% of labels were shared across independent dyads, and there was no consistency in how these dyads partitioned the continuum ($p < .001$, $n = 80$, Kruskal Wallis H Test). As a result, dyads varied not only with respect to the labels they

adopted for the same regions of the continuum, but also with respect to the regions of the continuum they successfully categorized. By contrast, large populations ($N=50$) generated remarkably similar vocabularies (50% Jaccard index, $n = 95$, $p < .001$, Wilcoxon rank sum) and similar partitions of the continuum ($p = 0.87$, $n = 15$, Kruskal Wallis H Test), indicating convergence in how these independent populations categorized the novel stimuli (Fig. 10B).

These findings appear puzzling at first since larger populations are expected to increase the unpredictability of category formation as a result of containing a greater diversity of individuals, and thus a greater number of categories that can be adopted. Yet, these results indicate that increasing population size – and thereby increasing the diversity of categories – can counterintuitively lead to convergent cultural trajectories across independent populations.

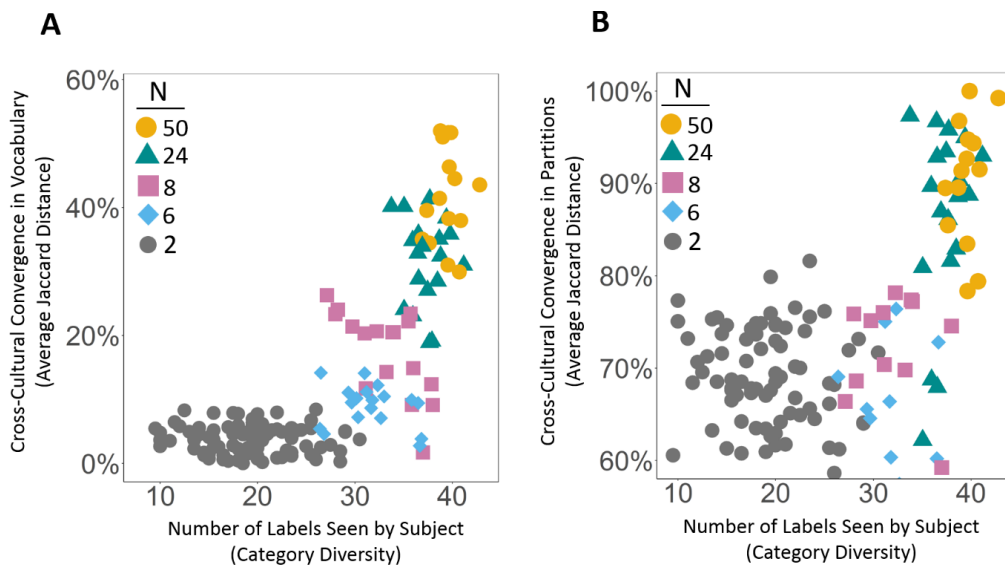


Figure 11. (A) Convergence in the vocabularies that emerged in populations of different sizes, for $N=2$ (black dots), $N=6$ (blue diamonds), $N=8$ (purple squares),

$N=24$ (green triangles), and $N=50$ (yellow circles). Vertical axis reports the average similarity in vocabulary (average Jaccard index) between each network trial and all other networks of the same population size. Horizontal axis displays category diversity, measured as the average number of unique labels encountered by subjects in a population. Data points represent experimental results (80 dyads and 15 social networks of each size). (B) Measuring cross-cultural convergence in terms of how the vocabularies of separate groups partitioned the continuum, where the y-axis displays the average overlap between the partitions in a given trial and all other trials of the same group size. The x-axis displays the average number of unique labels encountered over the course of the experiment by each subject within each group. Average partition overlap is quantified using centroid alignment. The centroid of a category is the median image in the range of the continuum to which this category referred. Average partition overlap is measured as $1 - (k_{ij} / m)$, where k_{ij} is the average minimum centroid distance between the categories of population i and j , and m is the maximum number of images that can separate two centroids (i.e. 1499).

The theoretical predictions for these convergence dynamics based on the model specified in chapter 2 provide an excellent fit with our experimental findings (Fig. 11). Across all experimental conditions, label diversity significantly increased with population size. Figure 11A shows that greater label diversity within populations predicts greater similarity in the category systems that emerge between populations (Jonckheere-Terpstra Test, $N=120$, $p < .001$). We find these convergence dynamics not just for the labels that were used, but also for how participants' partitioned the continuum into distinct regions (Figure 11B).

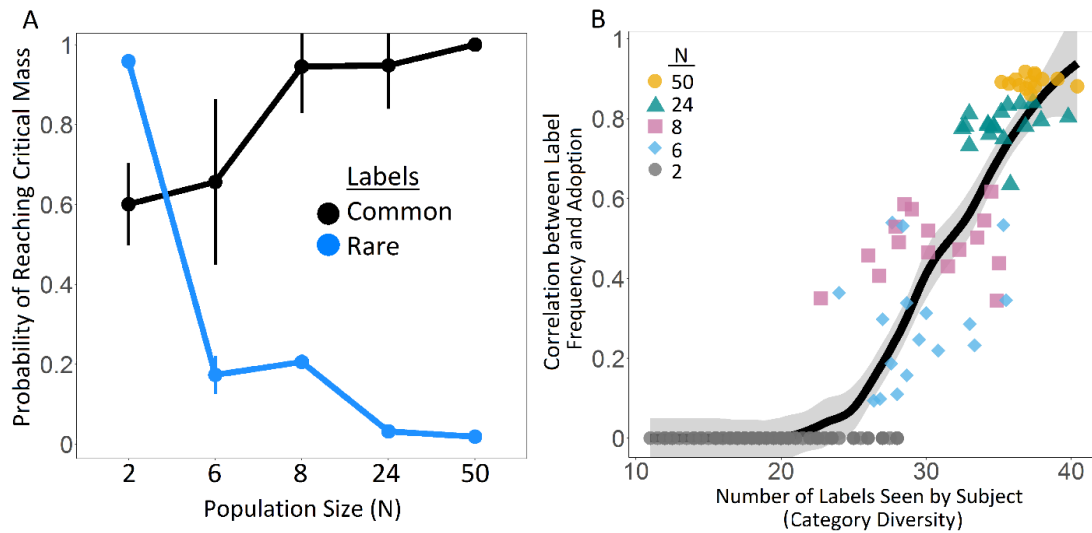


Figure 12. Larger populations amplify the spread of initially frequent labels. (A) Displaying the effect of population size on the ability for labels to reach critical mass (when at least 25% of subjects in a network independently introduce a label). Common labels are identified as outliers with high initial frequency (see Chapter 2 for model details). Data display the proportion of experimental trials in each condition for which each label type reached critical mass. Error bars display 95% confidence intervals. (B) The correlation between the initial frequency of a label in a population and the proportion of subjects in a population who adopted the label (vertical axis), where adopting a label implies that a subject produced a label after being exposed to it. Horizontal axis displays the diversity of categories in each trial, indicated as the average number of unique labels encountered by each subject in a network. All observations are independent and at the network-level. The data represent 80 dyads and 15 social networks of each size.

Here I test the simple mechanism proposed in chapter 2 as an explanation of these findings. I suggest that larger populations amplify the spread of initially more frequent labels, leading these common labels to reach a ‘tipping point’, after which they diffuse and become widely adopted. Figure 12A shows the relationship between population size and critical mass dynamics, empirically. In small populations, common labels were not sufficiently reinforced to reach the tipping point needed to trigger widespread adoption. Consequently, small populations were significantly more likely to adopt rare labels ($n = 80$, $p < .001$, Wilcoxon Signed Rank), leading these populations to follow divergent evolutionary trajectories. However, increasing population size significantly increased the likelihood that common labels (like “crab” and “bunny”) would be reinforced and adopted (Jonckheere-Terpstra Test, $n = 120$, $p < .001$), significantly reducing the likelihood that rare labels would spread (Jonckheere-Terpstra Test, $n = 120$, $p < .001$). Our findings indicate a direct relationship between population size and cross-cultural convergence (Fig. 12B). For large populations ($N=50$), the likelihood of common labels becoming widely adopted approaches unity, leading these populations to convergent cultural trajectories.

A crucial implication of our theory is that cultural convergence does not simply depend upon cognitively salient features of the labels themselves, but upon the labels’ frequency in the population. An established intuition is that certain categories gain popularity because they have intrinsic cognitive appeal (e.g., because of their ‘natural’ descriptive fit with the stimuli) (Rosch 1973; Winkielman et al. 2006). However, even when the most popular labels (e.g. “crab” and “bunny”) were attempted in dyads, they regularly failed to gain acceptance. We located the region of the continuum shared

among all uses of the common label “crab” in the $N=50$ networks (i.e. images in the range 500 – 600), and we examined the dyads in which “crab” was attempted for this region. Every time this label was introduced for this region in large networks ($N=50$), it gained adoption. By contrast, even in dyads where the label “crab” was attempted for this same region, a wide range of rare labels were adopted in the place of “crab”, including “baby”, “turtle”, “hotdog”, and “smile”. As a result, dyads were much less likely to adopt “crab” when this label was introduced compared to $N=50$ populations (Binomial Test, $p < .001$, $CI = [0.63, 0.83]$). This suggests that the adoption of these labels at the population level is not strictly determined by their cognitive appeal, but rather by the fact that they are more likely to be reinforced and reach critical mass in larger populations.

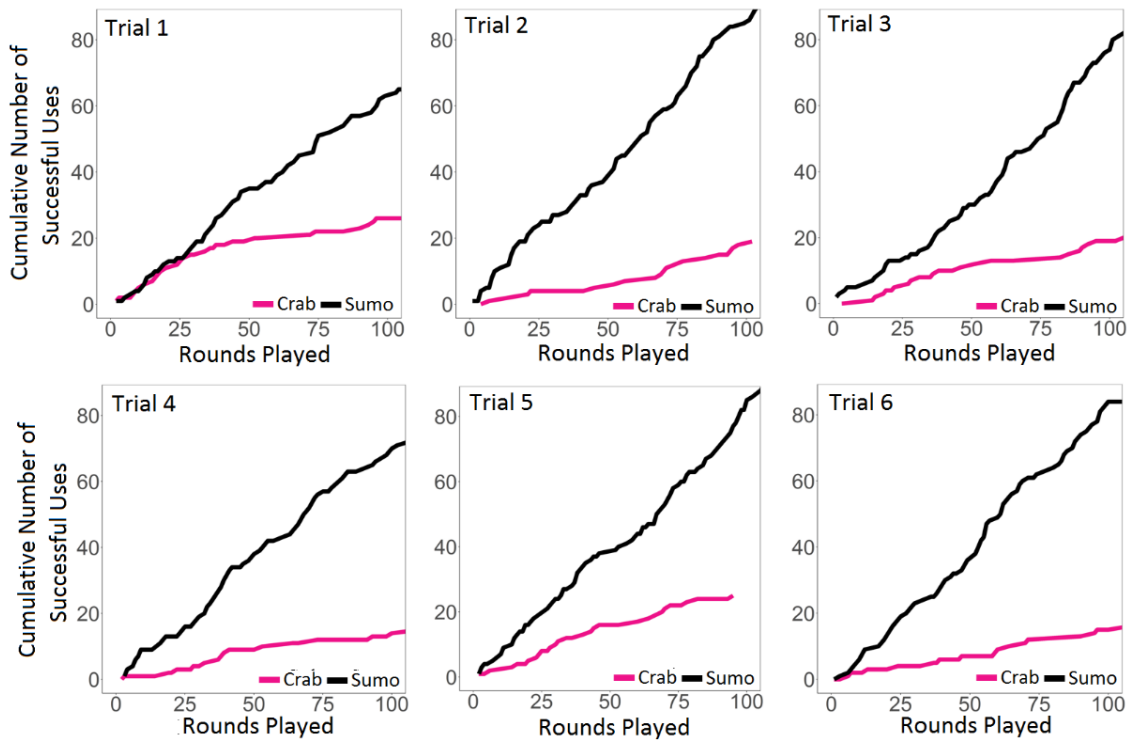


Figure 13. Time series showing the adoption of confederates’ rare label (“sumo”) by noncommitted subjects (i.e. experimental subjects). Pink lines

indicate the cumulative number of successful uses among experimental subjects of the label “crab”. Black lines indicate the cumulative number of successful uses among experimental subjects of the label “sumo”. Each round is measured as $N/2$ pairwise interactions, such that each player has one interaction per round. The data displayed exclude all interactions between confederates.

To evaluate this hypothesis directly, I experimentally tested the following counterfactual: if we artificially inflated the popularity of infrequent labels to reach critical mass, would it trigger convergence on those labels rather than on more cognitively appealing ones? I conducted six robustness trials ($N=24$) in which each network contained a minority of confederate subjects (37%) tasked with spreading a novel category system based on infrequent labels. For instance, I trained confederates to use the rare label “sumo” (Fig. 13) for the same regions of the visual continuum associated with the most popular label in our initial studies, “crab”. Figure 13 shows that although “crab” appeared in each robustness trial, “sumo” consistently outcompeted “crab”. In every robustness trial, populations adopted the confederates’ labels across each region of the continuum, yielding significantly more convergent category systems (58% Jaccard index) than those that emerged in $N=24$ populations without confederates (36% Jaccard index), ($n = 21$, $p < .001$, Wilcoxon rank sum).

3.4. Discussion

This study strongly supports the hypothesis that coordination dynamics in social networks can facilitate both the divergence and convergence of category systems, as a function of network size. Moreover, our results are highly consistent with our hypothesized mechanism, that large networks amplify population densities in which labels are originated by subjects. In smaller social groups, population densities do not allow labels to be meaningfully differentiated from others. Even when $N=8$, if a label is introduced by 20% of subjects, this only barely evaluates to 2 label originators, which places this label on roughly equal footing as idiosyncratic labels that are introduced by only one person. As a result, labels that are more popular at the population level have a substantially lower likelihood of succeeding in smaller networks, which can instead serve to locally reinforce highly idiosyncratic labels, leading to path-dependent trajectories in category formation. By contrast, in large networks (e.g. $N=50$), if a label is introduced by 20% of the population, this means that this label is introduced by 10 independent individuals, whereas the vast majority of labels are still introduced by only 1 person. As a result, these popular labels become much more likely to spread and gain adoption in separate social networks, leading to a nonlinear increase in cross-cultural convergence.

One possible objection is that given more time, dyads would approach the levels of cross-cultural convergence observed in large social networks. The concern is that many more interactions happen in parallel in the large networks, such that dyads are disadvantaged in the overall number of interactions informing their category systems. To test this, we allowed 30 dyads to play for an additional 25 rounds. We show that the average level of cross-cultural convergence in both vocabulary and image groupings

was not significantly higher than the original dyads, when comparing any number of topmost successful vocabulary items.

A final concern is that our results may be an artifact of the game algorithms used to select and display images to subjects. To address this concern, we first show that each shape in the continuum was equally likely to be shown across all trials (Kruskal-Wallis H Test, $P=0.89$). In general, there were no significant differences in the distribution of images that individuals saw across conditions (Kruskal-Wallis H Test, $P=0.48$). Secondly, we show that when subjects were shown the same image twice in social networks, no image was more likely repeat, thereby preventing algorithmic bias in the process of social reinforcement.

The “social constructivist” view of cultural evolution suggests that large communication networks contain greater individual variation, which leads to greater divergence and unpredictability in the evolution of category systems (Fay et al. 2010b; David 2007; DiMaggio 1987; Berger and Luckmann 1967; Salganik, Dodds, and Watts 2006; Macy et al. 2019). Here, we show that while increasing the size of communication networks does, in fact, significantly increase the diversity of categories that people encounter, it does not increase divergence. Rather, it increases cross-cultural convergence. Our results suggest that convergence in the categorization of novel phenomena across independent populations is significantly determined by the communication networks in which people are embedded.

These findings offer a new experimental interpretation of past observational data on cross-cultural similarities in category systems. We suggest that rather than cross-cultural convergence providing evidence of innate, universal cognitive categories

(Brown 1984, 2004; Malt 1995; Pinker 2003; Youn et al. 2016), instead it may indicate that communication in large social networks filters cognitive and lexical diversity in such a way that promotes the development of similar category systems across diverse populations.

CHAPTER 4: CONCLUSION

Is the way we categorize the world governed by innate cognitive universals, or socially constructed? This dissertation uses insights from social network dynamics to provide a unique answer to this question. Nativism, the view that innate cognitive universals account for cross-cultural category convergence, is not consistent with the creativity and diversity observed in how individuals categorize novel, continuous stimuli. Social constructivism views creativity and diversity as key factors in category formation, however these factors are said to result in highly path dependent category systems as a result of communication. Combined, these two views leaves us with a paradox concerning cross-cultural convergence: if individual vary in how they categorize novel continuous domains, and this individual-level variation leads to path-dependency, how do get cross-cultural convergence at all?

This dissertation argues that insights from the social network dynamics of category formation can resolve this paradox. Through the use of formal models and online experiments, this study shows that social processes of category formation need not give rise to path-dependency. The central and counterintuitive finding is that by increasing network size, and thereby increasing the diversity of categories in a population, communication can actually lead to predictable patterns of cross-cultural convergence when categorizing a novel continuum of objects. As such, this finding opens to a new domain of inquiry, concerning how social network structure underlies patterns of divergence and convergence in the social construction of category systems.

This research program has direct practical implications for a widespread problem in contemporary data production and management – that is, the problem of categorizing massive amounts of rapidly emerging and continuous content (e.g. new companies, art, and technology). Social groups are unable to ascribe value to these cultural products until they can effectively categorize them (Durand, Granqvist, and Tyllström 2017; Hannan et al. 2007; Zuckerman 1999, 2012). One domain where this is especially concerning is in content moderation over social media. Potentially harmful content floods social media websites like Facebook everyday in the millions, and yet most of this content is too new and culturally nuanced to be classified by existing machine learning algorithms (Gillespie 2018; Klonick 2018). Currently, social media companies are attempting to use large-scale crowdsourcing to manage content moderation, where online workers manually flag content as potentially harmful, with two major limitations (Gillespie 2018; Klonick 2018): (1) people vary wildly in how they interpret and apply the categories required by Facebook’s Community Standards (e.g. “crime”, “violence”, and “bullying”), and (2) people are slow to classify new content, leaving millions of users at risk everyday. The assumption in crowdsourcing is that people need to be kept independent to produce accurate classifications (Gillespie 2018; Lorenz et al. 2011; Pennycook and Rand 2019), because communication amplifies bias and leads to group think (Lorenz et al. 2011; Sunstein and Hastie 2014). The result is an unmanageable amount of diversity in the categories formed by independent coders, which in aggregate rarely provides social media companies with a reliable signal regarding which content they should and should not remove. The theoretical framework developed in this dissertation can be used to test an alternative approach in which category diversity can, counterintuitively, accelerate

the emergence of consensus in collective category formation. Specifically, the core theory in this dissertation suggests that allowing moderators to collaboratively classify content in social networks can accelerate and improve classification, enabling content moderation at scale. I refer to this as *networked crowdsourcing*.

Testing networked crowdsourcing in the domain of content moderation would provide an opportunity to address another longstanding debate at the boundary of nativism and social constructivism – namely, the question of whether people’s moral classifications are determined by innate moral instincts or by peer influence and conformity effects (Landy and Bartels 2018; Sinnott-Armstrong 2012). In the world of content moderation, variation among individuals’ categories for moral content is associated with a systematic problem that this experiment can both examine and potentially alleviate, i.e. bias (e.g. racial and gender) among coders undertaking crowdsourcing (Gillespie 2018; Noble 2018). In future work, it will be possible to build networks consisting of people who share a demographic trait linked to bias (e.g. “men”, in association with gender bias); this approach can then be used to grow and compare the classification systems produced by distinct, homogenous social networks (e.g. all male and all female networks) to measure how these groups differ in their perceptions of inappropriate content. Once separate category systems are grown in distinct social groups, the next step is to examine whether communication between these distinct social networks can reduce bias, similar to my earlier work, where bipartisan communication networks were shown to eliminate political bias in the interpretation of climate data (Guilbeault, Becker, and Centola 2018).

An important implication of recent collective intelligence work on bias reduction is that the structure of social networks plays a key causal role in whether networks amplify, or reduce, bias. Future work will also explore the role of topology in determining whether networked crowdsourcing entrenches or eliminates ideological bias. The aim is to discover whether there is an optimal level of connectivity to create between separate ideologically homogeneous communities so that their perspectives are integrated. In general, the question of how topological variation shapes category formation processes will be a pivotal focus in future research. A range of topological variables, including modularity and path length, are already linked with formal hypotheses posited in theoretical social science for how they should impact category formation (DiMaggio 1987; Gong et al. 2012; Milroy 1987), and these hypotheses are now amenable to direct empirical investigation within the experimental framework developed in this dissertation.

Moreover, content moderation is just one domain where crowdsourcing is applied to address extant categorization problems. Social media is not the only arena of social interaction flooded with massive amounts of novel and continuous content. Another prime and pertinent example is the domain of scientific classification. An illustrative story comes from the citizen science platform *Galaxy Zoo* (Salganik 2017; Watson and Floridi 2018). *Galaxy Zoo* was developed to address a pervasive classification problem in astronomy. While billions of dollars have been invested into satellites that take millions of images of space everyday, there is a paucity of scientists with sufficient time and expertise to examine these images. *Galaxy zoo* developed an online crowdsourcing system where anonymous users on the internet volunteer to classify a variety of stars and galaxies structured by continuous geometrical dimensions. This platform became shockingly successful when a large number of crowdworkers interacting over *Galaxy*

Zoo's discussion boards unexpectedly discovered a new kind of star – the *Green Pea*⁶ (Watson and Floridi 2018) – which became the subject of several technical publications. Since this discovery, Galaxy Zoo quickly expanded into *Zooniverse*, a platform with a diverse set of crowdsourcing efforts on a range of topics from gravity wave detection to coding primate behavior (Jackson et al. 2018; Zevin et al. 2017). Networked crowdsourcing has the potential to greatly enhance the emergence of consensus and accuracy in scientific classification schemes; and theoretically, it has the potential to significantly deepen our understanding of how scientific taxonomies emerge, clash, and grow in social groups of different sizes, structures, and demographic compositions (Wu, Wang, and Evans 2019), with broader implications for the structure of scientific revolutions (Collins 1998; Kuhn 1996).

Most fundamentally, the network dynamics of category formation has theoretical implications for our understanding of what constitutes social structure itself as an object of inquiry in social science. A foundational idea in sociology is that the structure of society, in terms of group relations and institutions, underlies the structure of category systems that emerge in society (Douglas 1986; Durkheim 1912; Simmel 1964)⁷. More recent work in sociology (Bowker and Star 2000) – generally influenced by the paradigm of *symbolic interactionism* (Blumer 1986) – has shifted focus to how social structures gain their influence through how people in social groups categorize social roles, actions,

⁶ Note how the term draws an explicit analogy to an existing concept, the green pea, suggesting that the *Green Pea* discovery was not the result of introducing an arbitrary label from the infinite void of possible arbitrary labels, but that instead it was sampled by a set of labels constrained by analogy, akin to the dynamics observed in the experiment discussed in chapter 2.

⁷ Durkheim ([1912] 2008) articulated this view particularly strongly in his supposition that even the abstract concept of a set, for which the elements of the set are members, derives from an original representation of social groups in which people are members.

and the conditions of communication, where changes in category systems are the primary vehicle through which social construction can lead to changes in social structure. This dissertation and the burgeoning body of work it evokes raises the broader question of whether, and if so, how the network dynamics of category formation mediate the collective interpretation of social structure itself, as a driving force in both the maintenance and creation of social order. These theoretical implications, along with the array of practical implications discussed above, will be the focus of future work to come.

BIBLIOGRAPHY

- Adamic, Lada, and Bernardo Huberman. 2002. "Zipf's Law and the Internet." *Glottometrics* 3:143–50.
- Allport, Gordon Willard. 1954. *The Nature of Prejudice*. Addison-Wesley Pub. Co.
- Anderson, P. W. 1972. "More Is Different." *Science* 177(4047):393–96.
- Atkinson, Quentin D., Andrew Meade, Chris Venditti, Simon J. Greenhill, and Mark Pagel. 2008. "Languages Evolve in Punctuational Bursts." *Science* 319(5863):588–588.
- Baronchelli, A., M. Felici, E. Caglioti, V. Loreto, and L. Steels. 2006. "Sharp Transition towards Shared Vocabularies in Multi-Agent Systems." *Journal of Statistical Mechanics: Theory and Experiment* 2006(06):P06014–P06014.
- Baronchelli, Andrea, Tao Gong, Andrea Puglisi, and Vittorio Loreto. 2010. "Modeling the Emergence of Universality in Color Naming Patterns." *Proceedings of the National Academy of Sciences* 107(6):2403–7.
- Becker, Howard Saul. 1984. *Art Worlds*. University of California Press.
- Berger, Peter L., and Thomas Luckmann. 1967. *The Social Construction of Reality: A Treatise in the Sociology of Knowledge*. New York: Anchor.
- Blumer, Herbert. 1986. *Symbolic Interactionism: Perspective and Method*. First edition. Berkeley, Calif.: University of California Press.

- Boone, Lalia Phipps. 1949. "Patterns of Innovation in the Language of the Oil Field." *American Speech* 24(1):31–37.
- Borges, Jorge Luis. 1973. *Other Inquisitions*. London: Souvenir Press Ltd.
- Borja, Mario Cortina, and John Haigh. 2007. "The Birthday Problem." *Significance* 4(3):124–27.
- Bowern, Claire. 2010. "Correlates of Language Change in Hunter-Gatherer and Other 'Small' Languages." *Language and Linguistics Compass* 4:665–79.
- Bowker, Geoffrey C., and Susan Leigh Star. 2000. *Sorting Things Out: Classification and Its Consequences*. MIT Press.
- Brennan, S. E., and H. H. Clark. 1996. "Conceptual Pacts and Lexical Choice in Conversation." *Journal of Experimental Psychology. Learning, Memory, and Cognition* 22(6):1482–93.
- Bromham, Lindell, Xia Hua, Thomas G. Fitzpatrick, and Simon J. Greenhill. 2015. "Rate of Language Evolution Is Affected by Population Size." *Proceedings of the National Academy of Sciences* 112(7):2097–2102.
- Brown, Cecil. 1979. "Folk Zoological Life-Forms: Their Universality and Growth." *American Anthropologist* 81(4):791–817.
- Brown, Cecil H. 1984. *Language and Living Things: Uniformities in Folk Classification and Naming*. Rutgers University Press.
- Brown, Cecil H., and Stanley R. Witkowski. 1981. "Figurative Language in a Universalist Perspective." *American Ethnologist* 8(3):596–615.

- Brown, Donald E. 2004. "Human Universals, Human Nature & Human Culture." *Daedalus* 133(4):47–54.
- Burr, Vivien. 2003. *Social Constructionism*. Psychology Press.
- Burris, Harold. 1979. "Geometric Figure Terms: Their Universality and Growth." *The Journal of Anthropology* 1(2):18–41.
- C B Mervis, and and E. Rosch. 1981. "Categorization of Natural Objects." *Annual Review of Psychology* 32(1):89–115.
- Centola, Damon. 2010. "The Spread of Behavior in an Online Social Network Experiment." *Science (New York, N. Y.)* 329(5996):1194–97.
- Centola, Damon. 2011. "An Experimental Study of Homophily in the Adoption of Health Behavior." *Science* 334(6060):1269–72.
- Centola, Damon. 2015. "The Social Origins of Networks and Diffusion." *American Journal of Sociology* 120(5):1295–1338.
- Centola, Damon. 2018. *How Behavior Spreads*. Princeton, N.J.: Princeton University Press.
- Centola, Damon, and Andrea Baronchelli. 2015. "The Spontaneous Emergence of Conventions: An Experimental Study of Cultural Evolution." *Proceedings of the National Academy of Sciences* 112(7):1989–94.
- Centola, Damon, Joshua Becker, Devon Brackbill, and Andrea Baronchelli. 2018. "Experimental Evidence for Tipping Points in Social Convention." *Science* 360(6393):1116–19.

- Cerulo, Karen A. 1995. *Identity Designs: The Sights and Sounds of a Nation*. Rutgers University Press.
- Clark, H. H., and D. Wilkes-Gibbs. 1986. "Referring as a Collaborative Process." *Cognition* 22(1):1–39.
- Collins, Randall. 1998. *The Sociology of Philosophies: A Global Theory of Intellectual Change*. Revised edition. Cambridge, Mass.: Belknap Press of Harvard University Press.
- David, Paul A. 2007. "Path Dependence: A Foundational Concept for Historical Social Science." *Econometrica* 1(2):91–114.
- DellaPosta, Daniel, Yongren Shi, and Michael Macy. 2015. "Why Do Liberals Drink Lattes?" *American Journal of Sociology* 120(5):1473–1511.
- Dereux, Maxime, Marie-Pauline Beugin, Bernard Godelle, and Michel Raymond. 2013. "Experimental Evidence for the Influence of Group Size on Cultural Complexity." *Nature* 503(7476):389–91.
- DiMaggio, Paul. 1987. "Classification in Art." *American Sociological Review* 52(4):440–55.
- DiMaggio, Paul. 1997. "Culture and Cognition." *Annual Review of Sociology* 23(1):263–87.
- Douglas, Mary. 1986. *How Institutions Think*. First Edition edition. Syracuse, N.Y: Syracuse University Press.

- Durand, Rodolphe, Nina Granqvist, and Anna Tyllström. 2017. *From Categories to Categorization: Studies in Sociology, Organizations and Strategy at the Crossroads*. Emerald Group Publishing.
- Durkheim, Emile. 1912. *The Elementary Forms of Religious Life*. 1 edition. edited by M. S. Cladis. Oxford: Oxford University Press.
- Emirbayer, Mustafa. 1997. "Manifesto for a Relational Sociology." *American Journal of Sociology* 103(2):281–317.
- Erikson, Emily. 2013. "Formalist and Relationalist Theory in Social Network Analysis." *Sociological Theory* 31(3):219–42.
- Fay, Nicolas, and T. Mark Ellison. 2013. "The Cultural Evolution of Human Communication Systems in Different Sized Populations: Usability Trumps Learnability." *PLOS ONE* 8(8):e71781.
- Fay, Nicolas, Simon Garrod, and Leo Roberts. 2008. "The Fitness and Functionality of Culturally Evolved Communication Systems." *Philosophical Transactions of the Royal Society B: Biological Sciences* 363(1509):3553–61.
- Fay, Nicolas, Simon Garrod, Leo Roberts, and Nik Swoboda. 2010a. "The Interactive Evolution of Human Communication Systems." *Cognitive Science* 34(3):351–86.
- Fay, Nicolas, Simon Garrod, Leo Roberts, and Nik Swoboda. 2010b. "The Interactive Evolution of Human Communication Systems." *Cognitive Science* 34(3):351–86.
- Fay, Nicolas, Bradley Walker, Nik Swoboda, and Simon Garrod. 2018. "How to Create Shared Symbols." *Cognitive Science* 42(S1):241–69.

- Flache, Andreas, and Michael W. Macy. 2006. "What Sustains Cultural Diversity and What Undermines It? Axelrod and Beyond." *ArXiv:Physics/0604201*.
- Flache, Andreas, and Michael W. Macy. 2011. "Local Convergence and Global Diversity: From Interpersonal to Social Influence." *The Journal of Conflict Resolution* 55(6):970–95.
- Fodor, Jerry A. 1998. *Concepts: Where Cognitive Science Went Wrong*. Oxford University Press.
- Foucault, Michel. 1988. *Madness and Civilization: A History of Insanity in the Age of Reason*. 1 edition. New York: Vintage.
- Foucault, Michel. 1990. *The History of Sexuality, Vol. 1: An Introduction*. Reissue edition. New York: Vintage.
- Freeberg, Todd M., Terry J. Ord, and Robin I. M. Dunbar. 2012. "The Social Network and Communicative Complexity: Preface to Theme Issue." *Philosophical Transactions of the Royal Society B: Biological Sciences* 367(1597):1782–84.
- Galantucci, Bruno, and Simon Garrod. 2011. "Experimental Semiotics: A Review." *Frontiers in Human Neuroscience* 5.
- Garrod, Simon, and Gwyneth Doherty. 1994. "Conversation, Co-Ordination and Convention: An Empirical Investigation of How Groups Establish Linguistic Conventions." *Cognition* 53(3):181–215.
- Gelman, Susan A. 2005. *The Essential Child: Origins of Essentialism in Everyday Thought*. Oxford: Oxford University Press.

- Gillespie, Tarleton. 2018. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. New Haven: Yale University Press.
- Gil-White, Francisco J. 2001. "Are Ethnic Groups Biological 'Species' to the Human Brain? Essentialism in Our Cognition of Some Social Categories." *Current Anthropology* 42(4):515–53.
- Goddard, Cliff. 2008. "Lexico-Semantic Universals: A Critical Overview." *Linguistic Typology* 5(1):1–65.
- Goffman, Erving, and Bennett Berger. 1986. *Frame Analysis: An Essay on the Organization of Experience*. Later Reprint edition. Boston: Northeastern University Press.
- Gong, Tao, Andrea Baronchelli, Andrea Puglisi, and Vittorio Loreto. 2012. "Exploring the Roles of Complex Networks in Linguistic Categorization." *Artificial Life* 18(1):107–21.
- Gordon, Robert. 1984. "Critical Legal Histories." *Faculty Scholarship Series*.
- Gorwa, Robert, and Douglas Guilbeault. 2018. "Unpacking the Social Media Bot: A Typology to Guide Research and Policy." *Policy & Internet*.
- Gould, S.J. 2002. *The Structure of Evolutionary Theory*. Cambridge, Mass: Harvard University Press.

- Guilbeault, Douglas, Joshua Becker, and Damon Centola. 2018. "Social Learning and Partisan Bias in the Interpretation of Climate Trends." *Proceedings of the National Academy of Sciences* 115(39):9714–19.
- Guilbeault, Douglas, Ethan Nadler, Mark Chu, Ruggiero Lo Sardo, Aabir Abubaker Kar, and Bhargav Srinivasa Desikan. 2020. "Color Associations in Abstract Semantic Domains." *Cognition (Forthcoming)*.
- Hannan, Michael T., László Pólos, and Glenn R. Carroll. 2007. *Logics of Organization Theory: Audiences, Codes, and Ecologies*. Princeton, N.J: Princeton University Press.
- Huang, Yi Ting, and Jesse Snedeker. 2009. "Semantic Meaning and Pragmatic Interpretation in 5-Year-Olds: Evidence from Real-Time Spoken Language Comprehension." *Developmental Psychology* 45(6):1723–39.
- Jackson, Corey Brian, Kevin Crowston, Carsten Østerlund, Mahboobeh Harandi, Mahboobeh Harandi, and Mahboobeh Harandi. 2018. "Folksonomies to Support Coordination and Coordination of Folksonomies." *Computer Supported Cooperative Work* 27(3–6):647–678.
- Jackson, Joshua Conrad, Joseph Watts, Teague R. Henry, Johann-Mattis List, Robert Forkel, Peter J. Mucha, Simon J. Greenhill, Russell D. Gray, and Kristen A. Lindquist. 2019. "Emotion Semantics Show Both Cultural Variation and Universal Structure." *Science* 366(6472):1517–22.
- James, William. 2000. *The Principles of Psychology: Volume 1*. New edition edition. New York: Dover Publications Inc.

- Jamieson, Kathleen Hall. 1996. *Packaging The Presidency: A History and Criticism of Presidential Campaign Advertising*. Oxford University Press.
- Kauffman, Stuart A. 1993. *The Origins of Order: Self-Organization and Selection in Evolution*. 1 edition. New York: Oxford University Press, U.S.A.
- Keller, Rudi. 2005. *On Language Change: The Invisible Hand in Language*. 1 edition. Routledge.
- Kemp, Charles, and Terry Regier. 2012. "Kinship Categories Across Languages Reflect General Communicative Principles." *Science* 336(6084):1049–54.
- Kirby, Simon, Hannah Cornish, and Kenny Smith. 2008. "Cumulative Cultural Evolution in the Laboratory: An Experimental Approach to the Origins of Structure in Human Language." *Proceedings of the National Academy of Sciences* 105(31):10681–86.
- Kirby, Simon, Mike Dowman, and Thomas L. Griffiths. 2007. "Innateness and Culture in the Evolution of Language." *Proceedings of the National Academy of Sciences* 104(12):5241–45.
- Klonick, Kate. 2018. "The New Governors: The People, Rules, and Processes Governing Online Speech." *Harvard Law Review* 131:1589–1670.
- Kuhn, Thomas S. 1996. *The Structure of Scientific Revolutions*. New ed of 3 Revised ed. Chicago, IL: University of Chicago Press.

- Landy, Justin F., and Daniel M. Bartels. 2018. "An Empirically-Derived Taxonomy of Moral Concepts." *Journal of Experimental Psychology: General* 147(11):1748–61.
- Lant, Theresa K., and Stephen J. Mezias. 1992. "An Organizational Learning Model of Convergence and Reorientation." *Organization Science* 3(1):47–71.
- Latour, Bruno. 1988. *Science in Action: How to Follow Scientists and Engineers through Society*. REP edition. Cambridge, Mass: Harvard University Press.
- Latour, Bruno, Steve Woolgar, and Jonas Salk. 1986. *Laboratory Life: The Construction of Scientific Facts, 2nd Edition*. 2nd edition. Princeton, N.J: Princeton University Press.
- Laurence, Stephen, and Eric Margolis. 2002. "Radical Concept Nativism." *Cognition* 86(1):25–55.
- Leach, Edmund. 1989. "Anthropological Aspects of Language: Animal Categories and Verbal Abuse." *Anthrozoös* 2(3):151–65.
- Lindsey, Delwin T., Angela M. Brown, David H. Brainard, and Coren L. Apicella. 2016b. "Hadza Color Terms Are Sparse, Diverse, and Distributed, and Presage the Universal Color Categories Found in Other World Languages." *I-Perception* 7(6).
- Lorenz, Jan, Heiko Rauhut, Frank Schweitzer, and Dirk Helbing. 2011. "How Social Influence Can Undermine the Wisdom of Crowd Effect." *Proceedings of the National Academy of Sciences* 108(22):9020–25.

- Macy, Michael, Sebastian Deri, Alexander Ruch, and Natalie Tong. 2019. "Opinion Cascades and the Unpredictability of Partisan Polarization." *Science Advances* 5(8):eaax0754.
- Malt, B. C. 1995. "Category Coherence in Cross-Cultural Perspective." *Cognitive Psychology* 29(2):85–148.
- Margolis, Eric, and Stephen Laurence. 2013. "In Defense of Nativism." *Philosophical Studies* 165(2):693–718.
- Medin, Douglas L., and Scott Atran. 2004. "The Native Mind: Biological Categorization and Reasoning in Development and across Cultures." *Psychological Review* 111(4):960–83.
- Medin, Douglas L., Elizabeth B. Lynch, John D. Coley, and Scott Atran. 1997. "Categorization and Reasoning among Tree Experts: Do All Roads Lead to Rome?" *Cognitive Psychology* 32(1):49–96.
- Medin, Douglas L., William D. Wattenmaker, and Sarah E. Hampson. 1987. "Family Resemblance, Conceptual Cohesiveness, and Category Construction." *Cognitive Psychology* 19(2):242–79.
- Medina, Tamara Nicol, Jesse Snedeker, John C. Trueswell, and Lila R. Gleitman. 2011. "How Words Can and Cannot Be Learned by Observation." *Proceedings of the National Academy of Sciences* 108(22):9014–19.
- Milroy, Lesley. 1987. *Language and Social Networks*. Wiley.

- Mohr, John W. 1998. "Measuring Meaning Structures." *Annual Review of Sociology* 24(1):345–70.
- Mufwene, Salikoko S. 2001. *The Ecology of Language Evolution*. Cambridge University Press.
- Munford, A. G. 1977. "A Note on the Uniformity Assumption in the Birthday Problem." *The American Statistician* 31(3):119–119.
- Nettle, Daniel. 1999. "Using Social Impact Theory to Simulate Language Change." *Lingua* 108(2):95–117.
- Nettle, Daniel. 2012. "Social Scale and Structural Complexity in Human Languages." *Philosophical Transactions of the Royal Society B: Biological Sciences* 367(1597):1829–36.
- Newberry, Mitchell G., Christopher A. Ahern, Robin Clark, and Joshua B. Plotkin. 2017. "Detecting Evolutionary Forces in Language Change." *Nature* 551(7679):223–26.
- Noble, Safiya Umoja. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. 1 edition. New York: NYU Press.
- Obukhova, Elena, Ezra W. Zuckerman, and Jiayin Zhang. 2014. "When Politics Froze Fashion: The Effect of the Cultural Revolution on Naming in Beijing." *American Journal of Sociology* 120(2):555–83.
- Pagel, Mark, Quentin D. Atkinson, and Andrew Meade. 2007. "Frequency of Word-Use Predicts Rates of Lexical Evolution throughout Indo-European History." *Nature* 449(7163):717.

- Pagel, Mark, Mark Beaumont, Andrew Meade, Annemarie Verkerk, and Andreea Calude. 2019. "Dominant Words Rise to the Top by Positive Frequency-Dependent Selection." *Proceedings of the National Academy of Sciences* 116(15):7397–7402.
- Pennycook, Gordon, and David G. Rand. 2019. "Fighting Misinformation on Social Media Using Crowdsourced Judgments of News Source Quality." *Proceedings of the National Academy of Sciences* 201806781.
- Peterson, Richard A. 1999. *Creating Country Music: Fabricating Authenticity*. First Edition. Chicago: University of Chicago Press.
- Pinch, Trevor, and Frank Trocco. 1998. "The Social Construction of the Early Electronic Music Synthesizer." *Icon* 4:9–31.
- Pinker, Steven. 1994. *The Language Instinct*. W. Morrow and Company.
- Pinker, Steven. 2003. *The Blank Slate: The Modern Denial of Human Nature*. Reprint edition. New York etc.: Penguin Books.
- Pinker, Steven, and Arthur Morey. 2014. *The Language Instinct: How the Mind Creates Language*. Unabridged edition. Brilliance Audio.
- Puglisi, Andrea, Andrea Baronchelli, and Vittorio Loreto. 2008. "Cultural Route to the Emergence of Linguistic Categories." *Proceedings of the National Academy of Sciences* 105(23):7936–40.
- Ranjan, Aparna, and Narayanan Srinivasan. 2010. "Dissimilarity in Creative Categorization." *The Journal of Creative Behavior* 44(2):71–83.

- Rawlings, Craig M., and Clayton Childress. 2019. "Emergent Meanings: Reconciling Dispositional and Situational Accounts of Meaning-Making from Cultural Objects." *American Journal of Sociology* 124(6):1763–1809.
- Regier, Terry, Alexandra Carstensen, and Charles Kemp. 2016. "Languages Support Efficient Communication about the Environment: Words for Snow Revisited." *PLOS ONE* 11(4):e0151138.
- Regier, Terry, Paul Kay, and Naveen Khetarpal. 2007. "Color Naming Reflects Optimal Partitions of Color Space." *Proceedings of the National Academy of Sciences* 104(4):1436–41.
- van de Rijt, Arnout. 2019. "Self-Correcting Dynamics in Social Influence Processes." *American Journal of Sociology* 124(5):1468–95.
- Rosch, Eleanor. 1975. "Cognitive Representations of Semantic Categories." *Journal of Experimental Psychology: General* 104(3):192–233.
- Rosch, Eleanor. 2002. *Principles of Categorization*. Cambridge, MA, US: MIT Press.
- Rosch, Eleanor H. 1973. "Natural Categories." *Cognitive Psychology* 4(3):328–50.
- Rosch, Eleanor, and Carolyn B. Mervis. 1975. "Family Resemblances: Studies in the Internal Structure of Categories." *Cognitive Psychology* 7(4):573–605.
- Rosch, Eleanor, Carolyn B. Mervis, Wayne D. Gray, David M. Johnson, and Penny Boyes-Braem. 1976. "Basic Objects in Natural Categories." *Cognitive Psychology* 8(3):382–439.
- Salganik, Matthew J. 2017. *Bit by Bit*. Princeton, N.J: Princeton University Press.

- Salganik, Matthew J., Peter Sheridan Dodds, and Duncan J. Watts. 2006. "Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market." *Science* 311(5762):854–56.
- Searle, John R. 1995. *The Construction of Social Reality*. Simon and Schuster.
- Shaw, Lynette. 2015. "Mechanics and Dynamics of Social Construction: Modeling the Emergence of Culture from Individual Mental Representation." *Poetics* 52:75–90.
- Shepard, Roger N., and Gregory W. Cermak. 1973. "Perceptual-Cognitive Explorations of a Toroidal Set of Free-Form Stimuli." *Cognitive Psychology* 4(3):351–77.
- Shwed, Uri, and Peter S. Bearman. 2010. "The Temporal Structure of Scientific Consensus Formation." *American Sociological Review* 75(6):817–40.
- Shweder, Richard A., and Edmund J. Bourne. 1982. "Does the Concept of the Person Vary Cross-Culturally?" Pp. 97–137 in *Cultural Conceptions of Mental Health and Therapy, Culture, Illness, and Healing*, edited by A. J. Marsella and G. M. White. Dordrecht: Springer Netherlands.
- Silvey, Catriona, Simon Kirby, and Kenny Smith. 2015. "Word Meanings Evolve to Selectively Preserve Distinctions on Salient Dimensions." *Cognitive Science* 39(1):212–26.
- Simmel, Georg. 1964. *Conflict / The Web Of Group Affiliations*. 1st Free Press Pbk. Ed edition. New York: Free Press.
- Sinnott-Armstrong, Walter. 2012. "Does Morality Have an Essence?" *Psychological Inquiry* 23(2):194–97.

- Smith, Christian. 2010. *What Is a Person?: Rethinking Humanity, Social Life, and the Moral Good from the Person Up*. University of Chicago Press.
- Spalding, Thomas, and Murphy Gregory. 1996. "Effects of Background Knowledge on Category Construction." *Journal of Experimental Psychology* 22(2):525–38.
- Steels, Luc, and Tony Belpaeme. 2005. "Coordinating Perceptually Grounded Categories through Language: A Case Study for Colour." *The Behavioral and Brain Sciences* 28(4):469–89; discussion 489-529.
- Sunstein, Cass R., and Reid Hastie. 2014. *Wiser: Getting Beyond Groupthink to Make Groups Smarter*. First Edition edition. Boston, Massachusetts: Harvard Business Review Press.
- Swidler, Ann. 1986. "Culture in Action: Symbols and Strategies." *American Sociological Review* 51(2):273–86.
- Tilly, Charles. 1997. "Parliamentarization of Popular Contention in Great Britain, 1758-1834." *Theory and Society* 26(2/3):245–73.
- Trehub, Sandra E. 2015. "Cross-Cultural Convergence of Musical Features." *Proceedings of the National Academy of Sciences* 112(29):8809–10.
- Watson, David, and Luciano Floridi. 2018. "Crowdsourced Science: Sociotechnical Epistemology in the e-Research Paradigm." *Synthese* 195(2):741–64.
- White, Harrison Colyar. 2012 [1964]. *An Anatomy Of Kinship: Mathematical Models For Structures Of Cumulated Roles*. edited by J. Coleman and J. March. Literary Licensing, LLC.

- Winkielman, Piotr, Jamin Halberstadt, Tedra Fazendeiro, and Steve Catty. 2006. "Prototypes Are Attractive Because They Are Easy on the Mind." *Psychological Science* 17(9):799–806.
- Witkowski, Stanley, and Harold Burris. 1981. "Societal Complexity and Lexical Growth." *Cross-Cultural Research* 16(1–2):143–59.
- Wittgenstein, Ludwig. 1965. *The Blue and Brown Books*. HarperCollins.
- Wittgenstein, Ludwig. 1973. *Philosophical Investigations*. 3rd edition. Englewood Cliffs, N.J: Pearson.
- Wu, Lingfei, Dashun Wang, and James A. Evans. 2019. "Large Teams Develop and Small Teams Disrupt Science and Technology." *Nature* 566(7744):378.
- Xu, Yang, and Terry Regier. 2014. "Numeral Systems across Languages Support Efficient Communication: From Approximate Numerosity to Recursion." *Proceedings of the Annual Meeting of the Cognitive Science Society* 36(36).
- Youn, Hyejin, Logan Sutton, Eric Smith, Cristopher Moore, Jon F. Wilkins, Ian Maddieson, William Croft, and Tanmoy Bhattacharya. 2016. "On the Universal Structure of Human Lexical Semantics." *Proceedings of the National Academy of Sciences* 113(7):1766–71.
- Zevin, M., S. Coughlin, S. Bahaadini, E. Besler, N. Rohani, S. Allen, M. Cabero, K. Crowston, A. K. Katsaggelos, S. L. Larson, T. K. Lee, C. Lintott, T. B. Littenberg, A. Lundgren, C. Østerlund, J. R. Smith, L. Trouille, and V. Kalogera. 2017. "Gravity Spy: Integrating Advanced LIGO Detector Characterization, Machine Learning, and Citizen Science." *Classical and Quantum Gravity* 34(No 6).

Zuckerman, Ezra W. 1999. "The Categorical Imperative: Securities Analysts and the Illegitimacy Discount." *American Journal of Sociology* 104(5):1398–1438.

Zuckerman, Ezra W. 2004. "Structural Incoherence and Stock Market Activity." *American Sociological Review* 69(3):405–32.

Zuckerman, Ezra W. 2012. "Construction, Concentration, and (Dis)Continuities in Social Valuations." *Annual Review of Sociology* 38(1):223–45.