Beyond "Social Contagion": Associative Diffusion and the Emergence of Cultural Variation*

Amir Goldberg
Stanford University

Sarah K. Stein
Stanford University

March 30, 2018

Abstract

Network models of diffusion predominantly think about cultural variation as a product of *social contagion*. But culture does not spread like a virus. In this paper, we propose an alternative explanation which we refer to as *associative diffusion*. Drawing on two insights from research in cognition—that meaning inheres in cognitive associations between concepts, and that such perceived associations constrain people's actions—we propose a model wherein, rather than beliefs or behaviors per-se, the things being transmitted between individuals are perceptions about what beliefs or behaviors are compatible with one another. Conventional contagion models require an assumption of network segregation to explain cultural variation. In contrast, we demonstrate that the endogenous emergence of cultural differentiation can be entirely attributable to social cognition and does not necessitate a clustered social network or a preexisting division into groups. Moreover, we show that prevailing assumptions about the effects of network topology do not hold when diffusion is associative.

---

[*] Direct all correspondence to Amir Goldberg: amirgo@stanford.edu; 650-725-7926. We thank Rembrand M. Koning, Daniel A. McFarland, Paul DiMaggio, Sarah Soule, John-Paul Ferguson, Jesper Sørensen, Glenn Carroll and Sameer B. Srivastava for helpful comments and suggestions. The usual disclaimer applies.

Introduction

Contemporary societies exhibit remarkable and persistent cultural differences on issues as varied as musical taste and gun control. This cultural variation has been a longstanding topic of sociological inquiry because it is central to how social order, and the inequalities it is founded on, is maintained (Lamont, Molnár, and Virág 2002; Bourdieu 1986; Goldberg, Hannan, and Kovács 2016). In particular, research has focused on the tendency of cultural practices to co-occur with one another (Martin 2002).[1] This cultural clustering—from clothing and lifestyle choices to consumption and religious behaviors—symbolically marks different categories of people. The different social identities these cultural boundaries delineate are typically undergirded by a variation in beliefs and dispositions. These divergent beliefs, such as those on the epistemic authority of science (Gauchat 2012) or the moral qualities of craft distillers (Ocejo 2017), structure differences in individuals' political, economic and health behaviors.

Where does this patterned cultural variation come from? Sociological work has predominantly studied cultural diffusion through the prism of "social contagion" (e.g. Christakis and Fowler 2007; Papachristos 2009). These network diffusion models commonly attribute cultural heterogeneity to structural boundaries to diffusion. Studies either assume the preexistence of a segregated social structure whereby cultural practices diffuse within, but not across network clusters (Centola and Macy 2007; Dandekar, Goel, and Lee 2013), or that structural disconnection emerges endogenously through actors' tendencies or incentives to preferentially interact with others who are culturally similar (Axelrod 1997; Centola, González-Avella, Eguíluz, and San Miguel 2007; Mark 2003; Baldassarri and Bearman 2007). Whatever the underlying mechanism, the end result is a balkanized world in which people interact within culturally homogenous, and structurally separated cliques. Cultural differentiation, in other words, is ultimately epiphenomenal of a structurally segmented world (DellaPosta, Shi, and Macy 2015).

But cultural diffusion often fails to trace network structure. Consider the recent surge in opposition to vaccinations as an example. Successful immunization campaigns virtually eradicated measles and other childhood diseases in the U.S. by the end of the twentieth century; but at the turn of the new millennium Americans' faith in vaccines began to erode (Horne, Powell, Hummel, and Holyoak 2015). The spread of anti-vaccination sentiments appears inconsistent with a network diffusion explanation. Parents who object to childhood vaccinations tend to be college educated and to have above average incomes. This is precisely the demographic that, up until recently, was most likely to comply with vaccination protocols. Indeed, as the passionate disagreements on California's strict 2015 immunization law demonstrated, beliefs on vaccines cut through tight-knit communities across the U.S. Hailing from the Oregon coast to the Texas heartland, *anti-vaxxers*, as they are colloquially called, have also become a constituency wooed by candidates on both sides of the political spectrum.[2] Attitudes on vaccines, in other words, do not appear to follow the contours of network segregation in America.

An important distinction is missing from the epidemiological imagery informing network diffusion models: whether an actor *adopts* a cultural practice is different from how an actor *interprets* it. Behaviors around vaccines are strongly rooted in cultural beliefs. Injecting a biological agent using a hypodermic needle—without being able to observe this action's purported effects—requires strong and unquestionable faith in the institutional authority of the medical profession.[3] As work by historians and sociologists has demonstrated, recent changes in attitudes toward vaccines in the U.S. relate to changes in how Americans, predominantly those on the higher end of the socioeconomic ladder, understand their roles as parents, their rights as consumers and their relationships with pediatricians (Conis 2014; Reich 2016). The rise in opposition to vaccines by educated and affluent parents cannot be explained without taking into account how vaccines have been reinterpreted.

Diffusion studies generally disregard this interpretative dimension. When they do not, they focus on differences in inherent appeal between cultural practices (e.g. Berger and Milkman 2012; Gamson and Modigliani 1989). But such differences in appeal explain why some practices diffuse more broadly than others, not why patterned cultural variation emerges *within* a population of interacting agents. In contrast, we argue that interpretations, and interpretative consensus, can emerge through the process of diffusion. Drawing on two insights from research in cognition—that meaning inheres in cognitive associations between concepts, and that such perceived associations constrain people's actions—we propose a theory wherein rather than beliefs or behaviors per-se, the cultural elements being transmitted between individuals are perceptions about which beliefs or behaviors are compatible with one another. People learn from their social environments how to associate between different cultural practices, and in their own behaviors enact, and therefore reproduce for others, these associations.

We formalize these two assumptions into a model of *associative diffusion*, and using agent-based modeling demonstrate that cultural differentiation emerges in a population even in the absence of an *a priori* segregated social structure or of homophilious interaction. We also explore a host of alternative explanations—including direct imitation, biased contagion, conformist contagion and homophilous contagion—and demonstrate that they cannot explain the emergence of cultural differentiation without assuming a preexisting or emergent structural division. We integrate these alternatives into our baseline model and further demonstrate that our findings are robust to the presence of nonconformists and to different network topologies. Our results also suggests that, contra the findings of previous work on networks and diffusion, a segregated network is least conducive to the emergence of pronounced cultural difference. Ultimately, our model turns the causal arrow in conventional accounts of social contagion on its head: we show that differentiation can emerge through the complex ways by which culture is cognitively represented and acted upon by individuals (Lizardo 2006).

Where Does Cultural Variation Come From?

Culture is often measured as the distribution of beliefs in a population. A consistent finding in the sociology of culture is that beliefs are not randomly distributed. Rather, people's cultural preferences and concomitant behaviors are strongly patterned, such that bundles of practices tend to co-occur with one another (Martin 2002). Parents who decide not to vaccinate their children, for example, often tend to embrace other health-related behaviors on ideological grounds. These parents commonly support home births, object to the consumption of genetically modified food and strongly believe in the health and developmental virtues of breastfeeding. As Reich (2016) demonstrates, these attitudes extend beyond the domain of health. The anti-vaccination parents she interviews tend to espouse strong individualism, to equate good parenting with intensive care-giving and to exhibit profound distrust toward big business.

Attitudes toward vaccination are not unique. Cultural practices tend to cluster together in all domains of social life. These cultural interdependencies are consequential because they delineate different social identities. From hackers' strong belief in individual liberty and admiration for the *Grateful Dead* (Turner 2008), to hipsters' anti-corporate activism and taste for craft beers (Carroll and Swaminathan 2000), cultural bundles carve cleavages in groups as small as adolescent sports teams and as large as national societies. The working-class Southerners who are the subject of Hochschild's (2016) study, for example, espouse a set of moral and ideological dispositions—religious devotion, pride in hard work, and a staunch opposition to government regulation—that, as they perceive, pit them in stark opposition to coastal liberals.

One explanation for this patterned clustering of cultural preferences is that different cultural elements are functionally dependent on one another. One might assume, for example, that a belief in the natural virtues of breastfeeding is logically consistent with an objection to the assumed unnaturalness of synthetic vaccines. But popular understandings of vaccines as unnatural are a relatively recent historical phenomenon.[4] Half a century ago, vaccines were predominantly understood as healthy and safe; vaccination was neither considered consistent with nor antithetical to breastfeeding. Narratives promoted by environmental activists since the 1960s, however, focused public discourse on issues such as pollution and industrial contamination. Vaccines, in turn, were reframed as toxic rather than safe (Conis 2015). The upper-middle class parents who, in the 1970s, would have enthusiastically vaccinated their children out of a sense of parental responsibility, are today most likely to invoke the same sense of responsibility to justify their objection to vaccinations (Conis 2014).

We use the term "culture" to refer to the social conventions that associate practices with meanings which—like in the case of vaccines—are not inherently derivative of these practices' formal properties. That is not to say that the patterned distribution of culture is entirely arbitrary; practices are limited by objective functional constraints (Zuckerman 2012). In some domains these constraints are fairly weak. There is no apparent functional reason why parents living in the mountains of Montana, for example, would be more likely to name their daughters *Jennifer* than those living on the Californian coast (Barucca, Rocchi, Marinari, Parisi, and Ricci-Tersenghi 2015). But in most realms of social life the distinction between the functional and symbolic attributes of cultural practices is less readily evident. The traditional association between high-brow music and intellectual sophistication, for example, has been challenged with the rise of cultural omnivorousness as a dominant logic of cultural consumption in Western societies (Peterson and Kern 1996). Similarly, the rationale that connects a belief in laissez-faire economics with an objection to legalized abortion is taken for granted in mainstream American political discourse but is not particularly prevalent elsewhere (Malka, Soto, Inzlicht, and Lelkes 2014; Baldassarri and Goldberg 2014). Though cultural clustering appears to be inexorable, the patterns it follows are not.

Contagion Models of Differentiated Cultural Diffusion

If the distributional patterns of cultural variation are not predetermined, then where do they come from? Recent sociological work has almost exclusively explored this question through a network diffusion lens. Cultural diffusion models rest on a well-established fact: humans exhibit an innate tendency to imitate others' behaviors and adopt their preferences and beliefs. The psychological reasons for this tendency are multifaceted, ranging from an evolutionary instinct for conformity to the need to resolve uncertainty in light of incomplete or complex information (Cialdini and Goldstein 2004). Whatever the underlying causes, social influence functions as the dyadic transmission channel through which cultural practices contagiously diffuse. Early adopters "virally" spread a newly acquired practice through their influence on others.

Simple and elegant as this epidemiological metaphor may be, it cannot explain the emergence of cultural differentiation. If individuals are straightforwardly influenced by their peers, then cultural practices should either saturate a population or fail to take off altogether. Indeed, a variety of sociological studies have explored the network topologies that facilitate, or hinder, the emergence of contagious cultural cascades (e.g. Granovetter 1978; Watts 2002; Centola 2015).

To explain systemic cultural variation social contagion theories need to assume the existence of structural boundaries to diffusion. These models generally provide two types of explanations for cultural differentiation. The first emphasizes the mechanism of *exposure*, assuming that adoption of a practice is a function of the structural opportunity to observe it. In their most rudimentary form, such models presuppose the existence of different social groups such that practices diffuse within, but not between them. The intuition behind this assumption is fairly straightforward: parents in Montana imitate their peers who name their daughters *Jennifer*, but this trend fails to reach structurally disconnected parents in California. Such differentiated diffusion can persist even in light of crosscutting connections between groups, as long as network ties are denser within groups than they are between them. When individuals require affirmation from multiple social connections before adopting a cultural practice then what Centola and Macy (2007) call "complex contagion" will lead to diffusion within, but not across, network clusters.

A second set of explanations focuses on the mechanism of *choice homophily*, namely, individuals' predisposition—due to intrinsic motivation or external rewards—toward culturally similar others. Such a proclivity leads to the emergence of culturally homogenous clusters either because individuals choose to interact homophilously, or because they are more susceptible to influence by culturally similar peers. Scholars of diffusion have suggested a variety of models that explore the complementary effects of homophily and social influence. Some emphasize the formation and dissolution of network ties (e.g. Carley 1991; Mark 1998; Centola et al. 2007), whereas others focus on changes in the strength of social influence as a function of cultural similarity (e.g. Dandekar et al. 2013; Flache and Macy 2011).[5] Whatever the differences in their underlying assumptions and emphases, all of these models describe a coevolutionary process whereby individuals become increasingly related to culturally similar others. Thus, even small and random initial variation gradually evolves into systematic cultural differentiation.

Essentially all contagion models—whether assuming the mechanisms of exposure or homophily, or both—describe cultural differentiation as a product of an underlying balkanized social network. Consider a recent study by DellaPosta et al. (2015). The authors propose an elegant and complex model in which an individual's likelihood of adopting a peer's cultural preference is proportional to the distance between the two agents in a socio-cultural space. Using this model, the authors demonstrate how the mutually reinforcing dynamics of influence and homophily can amplify minor cultural differences between individuals to generate the strong and seemingly arbitrary correlation between Americans' political ideology and lifestyle choices, creating the proverbial "latte liberals" and "bird-hunting conservatives." But to provide this explanation DellaPosta and colleagues' model presupposes a "connected caveman" small-world network in which individuals are segregated into sparsely interconnected and densely intraconnected clusters. Cultural differentiation, in other words, is epiphenomenal of an underlying and preexisting segmented social structure, a mere spurious byproduct of what the authors term "network autocorrelation."

The Cultural Conductivity of Superficial Interactions

A central assumption in theories of network diffusion is that cultural transmission only occurs through stable and meaningful relationships. Most models reify such relationships as network ties, and only allow agents who share a tie to exchange cultural knowledge.[6] Other models, such as the "constructural" one proposed by Carley (1991) and later extended by Mark (1998), do not explicitly model ties. Nevertheless, agents' likelihood of interacting, and therefore exchanging cultural information, is proportional to the knowledge they already share, effectively limiting such exchange to significant relationships.[7] Whether network ties are explicitly or implicitly modeled, culture does not diffuse between people who are not significantly acquainted with one another.

Though contagion models seldom provide explicit justification for this assumption (DiMaggio and Garip 2012), three explanations appear to be relevant. The first relates to the frequency and intensity of interpersonal interaction. Cursory and superficial interaction, it is argued, does not provide sufficient bandwidth for the exchange of cultural knowledge (Aral and Van Alstyne 2011). Second, the motivation for sharing information depends on tie strength. Individuals share their thoughts and intentions only with those with whom they have enduring relationships (Cowan 2014). Finally, susceptibility to cultural information also depends on tie strength. People are inclined to adopt practices they learn from others whom they trust and feel emotionally attached to (Miller and Prentice 2016; Centola 2011).

Together these mechanisms imply that people only learn culture through their significant and enduring relationships. Indeed, interaction depth is necessary when the cultural knowledge being shared is complex or costly. In the book club studied by Childress and Friedkin (2012), for example, club members engage in lengthy and animated discussions of their book evaluations. The intense and detailed debate affords them with the opportunity to influence each others' opinions.

But cultural information can be simple and easily transmittable. Every cultural exchange is, in essence, an exchange of symbolic representations (Berger and Luckmann 1967; Jablonka and Lamb 2006; Sperber 1996). Symbols are powerful and efficient tools of communication precisely because they parsimoniously convey complex and nuanced concepts. A three-piece suit, for example, connotes very different information about its wearer than does a pair of jeans. Though a great deal of symbolic information is communicated nonverbally, language is the dominant medium through which it is exchanged. When a speaker says she is a liberal, or that she listens to country music, the recipient of this information understands the speaker's intentions without the former having to explain what liberalism or country music are.

Because it is easily transmittable, the propagation of cultural information does not necessitate a long-lasting relationship or a meaningful discussion of attitudes and motivations. In fact, in most everyday settings we observe the symbolic behaviors of others—be them complete strangers or individuals with whom we have established relationships—without having unobstructed access to their underlying thoughts and intentions. We hear a co-worker mention that his child is not vaccinated; we see a mother in daycare nursing her child when picking up our own; we observe a classmate, a service provider or a fellow passenger on the train wearing a shirt professing her favorite band or political affiliation. Because it is observable and parsimonious, this information registers even in the absence of intentional sharing or interactive bandwidth.[8]

Moreover, cultural information does not need to be transmitted through strong ties to influence its receiver into action. A voluminous literature in psychology and sociology demonstrates that humans are innately attuned to the informational and normative cues in others' behaviors, even when interaction is transient or superficial (Cialdini and Goldstein 2004; Miller and Prentice 2016). Experiments by evolutionary psychologists, for example, show that children imitate unfamiliar adults' behaviors both as a means to reduce uncertainty (Lyons, Young, and Keil 2007) and out of concern for normative compliance (Kenward 2012). This sensitivity to others' behaviors generalizes to a variety of contexts. In Salganik and colleague's (2006) *Music Lab* experiment, for example, participants were randomly assigned into parallel artificial music markets and asked to download unfamiliar songs. Exposure to previous participants' choices dramatically influenced new participants' music consumption patterns, gradually amplifying minor initial differences in appeal between songs into large differences in popularity. In a similar experiment, Willer, Kuwabara, and Macy (2009) found that subjects were influenced by others to change their ratings of wines, and consequently enforced these adjusted opinions on others. Importantly, in both experiments subjects' behaviors were affected by others' despite the absence of prior familiarity, affinity or direct interaction between them.

The nature of the interpersonal relationship through which a cultural practice is observed becomes consequential for adoption only when the behavior it entails carries significant risk. In such instances, observers are more likely to be influenced by peers they know and trust. In a field experiment conducted by Paluck and Shepherd (2012), for example, a random intervention was designed to estimate peer effects on bullying in a high school. Risky behavior, such as defending harassed students, only diffused through strong ties. But exposure to others' declining bullying behavior was enough to reduce students' likelihood of engaging in bullying themselves, irrespective of whether the peers they observed were friends or mere classmates. In other words, students were responsive to the prevalence of bullying behavior they were exposed to, and adjusted their own behavior accordingly, irrespective of interaction valence or intensity.

A substantial proportion of cultural transmission, we argue, occurs through such transient observation of behaviors. This does not mean that durable relationships of the kind that are assumed in network diffusion studies are inconsequential. But if social influence can and often does operate through superficial interaction—if, in other words, culture is "contagious"—then why do easily transmittable cultural practices diffuse differentially when there are no barriers to observing each others' behaviors? Consider the adolescent *lads* in Willis' (1977) ethnography of a 1970s West Midlands school, who denigrate *ear'oles* for their compliance with behavioral expectations set by teachers. Though the *lads* intentionally smoke at the school gate in order to be seen by other pupils, smoking does not diffuse throughout the student population; rather, it becomes a strong marker of being a *lad*. Differentiated public displays of music consumption and dress similarly mark *decent* and *ghetto* social identities in the American inner-city. These cultural divisions endure despite being crosscut by an abundance of opportunities for interaction and mutual observation (Anderson 1999).

A contagion model cannot explain the emergence and endurance of cultural differentiation—whether in *Hammertown Boys School* or on the streets of Philadelphia—unless it assumes a preexisting and insular division into groups. Smoking is easily observable; if schoolboys merely adopt the behaviors they see, then it should have diffused throughout the school. But the fact that it does not suggests that the boys somehow know which behaviors they should, and should not, imitate. As Willis demonstrates, *lads* do not join the school as such. Rather, they become *lads* through their interactive experiences. If that is the case, then how does smoking become associated with being a *lad*?

The Missing Link: Meaning

Meaning is conspicuously absent from these epidemiological explanations. Contagion models necessitate structural complexity—that is, they need to assume a segregated social network and interaction depth—because they normally conceive of cultural transmission as a simple and straightforward interpersonal process. These models essentially conceptualize cultural practices as indistinguishable bits of information that, like viruses transmitted between individuals, are relayed across a social network. The human relay stations that make up this network either block or retransmit the signals they receive. Whether an agent retransmits depends only on signal strength; signal content is regarded as immaterial for the agent's decision to adopt.

But content, and its meaning, are highly consequential for cultural diffusion (Hargadon and Douglas 2001). Though culture is a fraught analytical construct most sociologists agree that it fundamentally relates to meaning-making; culture is often defined as interpersonally shared subjective understandings (Patterson 2014). By "meaning-making" we refer to the interpretative process whereby an individual assigns an observed stimulus with a location in a cognitively represented semantic web (e.g. when the act of child vaccination is associated with the cognitively represented concept of "unnatural" or "healthy"). Cultural meanings, as we define them here, are a subset of cognitive interpretations that are constructed through an individual's social experiences. Consider smoking as an example. Though cigarettes, cigars and pipes provide similar physiological utilities, these different forms of tobacco consumption are commonly associated with distinct cultural meanings. Whereas cigars connote masculinity and power, pipes are conventionally associated with contemplation and old age. It is not surprising that cigarettes, with their rebellious connotations, were adopted by the defiant *lads* in Willis' ethnography of *Hammertown Boys School*.

The disregard for meaning in conventional diffusion models leads to two important shortcomings. First, these models do not take into account that the perceived value of adopting a cultural practice is dependent on how this practice is interpreted. Residents of the Peruvian village of Las Molinas, for example, were resistant to a mid-century water boiling health campaign because they perceived hot water as something only appropriate for the sick (Rogers 2010). The decision to adopt a cultural practice is also implicitly—though often unselfconsciously—a decision about the cultural meaning being signaled to others. A Hammertown schoolboy's decision to adopt smoking is not merely related to the utility gained by inhaling nicotine; it is also an act of defiance.

A second shortcoming relates to how individuals infer meaning. Virtually all diffusion models treat adoption as a discrete event. These models conventionally represent culture as vectors of independent preferences. Social influence is modeled as the effect of one agent's behavior on another agent's cultural preference in isolation of other preferences. But cultural practices are not meaningful in and of themselves. Rather, meaning is a property of their relationship with other cultural elements. Phillips (2013), for example, demonstrates that the diffusion of various Jazz recordings in the 1920s was highly contingent on the narratives related to the conditions of their creation. German Jazz, he argues, failed to achieve popularity because of an incongruence between the meanings popularly associated with Jazz music and those associated with Berlin musicians.

The simple memetic imagery informing conventional contagion models does not account for this semantic complexity. It assumes that exposure to a cultural practice uniformly translates into adoptive behavior. But if an individual's propensity to retransmit a cultural practice is conditional on how that practice is interpreted, individuals who are exposed to the same information might still behave differently. One schoolboy might see smoking as a cool symbol of youthful rebelliousness, whereas another might predominantly associate it with masculinity. The two boys may reach different conclusions about the appeal of smoking. Even if the two pupils similarly interpret smoking as a form of anti-establishment behavior, they might still feel differently as to whether such behavior is desirable; one might be inclined to act in defiance of the teachers, while the other might not.

Moreover, if meanings are inferred from relationships between cultural practices then the diffusion of a cultural practice is dependent on the distribution of other practices in the population. These interdependencies are not derived from objective functional or logical relationships. Rather, they are an emergent product of context. Kaufman and Patterson (2005), for example, demonstrate that cricket experienced differentiated degrees of diffusion into British ex-colonies due to variation in the social conditions across receiving countries and the local cultural meanings these conditions gave rise to. In some cases cricket was popularly adopted (or rejected) because it connoted Britishness, whereas in others because it afforded the opportunity to resist British dominance.

Meaning decouples exposure to a cultural practice from the decision to adopt it. This decoupling, we contend, facilitates differentiated adoption of cultural practices even in the absence of structural barriers for diffusion. The differentiated diffusion of smoking in Hammertown Boys School is dependent on the cultural meanings associated with this practice which, in turn, is driven by the distribution of other cultural practices (and their meanings) among the student body. The cultural meaning of smoking emerges through the process of its diffusion.

From Contagious to Associative Diffusion

A Theory of Associative Diffusion

We have so far used the term practice, in a very broad sense, to denote a cultural element that is diffusing. But, as we have alluded to earlier, there is a difference between a behavior that is being enacted and the underlying cognition motivating it. Conventional models of social contagion generally disregard this distinction. They typically model cultural preferences as binary variables, and assume that social influence occurs when one agent simply adopts a practice she is exposed to.[9] We illustrate this conceptualization in Panel A of Figure 1. Agent A is performing a practice: she is smoking. Agent B, who is not a smoker, observes agent A and adopts the practice. Consequently agent B himself smokes.[10]

In contrast, our theory of cultural diffusion distinguishes behavior from cognition. We assume that agents have preferences for practices, which are operationalized as continuous variables ranging from negative to positive values, and that these preferences affect the likelihood that an agent will enact a given practice at a given moment in time. Agents observe each others' behaviors, but do not have direct knowledge of the preferences producing them. They can only infer their interlocutor's motivations. To simplify our model, we assume that preferences and behaviors correspond directly to one another, that is, that a preference and its corresponding behavior are, respectively, the private and public representations of the same object. In other words, all agents cognitively represent the same set of concepts, and agree what behavior each entails.[11]

While agents agree on the set of possible cultural practices, they might have different interpretations. Our model therefore assumes a two-stage process of diffusion. In the first stage an agent *interprets* another agent's behavior and in the second stage the agent *evaluates* the behavior (Trope and Liberman 2010; Goldberg 2011). These two cognitive mechanisms—interpretation and evaluation—together affect the agent's propensity to reenact the behavior she observes. Panel B of Figure 1 schematically illustrates this two-stage process of interpersonal transmission. Agent B observes agent A smoking. First, he updates his interpretation of smoking. We represent interpretation as the location of smoking in a semantic network. Second, the agent evaluates smoking by updating his preference for it. Finally, his probability of enacting smoking is proportional to his preference.[12]

—— Figure 1 about here ——

What do interpretation and evaluation entail, and how do they affect cultural transmission? In developing our two-stage model we draw on two established findings in cognitive science: semantic cognition and constraint satisfaction.

Interpretation: Semantic Cognition

The first relates to the cognitive underpinnings of interpretation. Theories of cognition generally agree that semantic knowledge is cognitively represented as a system of interdependencies between concepts, and that concepts are meaningful by virtue of their relationships of entailment and opposition with other concepts (D'Andrade 1995; Murphy 2004; Patterson, Nestor, and Rogers 2007; Jablonka and Lamb 2006). Semantic context affects how new information is interpreted (Moreau, Markman, and Lehmann 2001). The mere presence of male stereotypical objects such as *Star Trek* posters in a classroom, for example, are enough to suppress female undergraduates' interest in computer science. These cultural cues lead female students to construe computer science as a masculine—and therefore unappealing—academic field (Cheryan, Davies, Plaut, and Steele 2009).

We make two assumptions about agents' semantic cognition. First, we assume that people cognitively represent semantic knowledge as a matrix of associations between concepts. Interpretation is the process of assigning a stimulus with a location in this semantic web.[13] Second, we assume that they impute these associations by observing co-occurrences between cultural practices in others' behaviors. Research in cognition provides strong evidence that individuals learn associations from one another. Chain transmission experiments, for example, find that humans are biased to impute associations in others' behaviors even when such associations are merely random noise (Griffiths, Kalish, and Lewandowsky 2008; Kirby, Cornish, and Smith 2008). Interpretation is therefore interpersonally transmitted when an agent updates her cognitively presented associations when observing others' behaviors. A pupil notes that the *lads* smoking at the school gates are also wearing high-platform shoes, engaging in physical violence and generally 'having a laff' at the *ear'ole's* academic aspirations. Together, these behaviors form a gestalt that connotes resistance to the school's establishment, and the middle-class ideals it represents.

Evaluation: Constraint Satisfaction

A second finding in cognitive science that we build on, and that informs the evaluation phase of our two-stage model, relates to how individuals form preferences. We assume that people adapt their behavioral preferences to cohere with the associative patterns they perceive, namely, that they seek to form equal preferences for practices which they perceive to be positively associated with one another. We base this assumption on psychological research on *constraint satisfaction*, which has its roots in the well-established finding that humans have a psychological need for resolving cognitive dissonance (Shultz and Lepper 1996; Kunda and Thagard 1996). Constraint satisfaction is the connectionist conceptualization of cognitive consistency. It can be thought of as a process of balancing the activation of nodes connected by excitatory and inhibitory links in a neural-like network of relationships. Such a balancing would lead to activation of positively related nodes, and suppression of negatively related ones. As a variety of studies demonstrate, constraint satisfaction models provide compelling explanations for a variety of otherwise difficult to reconcile experimental findings on how people form impressions, make stereotypical attributions and are affected by priming (Kunda and Thagard 1996; Schröder and Thagard 2014; Freeman and Ambady 2011).

Two implications for cultural diffusion follow. The first is that preferences depend on other semantically related preferences and, consequently, that agents adapt their preferences when their observations of others' behaviors lead them to update semantic links. Indeed experimental evidence suggests that people adjust their preference to cohere with the information they observe and the choices they make (Holyoak and Simon 1999; Simon, Krawczyk, and Holyoak 2004). Second, people behave in ways that satisfy the semantic constraint that they observe. Participants in a story-retelling chain transmission experiment, for example, gradually eliminate information that is culturally incongruent (Hunzaker 2016). This tendency for cognitive consistency appears to explain macro behavioral trends as well. Entrenched beliefs about the the incompatibility between sexual activity and school attendance that are prevalent in Malawi, for example, induce girls to drop out of school despite there being no evidence that sexual activity undermines school success (Frye 2017).

Associative Diffusion

Taken together, we argue, semantic cognition and constraint satisfaction produce a self-reinforcing process wherein agents enact the associations they observe. Agents (1) impute a cultural order of interdependencies between practices by observing co-occurrences between them in others' behaviors and (2) adapt their behavioral preferences and consequent behaviors in a manner consistent with this order. We refer to this process as *associative diffusion.*

As illustration, imagine a schoolboy who, having observed the behaviors of his peers, perceives smoking and physical aggression to be positively associated with one another, and negatively associated with wearing school uniform and being studious. Imagine further that the pupil had also inferred from his peers' behaviors that the two latter practices are associated with one another. Constraint satisfaction would entail adopting preferences that balance these relationships: either being inclined to smoke and partake in physical violence, or to wear uniform and being studious. Note that the source of constraint is psychological, not ontological. Nothing about smoking makes it inherently more congruent with violence than with studiousness. Rather, by observing others the schoolboy has learned that smoking and being studious are socially incompatible.[14] Furthermore, by enacting perceptually congruent behaviors he reproduces this cultural order for other schoolboys to observe.

We argue that associative diffusion extends and improves on existing sociological literature in two important ways: it explains how cultural order emerges and why this emergence results in cultural differentiation. Recent work by cultural sociologist has paid increasing attention to semantic interdependencies between preferences and beliefs (e.g. Goldberg 2011) and to their behavioral implications through constraint satisfaction (e.g. Schröder, Hoey, and Rogers 2016). But these studies take the structure of interdependencies as a given, assuming that people are enacting a preordained cultural order. Work by psychologists has similarly explored how constraint satisfaction explains preferences and behaviors, but not how the cognitive associations that produce constraint are learned and adapted (but see Ehret, Monroe, and Read (2015) for an exception). These approaches can explain how cultural order is reproduced, but not why and how it emerges. Associative diffusion, in contrast, models cultural learning and should therefore provide an explanation for the emergence of interpretative consensus.

By paying attention to the cognitive underpinnings of cultural learning, our model of associative diffusion also departs from traditional theories of network diffusion. Conventional contagion models describe a mechanism of *social proof*, where agents seek affirmation for their own decisions, beliefs and assumptions in others' aggregate behaviors (Cialdini 2007). In contrast, our model describes a process of *social construction* (Berger and Luckmann 1967). People do not simply mimic others' behaviors; rather, they learn, by observing others, what these behaviors mean. In other words, we argue that interpretation coheres as behaviors diffuse through a population. People coordinate their interpretations by learning from one another which behaviors are compatible with each other. Meaning is implicitly communicated between individuals by affecting their perceived associations, leading to the emergence of interpretative consensus.

While contagion models would typically predict that in the absence of structural boundaries to diffusion actors will harmonize their cultural preferences, we expect interpersonal associative coordination to result in the emergence of cultural differentiation. We base this expectation on the insight, recently promoted by cultural sociologists, that interpretative consensus does not imply evaluative agreement. People who share the same cognitive association might still reach different evaluative conclusions. Free-market ideologues and anti-consumerists, for example, both agree that capitalism is driven by self-interest but disagree whether it is desirable or destructive (DiMaggio and Goldberg 2018). The mutually constraining dynamics of semantic connections accentuate evaluoatory divergence when there is interpretative consensus. Indeed, recent work exploring the properties of constraint satisfaction finds that when a connectionist model is allowed to learn (that is, it updates associative links in response to stimuli), preferences become increasingly entrenched and polarized (Monroe and Read 2008). The self-reinforcing dynamics of associative diffusion, we conjecture, should therefore lead to gradual differentiation in preferences and concomitant behaviors. Overall, we argue that:

Main Proposition*: Associative diffusion leads to the emergence of cultural differentiation even when agents have unobstructed opportunity to observe one another. Social contagion does not lead to cultural differentiation unless agents are structurally segregated.*

Model

To test this proposition, we implement the following agent-based model:

1. An actor and observer are randomly chosen to interact.
2. The actor enacts two practices (based on her preferences).
3. The observer updates his perception of associations between these practices.
4. The observer changes his preferences for one of these practices only if that change leads to an increase in constraint satisfaction.

We provide an overview of the model in Table 1 and discuss these steps in detail below.

Fundamentals

Let K be a finite and fixed set of cultural practices, and let there be N individual agents. Each agent is represented by two data structures:

1. Matrix R of size KxK corresponds to the agent's cognitive representation of associations. The value of each element $R_{ij} \in [0, \infty)$ represents the strength of the perceived association between practices i and j. R is initialized to $R_{ij} = 1 \; \forall i, j \in K$, such that practices are initially perceived by agents to be equally associated with one another.

2. The agent's behavioral preferences are represented as a vector, $V = (v_1, v_2, ...v_k)$ of length K, where $v_i \in [-\infty, \infty]$. V is initialized with random values drawn from a uniform distribution.

In each iteration t we randomly draw two interacting agents, A and B, from the population. We refer to them as the performer and observer, respectively. We make three assumptions about the nature of interaction.

First, we assume no existing network structure that affects the opportunity for interaction. All agents are therefore equally likely to interact with one another. Second, we assume that A only exhibits a subset of behaviors at each interaction. In other words, agents do not know their interaction partner's location in social space. They can only infer that location on the basis of the behaviors being displayed. For the sake of simplicity, we assume that B observes A performing exactly two behaviors, which we refer to as b_1 and b_2. Finally, we assume that interactions are anonymous, that is, that agents do not remember information about other agents even if they had interacted with them before. Our model can, in theory, be extended to account for memory in repeat interactions, such that B infers associations from A's recent m behaviors. Our simple setup is analogous to such memorable repeated interaction where B only remembers A's two most recent behaviors, i.e. where $m = 2$.

Agent A's likelihood of exhibiting behavior i is proportional to v_i. We use a variant of Luce's (2005) choice axiom such that:

$$P(i) = \frac{e^{v_i}}{\sum_{j=1}^{K} e^{v_j}} \tag{1}$$

Updating

Whenever agent B observes a co-occurrence between practices i and j, the agent increases the association between them such that $R_{ij} = R_{ij} + 1$. Associations in R decay at a rate $0 < \lambda < 1$. Thus, associations that are not reinforced through repeated observed co-occurrence asymptotically decrease toward zero. Upon observing a co-occurrence agent B also updates one of the preferences corresponding to the two co-occurring behaviors. Drawing on literature on attitude strength and cognitive dissonance (e.g. Petty and Krosnick 1995), we assume that the weaker of the two preferences, defined as the one whose absolute distance from the mean preference is smaller, is randomly updated with $\Delta v \sim N(0,1)$.[15]

Agent B retains the preference update if and only if it satisfies the constraint imposed by the associations represented in R; otherwise, no preference updating occurs. To calculate constraint satisfaction we need to measure the concordance between vector V and matrix R. We assume that an agent's preferences satisfy the constraint imposed by her associative perceptions if she exhibits similar degrees of preference for practices that are associated with one another (Simon et al. 2004). We therefore evaluate the concordance between V and R by computing the differences between all pairs of preferences comprising V, and comparing them to their corresponding elements in R. Constraint satisfaction increases as the difference in preferences between two practices i and j, for whom the association R_{ij} is strong, decreases.

To do so we transform V into a KxK sized distance matrix Ω that represents the similarity between the agent's preferences. Each element $\Omega_{ij} = |v_i - v_j|$ corresponds to the absolute difference between v_i and v_j. We standardize Ω by its maximal value such that $\Omega_{ij} = 0$ if the agent's preferences for i and j are identical, and nears 1 as they diverge. Similarly, we standardize R by its maximal value such that its elements range from 0 (corresponding to no perceived association between practices) to 1 (corresponding to maximal perceived association). *Constraint Satisfaction* is defined as:

$$CS(V,R) = \tfrac{K}{K(K-1)} \sum_{i=1}^{K} \sum_{j=1}^{K} |R_{ij} - \Omega_{ij}| \qquad (2)$$

The term $|R_{ij} - \Omega_{ij}|$ in eq. [2] nears 1 as the distance between the agent's preferences for i and j becomes inversely proportional to the perceived strength of their association. Ranging from 0 to 1, as $CS(V,R)$ nears 1, V is said to perfectly satisfy the constraint defined by R. Agent B retains the preference update only if $CS(V,R)$ increases.

Figure 2 provides an illustration of constraint satisfaction. In this hypothetical example, the agent's perceived associations (matrix R) exhibit two clusters of practices that are strongly associated within-cluster and dissociated between-cluster. Only one practice, labeled d, is weakly associated with practices outside its cluster. The two preference vectors with highest constraint satisfaction (labeled a & d in panel 3 of Figure 2) are those in which the agent has an equally strong preference for practices in one cluster, and a dislike for practices in the other. Constraint satisfaction decreases when the agent's preferences are similar for practices that are perceived to be dissociated.

—— Figure 2 about here ——

Constraint satisfaction, as we implement it here, is analogous to a connectionist cognitive process whereby an agent updates a preference only if by doing so this preference becomes more compatible with preferences for other practices with which the focal practice is strongly associated. Note that we assume that the agent does not fully satisfy constraint. Rather, consistent with research that demonstrates that people can tolerate cognitive inconsistency by compartmentalizing cognitive dissonance, we assume that only the dissonance made salient to B by A's behavior is being resolved. That is, B only updates the weaker of the preferences instantiated by A's behaviors. Other preferences remain unchanged.

The implications are illustrated in Figure 3, which provides a summary of the agent-based model. Imagine that A and B are two co-workers. B observes A exhibiting two practices—for illustration, imagine that A is mentioning that she vaccinates her children and is eating organic food. B has opposing preferences for these two practices: he is pro-vaccination, but has a dislike for organic food. Having observed his co-worker exhibiting both, his perceived association between these two practices increases. To accommodate this perceived increase B would need to decrease his dislike for organic food, so as to make his preferences for vaccination and organic food more compatible. Such an update, however, would be at odds with his strong perceived association between organic food, biking and hiking (practices e and f respectively), and is therefore rejected; the structure of associations thus constrains B's preferences. In other words, whether or not B updates his preferences is not merely a product of the behaviors A exhibits. Rather, it is constrained by the overall set of perceived associations that B had cumulatively inferred from his observational experiences.

—— Figure 3 about here ——

Measurement

We develop several measures to assess convergence and dissimilarity between agents. Drawing on our analytical framework, we distinguish between cognitive (relating to information only available to the agent) and behavioral (relating to information available to all agents) dimensions of convergence.

Cognitive Agreement

On the cognitive dimension we distinguish between interpretative and evaluatory agreement between agents. We measure ***interpretative agreement*** between agents by comparing the similarity in their perceptions about which practices are associated with one another. *Interpretative distance* between two agents is defined as the distance between their respective association matrices R and R^*. This distance is calculated as the pairwise absolute difference between all corresponding cells in the two matrices:

$$\|R, R^*\| = \frac{1}{K^2} \sum_{k=1}^{K} \sum_{l=1}^{K} |\tilde{R}_{kl} - \tilde{R^*}_{kl}| \tag{3}$$

where $\tilde{R} = R/max(R)$. Interpretative distance at the group level is defined as the mean interpretative distance between all pairs of agents:

$$\langle \|R, R^*\| \rangle = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \|R_i, R_j\| \tag{4}$$

As $\langle \|R, R^*\| \rangle$ decreases, the agents comprising the population increase their interpretative agreement; they perceive the world through the same associative lens.

Evaluatory agreement relates to agents' preferences: agents who evaluate practices similarly also have similar preferences. We distinguish between preference similarity and congruence. *Preference similarity*, measured as the correlation between two agents' preference vectors, quantifies the extent to which the two agents value the same practices. *Preference congruence*, in contrast, measures the extent to which agents' preferences follow the same pattern. We measure preference congruence as the absolute correlation between the two agents' preference vectors. Preference congruence quantifies the extent to which the two agents tend to like, or dislike, the same practices. We define group-level preference similarity as the mean correlation between all pairs of agents' preference vectors:

$$\langle \rho(V, V^*) \rangle = \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} \rho(V_i, V_j) \tag{5}$$

and group-level preference congruence as the mean absolute correlation between all pairs of agents' preference vectors:

$$\langle |\rho(V, V^*)| \rangle = \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} |\rho(V_i, V_j)| \tag{6}$$

Behavioral Agreement

On the behavioral dimension, we use mutual information to measure convergence in agents' behaviors. The mutual information between two variables X and Y measures the extent to which one variable predicts the other. It is calculated as follows:

$$I(X, Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \tag{7}$$

where $p(x)$ is the marginal probability of behavior x and $p(x, y)$ is the joint probability of behaviors x and y.

We apply mutual information to agents' behaviors such that $X = b_1$ and $Y = b_2$ (see Appendix A for more details). This allows us to measure the extent to which observing a random agent performing one behavior provides information about what her other behavior is likely to be. We interpret behavioral predictability as an indication that behaviors are becoming more *meaningful*: it is enough to observe an agent enacting one practice to make a reliable inference about her preferences for other practices. Imagine that the agents are schoolboys. As mutual information increases, seeing a schoolboy smoke also indicates that he is likely to wear platform shoes but unlikely to be studious. His smoking behavior implies an emergent identity as a *lad*. Mutual information, in other words, measures the extent to which behaviors are mutually implicated. It evaluates the strength of relationships between behaviors that an observer can infer from others' behaviors.

Results

Two Agent Model

We begin by restricting the model to two agents. These agents alternate between roles, such that at time t agent A is the performer and B the observer, and at time $t+1$ the agents swap roles. Our purpose at this stage is modest: to explore the dyadic dynamics of an interaction model in which two agents infer associations exclusively from the other's behaviors. Though unrealistically simplistic, the purpose of this setup is to understand the implications of our model at the interpersonal level, before aggregating them to the group level. In particular, we seek to explore whether the mutual observation of behaviors will lead the two agents to reach similar or opposing preferences. If associative diffusion leads to cultural differentiation, as we argue, then it should induce agents to adopt aligned or opposed preferences.

We generate 1,000 simulations between two agents with $K = 6$. Results are plotted in Figure 4. Panel A plots the final correlation between the two agents' preference vectors, V_A and V_B, as a function of their initial correlation, for 1,000 simulation runs. As it illustrates, the two agents' preferences either gradually converge or diverge. In other words, as the two agents interact they either adopt the same or the opposite preferences.

—— Figure 4 about here ——

This tipping toward either convergence or divergence is reflected in the gradual increase in the absolute correlation between the two agents' preference vectors, as plotted in Panel B. Whether the two agents are in agreement or opposition, their preferences become increasingly congruent. By observing each other's behaviors and updating their association matrices accordingly, the two agents gradually coordinate which behaviors are compatible with one another. Importantly, whether the agents adopt identical or opposing preferences, their behaviors become increasingly predictable: as Panel B illustrates, mutual information gradually increases. As time progresses, by observing an agent's discrete behavior we can increasingly predict which other practices she is likely, or unlikely, to enact. Thus behaviors become increasingly meaningful.

Why does interaction lead two agents toward agreement or differentiation? Imagine an observing an interlocutor who eats organic food and is an avid hiker. You learn from this conversation partner that the two are associated. If you are a hiking enthusiast yourself this information is likely to positively affect your evaluation of organic food. The opposite is also true, however; this interaction would negatively impact your evaluation of organic food if you were initially negatively disposed toward hiking. Experimental work demonstrates that, when presented with new information, people are inclined to strengthen their prior attitudes (Taber and Lodge 2006). Constraint satisfaction accentuates this tendency such that preferences spill over between practices that are perceived as related. Because the two agents are increasingly coordinating their perceptions of semantic relationships by observing each other's behaviors, this cross-preference alignment gradually leads them toward agreement or disagreement. If two agents agree that organic food and hiking are congruent, they might either espouse or reject both practices.

Importantly, as Panel A illustrates, the initial correlation between the agents' preferences is not predictive of the final correlation between them, except at the extremes (when agents are randomly initialized to have a strong positive or strong negative correlation). The two-agent simulation, in other words, does not simply intensify randomly assigned initial similarities or differences between agents. Rather, interpretative coordination and constraint satisfaction together lead agents either toward preference convergence or divergence. As we explore in more detail in Appendix B, whether agents converge or diverge is path-dependent, relating to stochastic decisions the agents make (i.e. preference update magnitude and direction).

Multiagent Model

If two interacting agents' preferences either converge or diverge, what would the dynamics be when more than two agents are interacting? We expect these mutually reinforcing and negating tendencies to lead to the emergence of a steady equilibrium of cultural differentiation. If, as we saw with the two-agent model, associative diffusion leads interacting agents toward agreement or opposition with equal probability, we expect a group of interacting agents to gradually sort into different cultural groups. To test this prediction, we conducted additional agent-based simulations with groups comprising $N = 30$ agents. We assumed no a-priori or emergent network structure constraining interaction between agents. At each modeling iteration two agents are randomly sampled from the group with equal probability, and are randomly assigned to either the performer or observer role.

Figure 5 summarizes the results of 1,000 such simulations, with $t = 100,000$. The three measures plotted in Panels A to C—preference congruence, interpretative distance and mutual information—indicate that the dynamics we saw at the interpersonal level aggregate into a group-level equilibrium. Two patterns are particularly informative. First, we see that preference similarity between agents remains steadily at 0 (Panel A, inset), indicating that agents do not adopt the same preferences. At the same time, their preferences become perfectly congruent: they gradually diverge toward opposing preferences, as indicated by the increase in the absolute correlation between their preferences. This patterned divergence leads to practices becoming more meaningful, as manifest in the gradual increase in mutual information (Panel B). As the agents interact, observing them perform one behavior provides increasing information about the subset of other behaviors they are likely to enact.

—— Figure 5 about here ——

Second, we see that interpretative distance between agents declines: they gradually come to perceive cultural order through similar associative lenses (Panel C). Though the agents do not adopt the same practices, they reach an *interpretative consensus*. They agree which practices go with one another, not which ones are preferable. Such interpretative agreement and evaluatory disagreement can result in a steady equilibrium only if agents' perceived associations cluster practices into densely associated subsets, and if different agents adopt different clusters of practices.

To see why this is the case, consider Figure 2 again. Imagine two agents who both share identical associative perceptions, as represented in the network illustrated in Panel 2, but who have different preference vectors (e.g. those labeled a & d in Panel 3). The clustered structure of the association network is what allows both agents to adopt different behaviors, but still be at identical levels of constraint satisfaction. In fact, it is precisely this clustering that makes the practices meaningful: if all pairs of practices were equally associated with one another, constraint satisfying agents would have had an equal likelihood of performing either practice. Such a pattern of co-occurrence would have 0 information value compared to randomly chosen behaviors. In contrast, the clustered associative pattern effectively partitions the set of practices into different implicit categories, each adopted by a different agent. We should therefore expect the group as a whole to gradually partition into subgroups of agents whose preferences correspond to the emergent clusters of practices. If the agents gradually converging on the associative structure depicted in Figure 2 were Hammertwom

Boys School pupils, then the clustering of various practices—smoking and 'having a laff,' uniform wearing and studiousness—would have implicitly designated different student as *lads* and *ear'oles*.

We evaluate the extent to which agents are clustered into subgroups of similar preferences. To do so, we use the K-means algorithm to partition the agents' preference vectors into clusters of similar preferences, and the gap statistic (Tibshirani, Walther, and Hastie 2001) to evaluate the optimal number of agent clusters (see Appendix A for more details). One advantage of using the gap statistic is that the method can estimate when there does not exist an optimal partition (namely, when the number of clusters equals one). We plot the mean number of clusters estimated by this procedure in Figure 5, Panel D.

Two patterns are immediately apparent. The first is that as the agents reach a stable interpretative consensus (when the curves in Panels A-C of Figure 5 plateau) they cluster into roughly two stable preference groups. The two-agent dynamic, which results either in convergence or divergence, aggregates into a group dynamic of differentiation. The second is that this period of stability is preceded by a period of turbulence and interpretative ambiguity, whereby the mean number of clusters rises from no clustering to a peak of 4. We observe a high variance in the peak number of clusters across simulations during this turbulent period, reaching upwards of 10 at the extremes. This dynamic corresponds to a complex social process whereby a set of tenuous preference clusters is gradually subsumed by an emergent division into two subsets of preferences.

An example of this gradual convergence is illustrated in Panel E of Figure 5, which plots a few snapshots from one random simulation run. Each panel in the figure depicts the agents' preference vectors at a different time. Columns correspond to the six practices, and rows correspond to individual agents. Preferences are color coded, ranging from strong negative (blue) to strong positive (yellow). As these snapshots illustrate, the group as a whole slowly partitions into two crisply bounded subgroups.

Alternative Explanations

The two- and multi-agent simulations demonstrate that as individuals coordinate their interpretations they also gradually divide into groups with opposing preferences. Earlier we had proposed that conventional social contagion models cannot explain the emergence of this kind of cultural differentiation unless they assume a segregated social structure that prevents groups from fully interacting with one another. We test this proposition in this section by considering two sets of alternatives to associative diffusion. In the first set we explore alternative contagion mechanisms wherein agents imitate their interlocutors with bias, either toward their prior preferences or toward the prevalent behaviors in the population. We demonstrate that neither leads to differentiation. In the second we consider what happens when contagion is conditional on homophily between performer and observer. We explore several network topologies and demonstrate that cultural variation emerges only when the network of interactions is segregated. In investigating both sets of alternatives we also explore their integration with our model of associative diffusion. We show that associative diffusion leads to cultural differentiation when agents are sensitive to practices' popularities and under different network configurations.

Contagion

Social contagion models generally assume that when two agents interact one agent adopts the other's preference. Let the two agents be, once again, A (performer) and B (observer), and let the preference in question be i. Essentially all contagion models assume the following adoption process:

$$V_{Bi}(t+1) = f(V_{Ai}(t)) \tag{8}$$

where $f(\cdot)$ is a function of actor A's preference. An overwhelming majority of contagion models assume ***naive contagion***, where B simply adopts A's preference (that is, $f(\cdot)$ in eq. 8 is the identity function). Such a simple contagion process obviously cannot produce cultural differentiation on its own. If agents simply imitate one another, and if there are no constraints on who they observe, they should gradually converge toward the same preferences. We therefore consider two additional contagion mechanisms that previous research suggests are prevalent, and can plausibly lead to cultural clustering.

The first, which we refer to as ***biased contagion***, relies on evidence from social psychology that people are motivated to adopt information that confirms, and reject information that disconfirms, their prior beliefs (Kunda 1990).[16] Experimental work demonstrates that this process pushes individuals toward extreme opinions, gradually leading to polarization (Lord, Ross, and Lepper 1979). We follow Dandekar et al. (2013) and operationalize biased contagion as a function of a bias parameter β such that $f \sim V_{Bi}^{\beta} \cdot V_{Ai}(t)$ in eq. 8. In other words, B's likelihood of adopting A's preference is moderated by her own preference for practice i. As long as $\beta > 1$, adoption is positively biased such that B becomes more likely to adopt A's preference as her prior preference for i increases (see Appendix C for more details). Intuitively, biased contagion should lead agents to be differentially influenced by their peers as a function of their prior preferences. Minor initial differences between agents might gradually compound toward polar differences. Biased contagion is therefore a plausible candidate for a contagion process that leads to cultural differentiation.

A second mechanism that might lead to cultural differentiation is ***conformist contagion***. By conformity we mean the tendency to preferentially adopt practices that are prevalent in a population. Though research demonstrates that people are universally disposed toward conformist behavior (Cialdini and Goldstein 2004) and that conformist learning is adaptive (Henrich and Boyd 1998), it also finds that individuals derive psychological utility from uniqueness (Snyder and Fromkin 1980; Chan, Berger, and Van Boven 2012). While most people resolve this tension by conforming to group norms, some are more likely than others to adopt counter-normative behaviors. Thus different people have different tastes for popularity (Lieberson 2000; Zuckerman 2012). The existence of nonconformists might lead to cultural differentiation when, for example, early adopters have different preferences than mainstream consumers (Moore 2006) or when avant-garde audiences exhibit unorthodox cultural tastes (Bourdieu 1993).[17]

We define an agent's level of conformity, $\omega \in [0, 1]$, as her degree of preference for popular practices; agents with $\omega = 1$ are conformists and those with $\omega = 0$ are nonconformists. We assume that agents update their perceptions of practice popularities by observing how much others perform them. A practice's perceived uniqueness, which we label ψ_i, is the inverse of this popularity (see Appendix C for details). When agent B observes agent A performing practice i her likelihood of updating her own preference is dependent on B's perception of the practice's uniqueness and its congruence with B's degree of conformism. If A is smoking, for example, and B—who, let us assume, has a strong taste for popularity—rarely sees others smoke, then we would want B to be unlikely to adopt A's preference for smoking. To meet this criterion we moderate B's likelihood of adopting A's preference by the distance between her degree of conformity and her perception of the practice's uniqueness. Formally, we define $f \sim |\omega_B - \psi_i| \cdot V_{Ai}(t)$ in eq. 8. The greater the distance between ω and ψ_i—such as when agent B is non-conformist ($\omega_B \to 0$), and perceives a practice to be unique ($\psi_i \to 1$)—the greater the probability of adopting the practice performed by A (see Appendix C for more details on how conformist contagion is implemented). We assume conformism is more prevalent in a population than nonconformism.

To test whether either of these contagion mechanisms can lead, on its own, to the emergence of cultural clustering, we ran a set of simulations, again with $N = 30$ agents, $K = 6$ practices and $t = 100,000$ iterations, where agents randomly interact with one another. We summarize the results in Figure 6. Panel A plots the mean number of clusters at the end of the simulation, with its 95% confidence interval. Neither biased contagion nor conformist contagion lead to the emergence of a greater number of clusters than would be expected when agents naively imitate one another. In all cases the number of clusters effectively converges on 1, as all agents adopt the same preferences.[18] In other words, cultural differentiation does not emerge even when agents are biased toward their existing preferences or when the population includes a mix of conformists and nonconformists, as long as agents are allowed to interact freely with one another.

—— Figure 6 about here ——

That biased and conformist contagion do not, by themselves, lead to cultural differentiation does not mean that they are inconsistent with the two-stage transmission mechanism that undergirds associative diffusion. To demonstrate that, we modify our associative diffusion model to account for variation in agents' conformity. In this extension of the associative diffusion model an agent's probability of performing a practice is a function of a combination of two parameters: (1) the agent's preference for that practice, which is determined by constraint satisfaction as previously, and (2) the distance between the agent's conformity and perceived uniqueness of this practice, as is the case in conformist contagion. We extend eq. 1 such that the probability to enact a practice is:

$$P(i) = \frac{e^{v_i|\omega-\psi_i|}}{\sum_{j=1}^{K} e^{v_j|\omega-\psi_j|}} \tag{9}$$

In other words, agents continue to collectively produce the cultural order by updating their preferences in a manner that satisfies the constraint they observe, but they also differ in the extent to which they are likely to perform popular or rare practices.

Results from 1,000 multi-agent simulations with $t = 150,000$ and variation in agents' tastes for popularity are plotted in Panel B of Figure 6. Though the conformity model takes longer to converge than the baseline multi-agent model (wherein agents are insensitive to the prevalence of cultural practices, see Figure 5), we once again see a by-now familiar pattern: a period of interpretative ambiguity in which the number of cultural clusters peaks at four is followed by a convergence toward two stable preference groups. As agents sort into two cultural clusters, practices become more meaningful. This is the case even when nonconformists outnumber conformists. Panel C plots meaningfulness—as measured by mutual information—as a function of the proportion of nonconformists in the population (see Appendix C for details on how this proportion is determined). Meaningfulness declines as the proportion of conformists declines, but practices remain meaningful as long as the proportion of nonconformists is less than 80%. That is, as long as nonconformists do not constitute an overwhelming majority, cultural order evolves along the pattern we saw earlier, whereby a period of interpretative ambiguity characterized by a steep increase in the number of clusters is followed by a gradual decrease toward interpretative consensus and cultural meaningfulness. Only when a vast majority of agents seek to maximize their uniqueness by performing rare behaviors interpretative consensus fails to emerge.

Homophily

A second alternative to associative diffusion relies on the mechanism of homophily, or the susceptibility to influence by others who are perceived to be socially similar. Previous work demonstrates that homophilious contagion leads to preference divergence especially when agents are negatively influenced by others who are socially different (Mäs and Flache 2013). But these dynamics are often explored in conjunction with a segmented social network (e.g. Flache and Macy 2011). It is therefore not obvious whether differentiation emerges due to homophily or due to a preexisting clustered social network.

To evaluate whether homophily leads to cultural differentiation irrespective of network clustering we explore three network topologies: (1) *fully connected network*, wherein each agent has equal likelihood of interacting with any other agent; (2) *scale-free network*, wherein network in-degree follows a power law distribution such that a minority of agents have many incoming ties and the majority have few incoming ties; and (3) *small-world network*, where agents are clustered into fully connected cliques with a handful of ties crosscutting clusters, and where the distribution of node in-degree is constant (this particular implementation of a small-world network is often referred to as the “connected caveman” topology). We assume that agents only observe others whom they have directed ties with, and that they are equally likely to interact with tied alters.

Fully connected networks represent our default setting, where agents have no restrictions on who they can interact with. Scale-free and small-world networks have both been demonstrated to be prevalent in myriad settings, but they generally correspond to different types of social relationships. Scale-free networks are characteristic of superficial and impersonal interaction structures, such as followship relationships on Twitter or academic paper citation networks. Small-world networks, especially as we implement them here, characterize stronger and more durable ties, such as those that connect friends, family members and co-workers. See Appendix C for more details on how we generate these networks.

Let A and B be, once again, the actor and observer, respectively. To allow for homophily, and to generate equivalence with the associative diffusion model, we assume that agent A performs two behaviors at a time, labeled i and j. Unlike models that assume that agents are fully aware of each others' set of preferences, we assume, in accordance with our argument about the cultural conductivity of superficial ties, that others' preferences are only partially available to observers. We therefore allow agents to observe only one additional behavior.^{19} B updates her preference for i as a function of her perceived social similarity with agent A, which she infers from the distance between her and A's preference for j. We label this similarity η_{ABj}. To allow for homophilious contagion we define $f \sim \eta_{ABj} \cdot V_{Ai}(t)$ in eq. 8. Following DellaPosta et al. (2015), we assume that negative influence occurs, but is less likely than positive influence (see Appendix C for further details on how homophilous contagion is implemented).

Panel A of Figure 7 reports results from 1,500 simulations of homophilous contagion, with three different network topologies. The coefficients represent the estimated number of clusters, and its 95% confidence interval, after 100,000 rounds. As is clearly apparent, homophilous contagion with a fully connected or scale-free network topology does not lead to cultural clustering. Consistent with our main proposition, cultural differentiation emerges only when agents are constrained to interact within densely intra-connected and sparsely inter-connected cliques. The existence of a small-world network topology does not always lead to the emergence of cultural differentiation (the average number of clusters is 1.26), but it often does. In contrast, fully connected or scale-free networks almost never facilitate such cultural clustering. Homophilous contagion, in other words, does not on its own lead to the emergence of different cultural groups. It is the network structure that determines whether or not cultural clustering will emerge.

—— Figure 7 about here ——

How does network topology affect the process of associative diffusion? Our previous results, as reported in Figure 5, illustrate the dynamics of associative diffusion under a fully connected network topology. Panel B of Figure 7 reports the number of cultural clusters, as a function of time, when we simulate associative diffusion with scale-free or small-world network topologies. Both network topologies lead to the emergence of cultural differentiation. But this sorting into cultural groups evolves along different trajectories. Scale-free networks exhibit the same pattern that we saw when there were no limits on interaction: an initial increase in the number of clusters is followed by a gradual decline. Small-world networks, in contrast, exhibit a very different pattern of steady increase in the number of clusters, beyond a dichotomous division into two groups.

Panel C plots the mutual information between practices under the two network topologies. As the two lines clearly indicate, practices become significantly more meaningful when agents' interaction patterns follow a scale-free structure than they do when agents are embedded in a small-world network. In other words, when agents are assumed to learn associations from each other, rather than merely imitating one another, cultural differentiation emerges irrespective of network topology. But the cultural boundaries between clusters of practices are crisper when culture diffuses through scale-free networks than when the population is divided into tightly-knit cliques. Network structure has an impact on the nature of cultural differentiation and the process through which it unfolds.

To summarize, our exploration of the effects of different network topologies leads to two important conclusions. First, our results indicate that contagion with homophily does not necessarily lead to cultural differentiation. Only when the network is already segmented into different cliques does homophily produce cultural differentiation. Second, we find that network topology also matters for associative diffusion. Scale-free networks support the emergence of crisp cultural differentiation, while small-world networks seem to make this process more subtle and fragmented. Unlike small-world networks, scale-free networks facilitate informational diffusion therefore leading to broad interpretative consensus. But when agents are embedded in weakly connected clusters, information does not freely travel between cliques and agents reach a weaker interpretative consensus. In the real world, individuals occupy multiplex network positions that embody both scale-free and small-world network properties. Our results suggests that crisp and unidimensional cultural differentiation is more likely to emerge when cultural information is interpersonally transmitted along scale-free patterns of interaction. When this process is undergirded by small-world networks, a more complex and interpretatively heterogeneous cultural order emerges.

Discussion

If culture is contagious, then how does cultural variation come about? Existing work typically assumes that cultural heterogeneity is the result of preexisting or emergent structural boundaries to diffusion. Cultural variation therefore merely mirrors an underlying segmented network structure. But as Barth (1969) pointed out half a century ago, the view that cultural difference is produced through social disconnection is simplistic and incomplete; cultural boundaries persist despite the constant flow of people across them.

We proposed a cognitively-informed diffusion model that overcomes this impasse. Our agent-based computational experiments demonstrate that associative diffusion leads to the emergence and endurance of patterned cultural variation even when people freely criss-cross emergent cultural boundaries in their interactions. Conventional diffusion models, in contrast, cannot explain cultural differentiation unless they assume a preexistent or emergent archipelago of near-isolated cliques.

When Does Culture Diffuse Associatively?

Contagion models require structural complexity to explain cultural variation because they conceptualize interpersonal transmission as a simple epidemiological process. Of course, network scholars do not interpret the contagion metaphor literally, as if culture spreads through mere exposure. They rely on this theoretical simplification, however, because they implicitly assume that culture is only transmitted through strong and homophilous relationships, and the deep and trustful interactions these relationships afford. Such an assumption implies that network topology is primarily consequential for diffusion when information is complex or costly—for example, when interlocutors exchange secrets (Cowan 2014) or novel information (Aral and Van Alstyne 2011).

But cultural information need not be complex nor costly; its diffusion therefore does not necessitate a strong network tie. Consider the diffusion of cycling. I do not need to know the man riding down the street on a bicycle to notice that he is wearing a suit. Nor do I need to observe his other social attributes—his occupation, culinary preferences or political ideology—to be influenced by his behavior. Such a cursory encounter would not allow me to reliably calculate my distance from this bicycle rider in a euclidean socio-cultural space, which is how contagion models conventionally operationalize network tie valence (e.g. DellaPosta et al. 2015; Baldassarri and Bearman 2007). The mere observation of a stranger I know nothing about and who is riding a bike down a city street would be unlikely to catalyze me into doing the same. Nevertheless, through my knowledge of the symbolic significance of wearing a suit I can infer what riding a bike means. Future encounters with bicycle riders would either reinforce or undermine this inference. Cultural symbols, in other words, are effective by virtue of being easily transmittable. Their complexity is a function of the intricate semantic webs into which they are interwoven. These webs exist as cognitive representations in the minds of those observing and enacting symbolic action.[20]

This does not mean that social networks are inconsequential for diffusion. In fact, as we illustrate in Figure 7, different network topologies lead to different cultural diffusion dynamics. Like conventional contagion models, our model of associative diffusion assumes that people learn culture from network alters with whom they interact. This model is distinctive not in what networks do but in what agents do with the information they receive through their network ties. Different types of information and different types of network relationships, we contend, afford different types of cultural diffusion dynamics and result in different forms of shared interpretation.

Strong and trustful relationships facilitate the exchange of complex and costly cultural knowledge, such as the cultural education that occurs when parents socialize their children. In contrast, superficial interactions, whether through ephemeral or durable network ties, can catalyze associative diffusion when two conditions hold. First, behaviors need to be observable, either because they cannot be done in private or because people choose to perform them in public. Second, there needs to be some uncertainty about the functional utility of adoption. When this functional utility is easily discoverable independently—for example when information about a job opportunity diffuses—then an individual exposed to new information does not need to rely on others in order to interpret it; under such conditions, simple contagion is likely to occur. But when there exists interpretative ambiguity—such as the toxicity of vaccines—people look at others to make sense of the information they had received. This, we suspect, is when observable behaviors acquire symbolic value and when associative diffusion processes are most likely to kick in.

Contribution to the Sociology of Networks and Diffusion

Overall, our findings have several implications for the study of networks and diffusion. First, our associative model shifts focus from the diffusion of practices to the diffusion of interpretation. Whereas in conventional models of cultural diffusion agents emulate others' discrete behaviors, in our model they learn which behaviors are compatible with one another. Thus, the diffusion of a practice depends not only on its first adopter's network position (Banerjee, Chandrasekhar, Duflo, and Jackson 2013), its inherent appeal (Berger and Milkman 2012), or its functional utility (Kolodny, Creanza, and Feldman 2015). Rather, it also depends on the distribution of other practices in the population. This implies that understanding the rise or decline of a cultural practice requires paying attention to other seemingly unrelated practices and their prevalence in the population. The diffusion of bicycles in Victorian England, for example, was inherently related to the modes of dress they afforded (Bijker 1995).

Second, as Strang and Meyer (1993) point out, traditional network models cannot explain why interaction some times leads to solidarity and other times begets conflict. Associative diffusion, in contrast, demonstrates how interaction between members of different "thought communities" (Zerubavel 1999) can, counter-intuitively, serve to entrench cultural boundaries and intensify preference polarization. Our two-stage diffusion model analytically distinguishes between interpretation and evaluation. Consequently, agents reach interpretative agreement—as reflected in the declining distance between their association matrices (Figure 5)—but adopt opposing preferences. Interaction leads individuals to coordinate their perceptions about the cultural order, not their preferences. Interlocutors might learn from each others' behaviors, for example, that individuals who support same-sex marriage also favor gun control, or that those who consume organic food also tend to object to childhood vaccinations. But the same information can lead to divergent preference updating. An exchange between two parents on the merits of immunizations might therefore strengthen their disagreement, rather than foster consensus. Existing network models cannot account for this phenomenon unless they assume the preexistence of negative network ties (e.g. Flache and Macy 2011).

Network structure plays a surprising role in associative diffusion. Contra conventional network theory wisdom, Figure 7 illustrates that a small-world segmented network topology inhibits, rather than facilitates, the emergence of crisp cultural differentiation. Though the number of cultural clusters slowly and gradually increases under a small-world architecture, the mutual information between behaviors remains significantly lower than when culture associatively diffuses over scale-free or fully connected networks. The demarcation of a cultural boundary, in other words, requires unconstrained interaction between members of the groups it separates (Fischer 1995). When such interaction is stymied, cultural differentiation emerges along fuzzier symbolic boundaries.

These results point to a potential synthesis between network-centric approaches that assume naive contagion and network-free associative diffusion. On their own, both approaches explain a process that results in cultural differentiation along a singular and crisp dimension of interpretative consensus. But recent work in cultural sociology demonstrates that individuals differ not only in their beliefs and preferences but also in the dimensions of meaning along which these beliefs and preferences are distributed (Goldberg 2011; Baldassarri and Goldberg 2014). Fused together, associative diffusion and network theory appear to explain the emergence of such a schematically heterogeneous world. When culture associatively diffuses over a small-world network people differentiate into multiple clusters that are structured along a multiplicity of cultural axes. The boundaries separating these different groups are not as pronounced as when interaction is free. In reality, people chronically intersect small-world networks of intense cultural transmission and more ephemeral scale-free networks that facilitate the associative diffusion of easily transmittable cultural information. We imagine that this multiplexity of ties, and the different diffusions they afford, is what enables the emergence of interpretative heterogeneity. We leave this exploration for future work.

Contribution to the Sociology of Culture

Our findings also inform sociological theories of culture. A variety of recent studies build on the symbolic interactioinst notion that meaning arises through social exchange (e.g. Hunzaker 2016). These models explain how cultural order is interactionally reproduced, but not where it comes from to begin with. In Schröder et al's (2016) elegant Bayesian model of Affect Control Theory, for example, agents' identities are situationally produced through people's motivation to reduce inconsistencies between behaviors being enacted and the meanings these behaviors connote. A fundamental assumption in Affect Control Theory, however, is that all agents associate the same meanings with the same behaviors.[21]

Agents therefore reproduce a given cultural order. Our model, in contrast, makes no such assumption. Rather, we demonstrate how meaning arises through the process of associative diffusion. Agents' sensitivity to associations between practices, as well as their adherence to constraint satisfaction, leads them to partition practices into different clusters. These clusters effectively constitute different categories, the enactment of which divides the population into different emergent groups.

Few sociological works similarly consider how meaning emerges through the process of diffusion. A prominent exception is Strang and Meyer's (1993) theory of institutional diffusion. The authors point to *theorization*—"the self-conscious development and specification of abstract categories and the formulation of patterned relationships" (p. 492)—as a central catalyst for the emergence and diffusion of cultural meanings. But theorization requires the intentional actions of theorists. In our model, in contrast, no agents have such intentionality or the institutional authority to theorize. Rather, categories implicitly emerge though the gradual clustering of practices. As clusters cohere, behaviors' information content—namely, their cultural meaningfulness—increases.

Our model of associative diffusion stops here. But in reality agents go beyond mere cognitive association. Clusters of practices become reified when agents employ labels—such as *lads*, *ear'ols* or *anti-vaxxers*—to clearly denote these emergent and hitherto unnamed categories. As these categories are attributed to people, and internalized by those they are applied to, they become identities. Our rudimentary associative diffusion model does not explicitly assume identities or higher level categorizations. Relying on basic socio-cognitive building blocks it nevertheless demonstrates how such identities can emerge through interaction. We leave the investigation of the effects of labeling and identity formation on cultural diffusion for future work.

While our model explains how cultural meanings emerge, it does not account for cultural change once these meanings become solidified. In fact, we demonstrate that the emergence of cultural differentiation is robust to the behaviors of nonconformists—who are often assumed to be agents of change—as long as there exist at least a handful of conformist individuals (Figure 6). The vast majority of conventional contagion models assume that all agents are perfectly conformist. When they do not, they find that nonconformist agents are conducive to dramatic—but rare—behavioral cascades, such as when risky collective action takes off or when costly and widely held conventions suddenly dissipate (Mackie 1996). But a significant portion of cultural change occurs through gradual and cyclic endogenous evolution

(Lieberson 2000). When extended to account for variation in conformity, our model explains how culture is both durable and constantly evolving (Hays 1994). Cultural durability stems from emergent interrelationships between practices, keeping their categorical meanings stable. But the presence of nonconformists catalyzes cycles of popularity whereby different practices ebb and flow in their pervasiveness. Thus, fundamental social identities such as liberal and conservative or high- and low-brow are historically durable, even if their behavioral manifestations—for example, parents' inclination to vaccinate their children—slowly evolve.

Conclusion: Associative Diffusion as Social Construction

Sociologists mostly agree that culture is a system of shared understandings. But students of cultural diffusion have overwhelmingly left meaning out of their epidemiologically-inspired models. Consequently, they attribute population-level variation in cultural preferences and beliefs to an underlying clustered network structure. We propose an alternative model of associative diffusion whereby arbitrary cultural meanings emerge and become consensually accepted through social interaction. Sociologists often refer to this process as *social construction* (Berger and Luckmann 1967).

Network scholars' inattention to meaning and interpretation has led them to overlook two important aspects of cultural diffusion that complicate the relationship between network structure and behavioral outcomes. First, interpretation decouples exposure to a practice from its adoption. Two structurally equivalent individuals might therefore react differently even if they observe the same behaviors by others. Second, because meaning inheres in associative relationships, cultural practices are semantically dependent on one another. The diffusion of a cultural practice is consequently affected by the distribution of other practices in the population. Network studies often overlook these complications, assuming that network topology alone determines patterns of diffusion. In contrast, we demonstrate that associative diffusion leads to differential adoption of practices even when there are no constraints on interaction. Moreover, we show that small-world clustered networks impede, rather than facilitate, the emergence of clear-cut cultural differentiation.

Though contagion models ordinarily treat cultural practices as discrete meme-like entities, there are a few exceptions. A handful of recent studies—mostly outside of sociology—have proposed, as we do, that cultural practices are interdependent (e.g. Kolodny et al.

2015; Enquist, Ghirlanda, and Eriksson 2011). These models nevertheless assume that such interdependencies are a function of these practices' inherent functional or logical attributes. Cultural order, in other words, is given a priori by natural constraints. But culture is, by definition, anything but natural. The interpretation of vaccines as healthy, or unnatural, is historically and socially contingent. Our model of associative diffusion explains how such shared interpretations emerge organically through interaction and why continued unobstructed interaction between differently opinionated individuals only serves to deepen cultural cleavages. Thus, once opposition to vaccinations becomes associated with other practices such as organic food consumption, social interaction between anti-vaxxers and parents who vaccinate their children only entrenches, rather than defuses, the cultural boundary that separates them.

Although the imagery of social contagion dominates the sociological imagination, cultural evolution is not analogous to epidemiological diffusion. From lifestyle choices such as musical or aesthetic taste (Bourdieu 1986) to political and religious ideology (Baldassarri and Gelman 2008), societies exhibit persistent cultural differentiation. Our agent-based model demonstrates that these divisions do not depend on the preexistence of a segregated social structure or of primordial social groups. Rather, clustered cultural variation can emerge as a consequence of the connectionist nature of human cognition.

Notes

[1] Following Strang and Soule (1998), we use the term 'practice' to denote, in a very general sense, a cultural element such as a belief, attitude or behavioral preference that can be transmitted between individuals. In the model that we develop below we formally distinguish between a preference, which represents the proclivity toward a cultural practice, and the behavior such a practice entails.

[2] During the 2016 presidential election, for example, skepticism about vaccines was voiced by conservative candidates such as Donald Trump and Carly Fiorina, as well as liberal candidates such as Jill Stein, the Green Party's presidential nominee.

[3] Beliefs on vaccination are a prime example of what sociologists often refer to as the social construction of rationality. Despite mounting and consistent evidence that vaccines are safe and are not associated with developmental disorders, and despite the undeniable potential lethality of childhood diseases such as measles, parents' choice not to vaccinate their children is commonly couched in rationalized calculations of risk (Reich 2016). Popular accounts often trace the rise of the anti-vaccination movement to a scientific study published by Andrew Wakefield in 1998, and later denounced as "the most damaging medical hoax of the last 100 years" (Flaherty 2011, p. 1302), which argued for a causal relationship between the MMR vaccine and autism. But, as Conis (2015, 2014) cogently argues, this study was the product of brewing skepticism

toward immunizations, rather than its catalyst. What is so striking about the anti-vaccination movement is that college educated individuals who are otherwise most receptive to scientific evidence continue to draw on this study, which has since been retracted, while rejecting the dozens of other studies refuting its conclusions.

[4] Objection to vaccines is not new. Vaccines were met with fierce resistance, especially when they were first introduced in the nineteenth century. But by the second half of the twentieth century this resistance was mostly subdued.

[5] Though the distinction between tie formation and tie strength is analytically important, it is often inconsequential for diffusion models. What these two constructs affect is the likelihood of social transmission between individuals. Thus the distinction between opportunity for or susceptibility to social influence is often semantic rather than substantive, unless negative influence is assumed (e.g. Flache and Macy 2011; Mark 2003).

[6] DellaPosta et al. (2015), for example, assume that ego networks are limited in size by Dunbar's number, which is presumably the upper limit on the number of durable social relationships that humans can cognitively maintain.

[7] Though Carley never uses the imagery of contagion, her model assumes that once interaction occurs, cultural knowledge is invariably exchanged. In that respect "constructural" models are no different from other contagion models.

[8] As Strang and Soule (1998) point out, there is much ambiguity in diffusion research about what is being observed. By observability we mean the opportunity to be exposed to symbolic information, whether verbal or nonverbal.

[9] In some models preferences are modeled as continuous values, and social influence is operationalized as one agent adopting another agent's preference (e.g. Friedkin and Johnsen 1990). Though these models do not treat adoption as a binary outcome, they nevertheless assume that the receiving agent observes the transmitting agent's preference and adopts it.

[10] Figure 1 illustrates a simple contagion. When contagion is complex, adoption necessitates exposure to more than one individual.

[11] As others have noted (e.g Sperber 1996), this assumption is often incorrect as public representations are symbolic simplifications of more elaborately represented cognitive concepts. Thus, culture can evolve through the process of diffusion if interpersonal transmission is imperfect. We leave this implication outside the scope of our model.

[12] We recognize that interpretation and evaluation are causally intertwined, and that how one evaluates a behavior often affects how that behavior is construed, rather than the other way around. Yet as analytical moments, these are two distinct phases in the process of assessing a cultural practice.

[13] We acknowledge that our operationalization of cognition is simplified. It is an analytical abstraction of an underlying complex neurophysiology, the details of which is beyond the scope of this study.

[14] It is important to point out that even if the constraint is psychological, it may be subjectively experienced as ontological. Prentice and Miller (2006) for example demonstrate that people essentialize observed behavioral regularities as natural. This reification is cognitively important and therefore consequential for social processes, but is beyond the scope of our study.

[15] If instead of randomly updating the weaker preference we set Δv to be in a direction that reduces cognitive dissonance, the simulations reported below naturally converge significantly faster. Nevertheless, we allow for the possibility that other cognitive mechanisms may be inconsistent with constraint satisfaction. To do so, we assume that our agents are random updaters. This ensures that our results are not driven by the assumption that agents are perfect constraint maximizers.

[16] This tendency is often referred to in the literature as biased assimilation, and is assumed to be facilitated by motivated reasoning, a cognitive bias that leads people to process information in a way that serves their interests and preserves their self-image.

[17] We thank anonymous Reviewer 4 for pointing out early adopters as an alternative to associative diffusion.

[18] Roughly 1% of simulations in the naive and conformist conditions, and 2% of cases in the biased condition, converge on more than one cluster. We introduce stochasticity into the naive model such that agents do not perfectly observe their interlocutors' preferences but rather infer them on the basis of having observed their corresponding behaviors being performed. Otherwise, all agents would have converged on the exact same preferences and the number of clusters would have invariably been 1. See Appendix C for more details.

[19] This assumption does not affect the results we report in this section.

[20] The bicycle sharing system launched in New York City in 2013 provides an interesting case in point. Initially enthusiastically endorsed by social activists and environmentalists, the program became the target of culture jammers' ridicule and criticism once it was announced that it would be sponsored by Citi Bank and correspondingly named *CitiBike*. Nothing about the program had changed; its cultural meaning, however, was transformed dramatically by virtue of its association with a major U.S. bank. People's propensity to adopt the program was shaped by its emergent cultural meaning, not by their awareness of its existence.

[21] These associations are given by empirically derived "affective dictionaries" that presumably represent a fundamentally shared cultural grammar.

References

Anderson, Elijah. 1999. *Code of the Street: Decency, Violence, and the Moral Life of the Inner City*. W. W. Norton & Company.

Aral, Sinan and Marshall Van Alstyne. 2011. "The Diversity-Bandwidth Trade-Off." *American Journal of Sociology* 117:90–171.

Axelrod, Robert. 1997. "The Dissemination of Culture: A Model with Local Convergence and Global Polarization." *Journal of Conflict Resolution* 41:203–226.

Baldassarri, Delia and Peter Bearman. 2007. "Dynamics of Political Polarization." *American Sociological Review* 72:784–811.

Baldassarri, Delia and Andrew Gelman. 2008. "Partisans without Constraint: Political Polarization and Trends in American Public Opinion." *American Journal of Sociology* 114:408–446.

Baldassarri, Delia and Amir Goldberg. 2014. "Neither Ideologues nor Agnostics: Alternative Voters' Belief System in an Age of Partisan Politics." *American Journal of Sociology* 120:45–95.

Banerjee, Abhijit, Arun G. Chandrasekhar, Esther Duflo, and Matthew O. Jackson. 2013. "The Diffusion of Microfinance." *Science* 341:1236498.

Barth, Fredrik (ed.). 1969. *Ethnic Groups and Boundaries: The Social Organization of Culture Difference*. Long Grove, Ill: Waveland Press.

Barucca, Paolo, Jacopo Rocchi, Enzo Marinari, Giorgio Parisi, and Federico Ricci-Tersenghi. 2015. "Cross-Correlations of American Baby Names." *Proceedings of the National Academy of Sciences* 112:7943–7947.

Berger, Jonah and Katherine L Milkman. 2012. "What Makes Online Content Viral?" *Journal of Marketing Research* 49:192–205.

Berger, Peter L. and Thomas Luckmann. 1967. *The Social Construction of Reality: A Treatise in the Sociology of Knowledge*. New York: Anchor.

Bijker, Wiebe E. 1995. *Of Bicycles, Bakelites, and Bulbs: Toward a Theory of Sociotechnical Change*. MIT Press.

Bourdieu, Pierre. 1986. *Distinction: A Social Critique of the Judgement of Taste*. London: Routledge, 1 edition edition.

Bourdieu, Pierre. 1993. *The Field of Cultural Production*. New York: Columbia University Press, 1st edition edition.

Carley, Kathleen. 1991. "A Theory of Group Stability." *American Sociological Review* 56:331–354.

Carroll, Glenn R. and Anand Swaminathan. 2000. "Why the Microbrewery Movement? Organizational Dynamics of Resource Partitioning in the U.S. Brewing Industry." *American Journal of Sociology* 106:715–762.

Centola, Damon. 2011. “An Experimental Study of Homophily in the Adoption of Health Behavior.” *Science* 334:1269–1272.

Centola, Damon. 2015. “The Social Origins of Networks and Diffusion.” *American Journal of Sociology* 120:1295–1338.

Centola, Damon, Juan Carlos González-Avella, Víctor M. Eguíluz, and Maxi San Miguel. 2007. “Homophily, Cultural Drift, and the Co-Evolution of Cultural Groups.” *Journal of Conflict Resolution* 51:905–929.

Centola, Damon and Michael Macy. 2007. “Complex Contagions and the Weakness of Long Ties.” *American Journal of Sociology* 113:702–734.

Chan, Cindy, Jonah Berger, and Leaf Van Boven. 2012. “Identifiable but Not Identical: Combining Social Identity and Uniqueness Motives in Choice.” *Journal of Consumer Research* 39:561–573.

Cheryan, Sapna, Paul G. Davies, Victoria C. Plaut, and Claude M. Steele. 2009. “Ambient Belonging: How Stereotypical Cues Impact Gender Participation in Computer Science.” *Journal of personality and social psychology* 97:1045–1060.

Childress, C. Clayton and Noah E. Friedkin. 2012. “Cultural Reception and Production: The Social Construction of Meaning in Book Clubs.” *American Sociological Review* 77:45–68.

Christakis, Nicholas A. and James H. Fowler. 2007. “The Spread of Obesity in a Large Social Network over 32 Years.” *New England Journal of Medicine* 357:370–379.

Cialdini, Robert B. 2007. *Influence: The psychology of persuasion*. Collins New York.

Cialdini, Robert B. and Noah J. Goldstein. 2004. “Social Influence: Compliance and Conformity.” *Annual Review of Psychology* 55:591–621.

Conis, Elena. 2014. *Vaccine Nation: America’s Changing Relationship with Immunization*. Chicago ; London: University Of Chicago Press, 1 edition edition.

Conis, Elena. 2015. “The American Historian: Vaccination Resistance in Historical Perspective.” *The American Historian* .

Cowan, Sarah K. 2014. "Secrets and Misperceptions: The Creation of Self-Fulfilling Illusions." *Sociological Science* 1:466–492.

Dandekar, Pranav, Ashish Goel, and David T. Lee. 2013. "Biased Assimilation, Homophily, and the Dynamics of Polarization." *Proceedings of the National Academy of Sciences* 110:5791–5796.

D'Andrade, Roy G. 1995. *The Development of Cognitive Anthropology*. Cambridge ; New York: Cambridge University Press.

DellaPosta, Daniel, Yongren Shi, and Michael Macy. 2015. "Why Do Liberals Drink Lattes?" *American Journal of Sociology* 120:1473–1511.

DiMaggio, Paul and Filiz Garip. 2012. "Network Effects and Social Inequality." *Annual Review of Sociology* 38:93–118.

DiMaggio, Paul and Amir Goldberg. 2018. "Searching for Homo Economicus: Variation in Americans' Construals of and Attitudes toward Markets." *European Journal of Sociology / Archives Européennes de Sociologie* pp. 1–39.

Ehret, Phillip J., Brian M. Monroe, and Stephen J. Read. 2015. "Modeling the Dynamics of Evaluation: A Multilevel Neural Network Implementation of the Iterative Reprocessing Model." *Personality and Social Psychology Review* 19:148–176.

Enquist, Magnus, Stefano Ghirlanda, and Kimmo Eriksson. 2011. "Modelling the Evolution and Diversity of Cumulative Culture." *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 366:412–423.

Fischer, Claude S. 1995. "The Subcultural Theory of Urbanism: A Twentieth-Year Assessment." *American Journal of Sociology* 101:543–577.

Flache, Andreas and MICHAEL W. Macy. 2011. "Small Worlds and Cultural Polarization." *The Journal of Mathematical Sociology* 35:146–176.

Flaherty, Dennis K. 2011. "The Vaccine-Autism Connection: A Public Health Crisis Caused by Unethical Medical Practices and Fraudulent Science." *Annals of Pharmacotherapy* 45:1302–1304.

Freeman, J. B. and N. Ambady. 2011. "A Dynamic Interactive Theory of Person Construal." *Psychological review* 118:247–279.

Friedkin, Noah E. and Eugene C. Johnsen. 1990. "Social Influence and Opinions." *The Journal of Mathematical Sociology* 15:193–206.

Frye, Margaret. 2017. "Cultural Meanings and the Aggregation of Actions: The Case of Sex and Schooling in Malawi." *American Sociological Review* 82:945–976.

Gamson, William A. and Andre Modigliani. 1989. "Media Discourse and Public Opinion on Nuclear Power: A Constructionist Approach." *American Journal of Sociology* 95:1–37.

Gauchat, Gordon. 2012. "Politicization of Science in the Public Sphere: A Study of Public Trust in the United States, 1974 to 2010." *American Sociological Review* 77:167–187.

Goldberg, Amir. 2011. "Mapping Shared Understandings Using Relational Class Analysis: The Case of the Cultural Omnivore Reexamined." *American Journal of Sociology* 116:1397–1436.

Goldberg, Amir, Michael T. Hannan, and Balázs Kovács. 2016. "What Does It Mean to Span Cultural Boundaries? Variety and Atypicality in Cultural Consumption." *American Sociological Review* 81:215–241.

Granovetter, Mark. 1978. "Threshold Models of Collective Behavior." *American Journal of Sociology* 83:1420–1443.

Griffiths, Thomas L., Michael L. Kalish, and Stephan Lewandowsky. 2008. "Theoretical and Empirical Evidence for the Impact of Inductive Biases on Cultural Evolution." *Philosophical Transactions of the Royal Society B: Biological Sciences* 363:3503–3514.

Hargadon, Andrew B. and Yellowlees Douglas. 2001. "When Innovations Meet Institutions: Edison and the Design of the Electric Light." *Administrative Science Quarterly* 46:476–501.

Hays, Sharon. 1994. "Structure and Agency and the Sticky Problem of Culture." *Sociological Theory* 12:57–72.

Henrich, Joe and Robert Boyd. 1998. "The Evolution of Conformist Transmission and the Emergence of Between-Group Differences." *Evolution and Human Behavior* 19:215–241.

Hochschild, Arlie Russell. 2016. *Strangers in Their Own Land: Anger and Mourning on the American Right*. New York: The New Press, reprint edition edition.

Holyoak, Keith J. and Dan Simon. 1999. "Bidirectional Reasoning in Decision Making by Constraint Satisfaction." *Journal of Experimental Psychology: General* 128:3.

Horne, Zachary, Derek Powell, John E Hummel, and Keith J Holyoak. 2015. "Countering antivaccination attitudes." *Proceedings of the National Academy of Sciences* 112:10321–10324.

Hunzaker, M. B. Fallin. 2016. "Cultural Sentiments and Schema-Consistency Bias in Information Transmission." *American Sociological Review* 81:1223–1250.

Jablonka, Eva and Marion J. Lamb. 2006. *Evolution in Four Dimensions: Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life*. Cambridge, Mass.: A Bradford Book, 1 edition edition.

Kaufman, Jason and Orlando Patterson. 2005. "Cross-National Cultural Diffusion: The Global Spread of Cricket." *American Sociological Review* 70:82–110.

Kenward, Ben. 2012. "Over-Imitating Preschoolers Believe Unnecessary Actions Are Normative and Enforce Their Performance by a Third Party." *Journal of Experimental Child Psychology* 112:195–207.

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* 105:10681–10686.

Kolodny, Oren, Nicole Creanza, and Marcus W. Feldman. 2015. "Evolution in Leaps: The Punctuated Accumulation and Loss of Cultural Innovations." *Proceedings of the National Academy of Sciences* 112:E6762–E6769.

Kunda, Ziva. 1990. "The Case for Motivated Reasoning." *Psychological Bulletin* pp. 480–498.

Kunda, Ziva and Paul Thagard. 1996. "Forming Impressions From Stereotypes, Traits, and Behaviors: A Parallel-Constraint-Satisfaction Theory." *Psychological Review* 103:284–308.

Lamont, Michèle, Molnár, and Virág. 2002. "The Study of Boundaries in the Social Sciences." *Annual Review of Sociology* 28:167–195.

Leskovec, Jure, Anand Rajaraman, and Jeffrey David Ullman. 2014. *Mining of Massive Datasets*. Cambridge: Cambridge University Press, 2nd edition.

Lieberson, Stanley. 2000. *A Matter of Taste: How Names, Fashions, and Culture Change*. New Haven, CT: Yale University Press.

Lizardo, Omar. 2006. "How Cultural Tastes Shape Personal Networks." *American Sociological Review* 71:778–807.

Lord, Charles G., Lee Ross, and Mark R. Lepper. 1979. "Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence." *Journal of Personality and Social Psychology* 37:2098–2109.

Luce, R. Duncan. 2005. *Individual Choice Behavior: A Theoretical Analysis*. Mineola, N.Y: Dover Publications.

Lyons, Derek E., Andrew G. Young, and Frank C. Keil. 2007. "The Hidden Structure of Overimitation." *Proceedings of the National Academy of Sciences* 104:19751–19756.

Mackie, Gerry. 1996. "Ending Footbinding and Infibulation: A Convention Account." *American Sociological Review* 61:999–1017.

Malka, Ariel, Christopher J. Soto, Michael Inzlicht, and Yphtach Lelkes. 2014. "Do Needs for Security and Certainty Predict Cultural and Economic Conservatism? A Cross-National Analysis." *Journal of Personality and Social Psychology* 106:1031–1051.

Mark, Noah. 1998. "Beyond Individual Differences: Social Differentiation from First Principles." *American Sociological Review* 63:309–330.

Mark, Noah P. 2003. "Culture and Competition: Homophily and Distancing Explanations for Cultural Niches." *American Sociological Review* 68:319–345.

Martin, John Levi. 2002. "Power, Authority, and the Constraint of Belief Systems." *American Journal of Sociology* 107:861–904.

Mäs, Michael and Andreas Flache. 2013. “Differentiation without Distancing. Explaining Bi-Polarization of Opinions without Negative Influence.” *PLOS ONE* 8:e74516.

Miller, Dale T. and Deborah A. Prentice. 2016. “Changing Norms to Change Behavior.” *Annual Review of Psychology* 67:339–361.

Monroe, B. M. and S. J. Read. 2008. “A General Connectionist Model of Attitude Structure and Change: The ACS (Attitudes as Constraint Satisfaction) Model.” *Psychological review* 115:733–759.

Moore, Geoffrey A. 2006. *Crossing the Chasm: Marketing and Selling High-Tech Products to Mainstream Customers*. New York, NY: HarperBusiness, revised edition edition.

Moreau, C. Page, Arthur B. Markman, and Donald R. Lehmann. 2001. ““What Is It?” Categorization Flexibility and Consumers’ Responses to Really New Products.” *Journal of Consumer Research* 27:489–498.

Murphy, Gregory. 2004. *The Big Book of Concepts*. Cambridge: A Bradford Book, revised edition edition.

Ocejo, Richard E. 2017. *Masters of Craft: Old Jobs in the New Urban Economy*. Princeton University Press.

Paluck, E. L. and H. Shepherd. 2012. “The Salience of Social Referents: A Field Experiment on Collective Norms and Harassment Behavior in a School Social Network.” *Journal of personality and social psychology* 103:899–915.

Papachristos, Andrew V. 2009. “Murder by Structure: Dominance Relations and the Social Structure of Gang Homicide.” *American Journal of Sociology* 115:74–128.

Patterson, Karalyn, Peter J. Nestor, and Timothy T. Rogers. 2007. “Where Do You Know What You Know? The Representation of Semantic Knowledge in the Human Brain.” *Nature Reviews Neuroscience* 8:976–987.

Patterson, Orlando. 2014. “Making Sense of Culture.” *Annual Review of Sociology* 40:1–30.

Peterson, Richard A. and Roger M. Kern. 1996. “Changing Highbrow Taste: From Snob to Omnivore.” *American Sociological Review* 61:900–907.

Petty, Richard E. and Jon A. Krosnick (eds.). 1995. *Attitude Strength: Antecedents and Consequences*. Mahwah, N.J: Psychology Press, 1 edition edition.

Phillips, Damon J. 2013. *Shaping Jazz: Cities, Labels, and the Global Emergence of an Art Form*. Princeton University Press.

Prentice, Deborah A. and Dale T. Miller. 2006. "Essentializing Differences Between Women and Men." *Psychological Science* 17:129–135.

Reich, Jennifer A. 2016. *Calling the Shots: Why Parents Reject Vaccines*. New York: NYU Press, 1 edition edition.

Rogers, Everett M. 2010. *Diffusion of Innovations, 4th Edition*. Simon and Schuster.

Salganik, Matthew J., Peter Sheridan Dodds, and Duncan J. Watts. 2006. "Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market." *Science* 311:854–856.

Schröder, Tobias, Jesse Hoey, and Kimberly B. Rogers. 2016. "Modeling Dynamic Identities and Uncertainty in Social Interactions: Bayesian Affect Control Theory." *American Sociological Review* 81:828–855.

Schröder, Tobias and Paul Thagard. 2014. "Priming: Constraint Satisfaction and Interactive Competition." *Social Cognition* 32:152–167.

Shultz, Thomas R. and Mark R. Lepper. 1996. "Cognitive Dissonance Reduction as Constraint Satisfaction." *Psychological Review* 103:219–40.

Simon, Dan, Daniel C. Krawczyk, and Keith J. Holyoak. 2004. "Construction of Preferences by Constraint Satisfaction." *Psychological Science* 15:331–336.

Snyder, C. R. and Howard L. Fromkin. 1980. *Uniqueness, the Human Pursuit of Difference*. Plenum Press.

Sperber, Dan. 1996. *Explaining Culture: A Naturalistic Approach*. Oxford, UK ; Cambridge, Mass: Blackwell Publishers, 1 edition edition.

Strang, David and John W. Meyer. 1993. "Institutional Conditions for Diffusion." *Theory and Society* 22:487–511.

Strang, David and Sarah A. Soule. 1998. “Diffusion in Organizations and Social Movements: From Hybrid Corn to Poison Pills.” *Annual Review of Sociology* 24:265–290.

Taber, Charles S. and Milton Lodge. 2006. “Motivated Skepticism in the Evaluation of Political Beliefs.” *American Journal of Political Science* 50:755–769.

Tibshirani, Robert, Guenther Walther, and Trevor Hastie. 2001. “Estimating the Number of Clusters in a Data Set via the Gap Statistic.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63:411–423.

Trope, Yaacov and Nira Liberman. 2010. “Construal-Level Theory of Psychological Distance.” *Psychological review* 117:440–463.

Turner, Fred. 2008. *From Counterculture to Cyberculture: Stewart Brand, the Whole Earth Network, and the Rise of Digital Utopianism*. Chicago, Ill.: University Of Chicago Press, 60265th edition edition.

Watts, Duncan J. 1999. “Networks, Dynamics, and the Small-World Phenomenon.” *American Journal of Sociology* 105:493–527.

Watts, Duncan J. 2002. “A Simple Model of Global Cascades on Random Networks.” *Proceedings of the National Academy of Sciences* 99:5766–5771.

Willer, Robb, Ko Kuwabara, and Michael W. Macy. 2009. “The False Enforcement of Unpopular Norms.” *American Journal of Sociology* 115:451–490.

Willis, Paul E. 1977. *Learning to Labor: How Working Class Kids Get Working Class Jobs*. Columbia University Press.

Zerubavel, Eviatar. 1999. *Social Mindscapes: An Invitation to Cognitive Sociology*. Cambridge, MA: Harvard University Press, reprint edition edition.

Zuckerman, Ezra W. 2012. “Construction, Concentration, and (Dis)Continuities in Social Valuations.” *Annual Review of Sociology* 38:223–245.

Tables

Table 1: Model Overview

Agent Initialization

Each agent holds two types of information:
1. associations: $R_{ij} = 1, \forall i, j \in K$
2. preferences: $V_i \sim U(-1, 1)$

Modeling Sequence

1. Select agents A and B at random
2. B observes A exhibiting practices i and j with probabilities $P(i)$ and $P(j)$
3. B updates $R_{ij} = R_{ij} + 1$
4. B selects preference k to update, where k is the weaker of v_i and v_j
5. B updates preferences, V', by setting $v'_k = v_k + \sim N(0, 1)$
6. iff $CS(V', R) > CS(V, R)$, V' is retained, otherwise revert to V
7. Apply decay function $R_{ij} = \lambda R_{ij}$

Figures

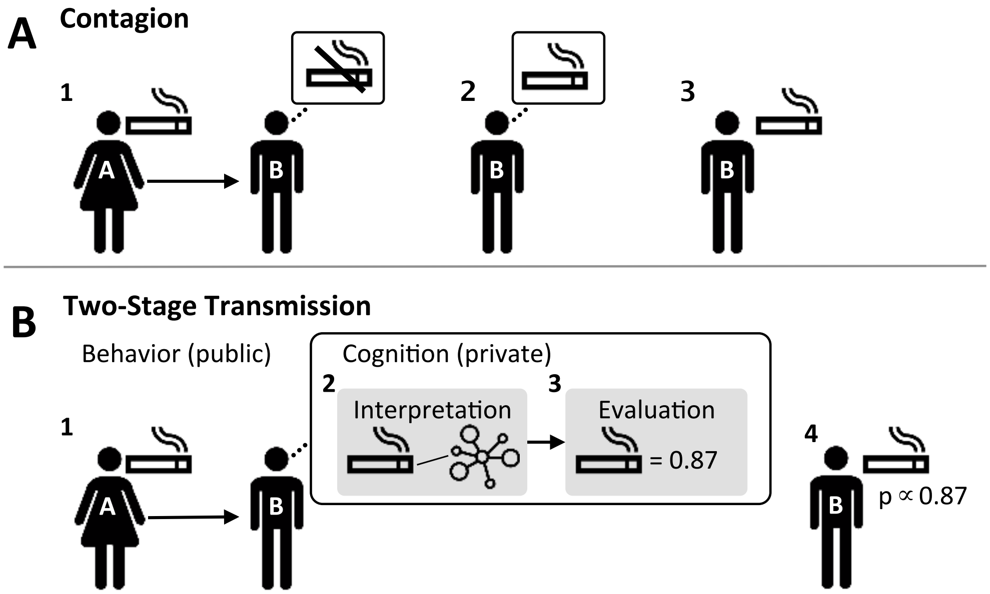

Figure 1: The process of cultural transmission in the contagion (A) and two-stage transmission (B) models. In both illustrations, agent B is observing agent A smoking. Square callouts relate to B's cognition. In (A), B changes his preference from anti-smoking to smoking, and consequently smokes. In (B), he updates his interpretation of smoking and his preference for smoking, and consequently smokes with an illustrative probability of 0.87.

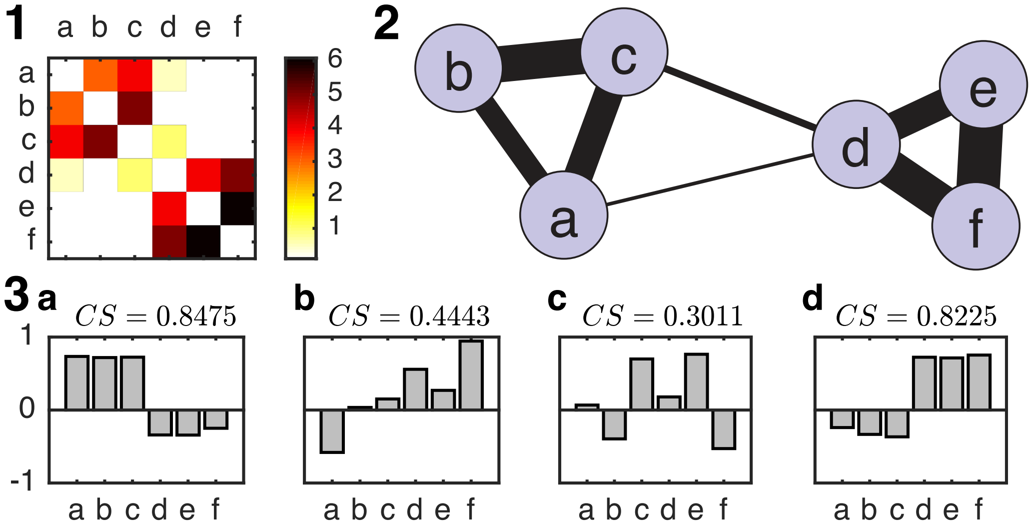

Figure 2: A hypothetical example of an agent's associative matrix R represented as (1) a heat map and as (2) a network. as well as (3) example of four preference vectors and their respective levels of constraint satisfaction, with respect this associative matrix.

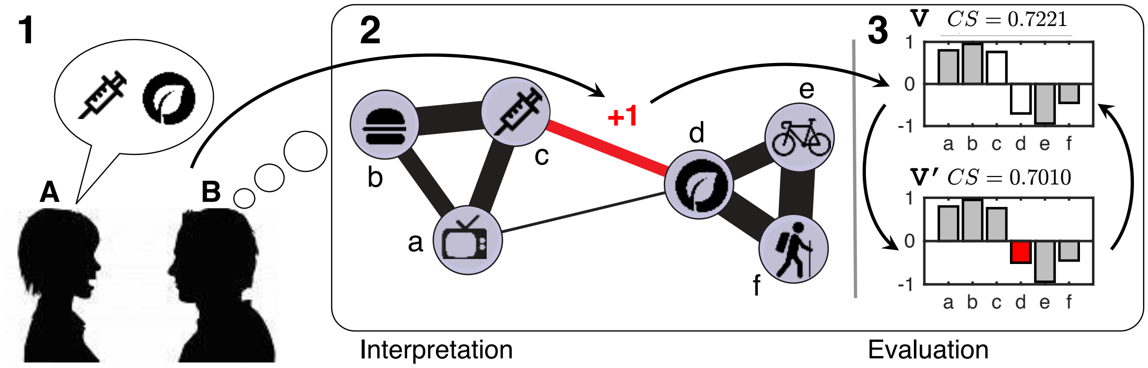

Figure 3: An illustration of the agent-based model sequence: (1) agent B observes A express support for vaccinations and organic food (practices c & d); (2) B updates the corresponding element in his associative matrix, R (the edge connecting nodes c & d in the network representation of R); and (3) randomly updates his preference for organic food (practice d, resulting in preference vector V'), which is the weaker preference of the pair {c,d} in his preference vector V. Because constraint satisfaction is reduced from 0.7221 to 0.7010, this preference update is rejected, and B's preference vector V remains unchanged.

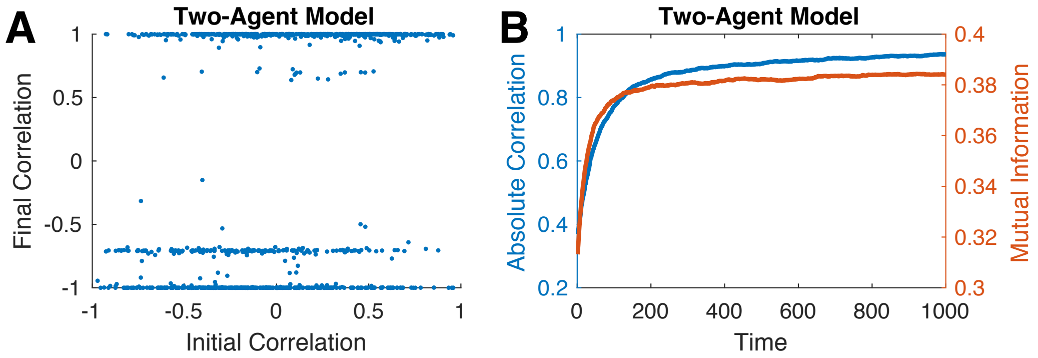

Figure 4: Two agent model with $K = 6$ and $t = 1,000$: (A) Final Pearson correlation between agents' preference vectors as a function of their initial correlation. (B) Absolute correlation between preference vectors (blue) and mutual information between the behaviors performed by each agent (red), as a function of time, averaged across all simulations.

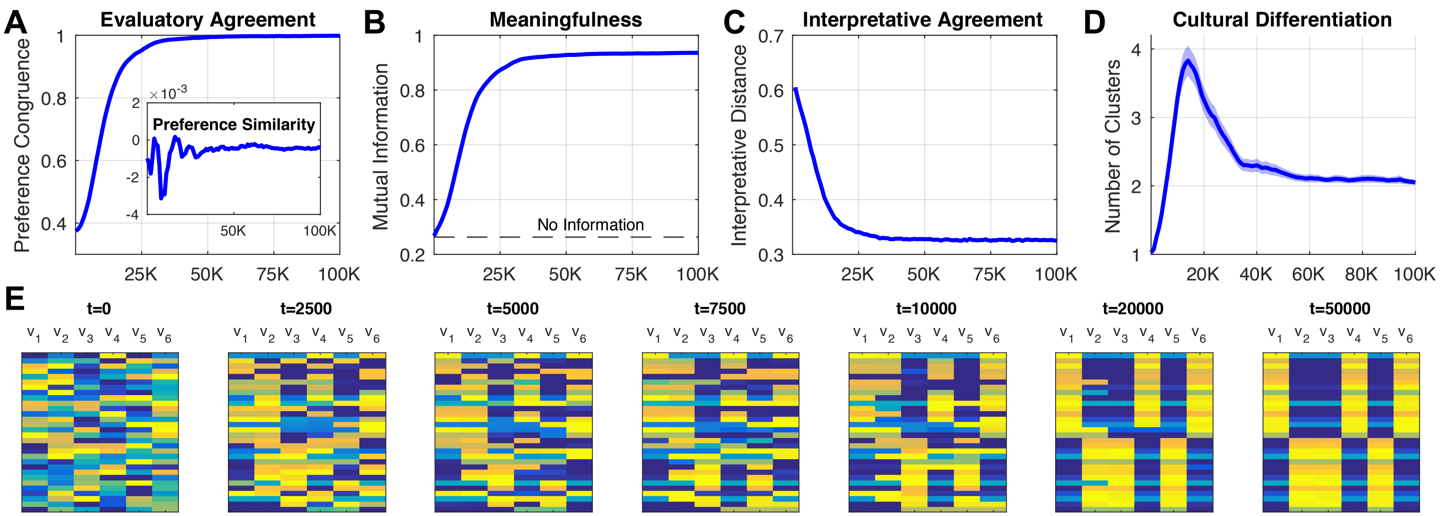

Figure 5: Multiagent models with 30 agents. (A) Mean preference congruence between agents (measured as absolute correlation between agents' preference vectors), preference similarity (measured as mean correlation between agents' preference vectors) is in the inset); (B) Mutual information between agents' behaviors; (C) Mean distance between all agents' associative matrices; (D) number of agent clusters estimated by the gap statistic, based on agents' preferences (with shaded confidence intervals); (E) snapshots of preference vectors for one simulation run (each heatmap represents the preferences of 30 agents for 6 practices, ranging from strong negative in blue to strong positive in yellow).

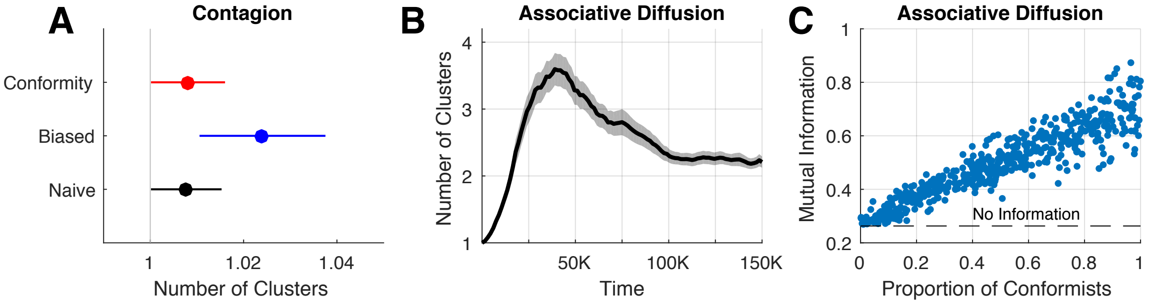

Figure 6: Alternative contagion models. (A) Number of clusters at end for contagion models with different transmission mechanisms; (B) Number of clusters for associative diffusion model with conformity; (C) Mutual information between behaviors at end for associative diffusion model with conformity and with varying proportions of conformists.

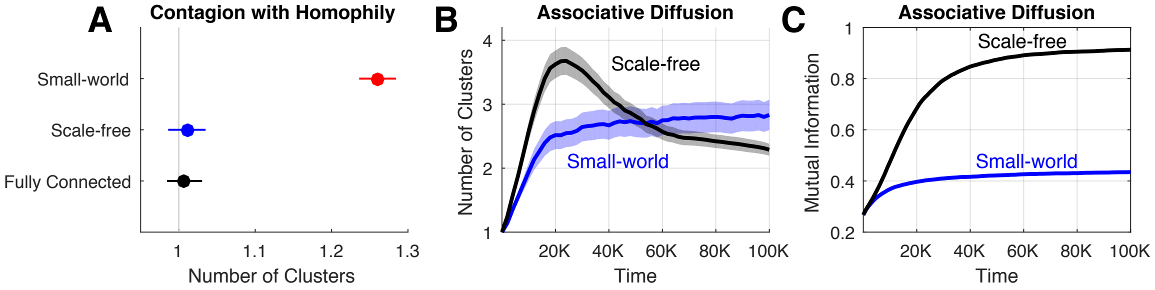

Figure 7: Different network topologies. (A) Number of clusters at end for contagion models with homophily and different network topologies; (B) Number of clusters for associative diffusion model with scale-free or small-world networks; (C) Mutual information between behaviors for associative diffusion model with scale-free or small-world networks.

A Measurement

Calculating Mutual Information

We use mutual information to measure agents' behavioral convergence, and interpret it as an indicator of the meaningfulness of practices. We calculate the mutual information at the population level as the mutual information between two behaviors performed by an agent randomly drawn from the population.

The expected behavioral probabilities can be analytically derived. To calculate the mutual information between the two practices that agents enact we need to calculate the marginal probabilities of choosing each practice as the first and second practice respectively, as well as the joint probability of choosing both practices in sequence (see eq. 7). Because agents are constrained to choose two different practices, these probabilities are not independent.

Let b_1 and b_2 denote the first and second practices enacted by a random agent. For a random agent i, $P_i(b_1 = x)$ denotes the probability that the first practice she exhibits is practice x. This probability equals the agent's baseline probability of choosing x, $P_i(x)$, and is given by the agent's preference for that practice (as defined by eq. 1 or eq. 9, depending on whether or not agent conformity is taken into account).

The probability that agent i chooses y as the second practice to exhibit, $P_i(b_2 = y)$, is conditional on the practice chosen as b_1. This probability is given by:

$$P_i(b_2 = y) = \sum_{x \in K, x \neq y} P_i(b_1 = x, b_2 = y) \tag{A1}$$

The joint probability of agent i choosing practices x and y in sequence is given by:

$$P_i(b_1 = x, b_2 = y) = P_i(b_1 = x) P_i(b_2 = y \mid b_1 = x) \tag{A2}$$

where $P_i(b_2 = y \mid b_1 = x) = \frac{P_i(y)}{1 - P_i(x)}$. Because the model restricts agents to choose two different practices at each iteration, $P_i(b_1 = x, b_2 = x) = 0$.

Agents are drawn uniformly at random from the population. Consequently, the probability that a random agent will enact a practice is equal to the mean probability over all agents. For example, the probability that a random agent chooses practice y as the second practice is $P(b_2 = y) = \frac{1}{N} \sum_{i \in N} P_i(b_2 = y)$, and the joint probability that a random agent enacts the sequence x, y is $P(b_1 = x, b_2 = y) = \frac{1}{N}\sum_{i \in N} P_i(b_1 = x, b_2 = y)$. These aggregate probabilities represent the probabilities experienced by an observer randomly observing agents in the population. Overall, the mutual information between the behaviors of the agents comprising the population are given by:

$$I(b_1, b_2) = \sum_{x \in b_1} \sum_{y \in b_2} P(b_1 = x, b_2 = y) \log \frac{P(b_1 = x, b_2 = y)}{P(b_1 = x)P(b_2 = y)} \tag{A3}$$

We define "no information" as the mutual information between behaviors if agents randomly perform behaviors irrespective of their preferences (subject to the restriction that they perform two behaviors consecutively). Under such conditions, the marginal probability of choosing a practice either as the first or second choice is $P(b_1 = x) = P(b_2 = y) = \frac{1}{K}$, and the joint probability is $P(b_1 = x, b_2 = y) = \frac{1}{K(K-1)}$. The mutual information is therefore $I(b_1, b_2)_{no\ information} = \log \frac{K}{K-1}$.

Estimating Number of Agent Clusters

As the computational simulations unfold, we seek to partition the population of agents into an optimal number of clusters such that preference pattern similarities between agents are maximized within cluster and minimized between clusters. For each pair of agents A and B we use $1 - \rho(V_A, V_B)$, where $\rho(\cdot, \cdot)$ is the Pearson correlation coefficient, as the distance metric between agents. The closer the correlation between agents' preference vectors is to 1, the closer their distance is to 0.

Estimating the optimal number of clusters in a population is computationally difficult (formally, it is an NP-hard problem). We use a common partitioning method, K-means, to find these clusters. Given a number of clusters, k, the K-means algorithm initializes k cluster centroids and iteratively adjusts cluster membership by assigning observations to the cluster whose centroid they are closest to (for more details, see Leskovec, Rajaraman and Ullman [2014]). The algorithm is efficient but non-deterministic.

Estimating the correct number of clusters is not a trivial task. Unless observations in a dataset are identical, increasing the number of clusters by 1 monotonically reduces within-cluster distance even if the data are randomly distributed. The "true" number of clusters is the maximal number that reduces within-cluster distance more than would be monotonically gained merely by increasing the number of clusters. We use the gap statistic (Tibshirani et al. 2001) to estimate this number. Using K-means we produce partitions with number of clusters ranging from 1 to $2K$ (twice the number of practices), and use the gap statistic to estimate the optimal partition.

The gap statistic computes partition compactness, W_k, for a partition into k clusters, which equals the normalized sum of distances between observations in each class. Formally:

$$W_k = \sum_{r=1}^{k} \frac{1}{2N_r} D_r \tag{A4}$$

where k is the number of classes, N_r is the size of class r, and D_r is the sum of pairwise distances between observations in r. We use $1 - \rho(V_A, V_B)$ as the distance between two agents, A and B. The gap statistic method compares the observed compactness to that obtained from a null reference distribution:

$$Gap_N(k) = E_N^*\{\log W_k\} - \log W_k \tag{A5}$$

where E_N^* denotes expectation under a sample size N. The optimal number of clusters is the smallest k that satisfies:

$$Gap_N(k) \geq Gap_N(k+1) - s_{k+1} \tag{A6}$$

where s_{k+1} is the standard error of compactness over the reference distribution. To obtain the null reference distribution, we generate 100 reference datasets where agent preferences are generated from a uniform distribution over a box aligned with the principal component of the data. For details on how this box is constructed, see Tibshirani et al. (2001).

B Path Dependence

What causes agents to gravitate in one direction over the other? In this section we demonstrate that agents' preferences at equilibrium are path dependent and that they are not merely determined by agents' initial random preferences. The results plotted in Panel A of Figure 4 already point in that direction. The figure plots the final correlation between agents' preferences in a two-agent simulation, as a function of the initial correlation between their randomly generated preferences. As the plot clearly demonstrates, the final correlation is not determined by the initial correlation, except in extreme cases where the initial correlation is either strongly positive or strongly negative.

As illustration that initial correlation between agents' preferences does not determine the final correlation, we plot three randomly selected runs of the two-agent model in Panel A of Figure A1. The diagram plots the inter-agent preference correlation as a function of time (we plot only the first 200 iterations of the model for visualization purposes as correlations tend to lock in beyond that point). The correlation patterns follow an erratic non-linear path early on, often moving between negative and positive values. In one run, for example, the correlation increases beyond 0.5 before changing course and dropping toward -1. Eventually, all correlation patterns settle on a steady state once correlation nears 1 or -1.

What determines these changes in inter-agent preference correlations? Panel B plots the magnitude of change in the inter-agent preference correlation between two subsequent model iterations as a function of the magnitude of change in agents' preferences (on a log scale). As we described earlier, agents update their preferences in reaction to other agents' behaviors. The magnitude and direction of this update is random, drawn from a normal distribution. Agents retain this update only if it does not decrease constraint satisfaction. As the plot in Panel B demonstrates, shifts in preference correlations are almost entirely driven by the magnitude of changes in agents' preferences (r=0.866). In other words, agents' stochastic preference updating behavior drives changes in their congruence with others.

Finally, to demonstrate that our results are not driven by initial inter-agent preference correlations, we ran a series of simulations of the two-agent model where agents' preferences are initialized to be zero for all preferences (namely, agents begin the simulation with neutral preferences, and all agents have the exact same preferences). Panel C plots the proportion of negative and positive final inter-agent preference correlations (based on 1,000 simulations). As it illustrates, half of the simulations tilt toward preference similarity, and half toward preference opposition. Thus, even when agents begin with identical preferences the system still evolves stochastically. As the examples in Panel A illustrate, this stochasticity is not merely a function of the first several steps of the model.

Together, these three analyses demonstrate that our results are not simply determined by initial random preferences. Rather, the evolution of preferences and their inter-agent correlations is path dependent, driven predominantly by stochastic preference updates that are motivated by agents' desire to increase constraint satisfaction. Initial preferences determine outcomes only when they are extremely strongly correlated.

C Alternative Specifications

In the main paper we report the results of simulations modeling a variety of alternatives to associative diffusion, using different network topologies. We demonstrate that these alternatives cannot explain the emergence of cultural variation unless a segregated small-world network structure is assumed. In this section we provide details on how we implement these different contagion mechanisms and network topologies.

Naive Contagion

All interpersonal transmission mechanisms assume that contagion occurs when agent B's preference for practice i changes as a function of agent A's preference, as described in eq. 8. The basic interpersonal transmission mechanism, which we refer to as naive contagion, occurs when $V_{Bi}(t+1) = V_{Ai}(t)$. Most diffusion models in the literature assume that contagion is perfectly naive.

The diffusion dynamics generated by this simple contagion mechanism are uninteresting for our purposes given that, in the absence of structural barriers to diffusion, they will always lead to complete cultural homogeneity. We therefore introduce two additions to the transmission model. First, and drawing on existing literature (e.g. Friedkin and Johnsen 1990; Dandekar et al. 2013), we assume that a social susceptibility parameter α, ranging from 0 to 1, determines the extent to which agents are susceptible to influence by others' behaviors. When $\alpha = 0$ agents are not affected by others' behaviors, whereas when $\alpha = 1$ they fully adapt their preferences to others'.

Second, we assume that agents are unaware of others' private preferences. Rather, they observe others' behaviors, and make inferences about their preferences. We define γ as the standard inference that an agent makes about another agent's private preference for a practice i when she observes that agent performing practice i. Together, we define naive contagion as:

$$V_{Bi}(t+1) = (1-\alpha)V_{Bi}(t) + \alpha\gamma \tag{A7}$$

When $\gamma = V_{Ai}(t)$ agents have full knowledge of their interlocutors' preferences. Such a model always leads to full preference convergence. When γ is fixed for agents, namely when agents always infer that another agent's preference is fixed, all preferences eventually (and unsurprisingly) converge toward γ.

To add stochasticity, we assume that γ is randomly and uniformly drawn from the range $[0.1, 1]$, i.e. that agents randomly infer other agents' preferences. The results plotted in Panel A of Figure 6 are based on a specification of $\alpha = 0.5$ and random γ. They are robust to different positive values of α.

In the specifications that follow, we always assume that γ ranges from 0.5 to 1 (i.e., agents infer a moderate to strong preference). The results reported in the main text are robust to this assumption and are reproduced when we assume that agents have full access to others' preferences. Nevertheless, we believe that a model that assumes preference inference is more realistic than one in which agents are assumed to have full access to others' preferences, especially when interaction is assumed to be superficial, as is the case in our model.

Biased Contagion

In the biased contagion condition we aspire to model a process whereby agent B's preexisting preferences mediate the effects of A's behaviors. Following Dandekar et al. (2013), we implement biased contagion as a function of B's prior preference for i, weighted by a bias parameter $\beta > 0$. β defines the extent to which B's existing preferences mediate social transmission from A. As long as $\beta > 1$, bias is positive. Like Dandekar et al., we define biased contagion as a ratio between A's positive and negative effects on B's preference for i, weighted by β, as follows:

$$\check{V}_{Bi}(t+1) = \frac{(1-\alpha)\hat{V}_{Bi}(t) + \alpha\hat{V}_{Bi}(t)^{\beta}\gamma}{(1-\alpha) + \alpha\hat{V}_{Bi}(t)^{\beta}\gamma + \alpha(1-\hat{V}_{Bi}(t))^{\beta}(1-\gamma)} \tag{A8}$$

where α is again a social susceptibility parameter ranging from 0 to 1. In the results reported in the main text we assume that $\alpha = 0.5$, but these results are robust to different values of α as long as it is reasonably above 0 (roughly $\alpha > 0.1$) such that some social influence occurs.

Because the effect of bias is implemented as an exponentiation of B's existing preference, following Dandekar et al. we transform this preference to a 0 to 1 range using the logistic function, such that $\hat{V}_{Bi}(t) = \frac{1}{1+e^{-V_{Bi}(t)}}$. We then transform $\check{V}_{Bi}(t+1)$ back into an infinite range using the logit function, $V_{Bi}(t+1) = \log \frac{\check{V}_{Bi}(t+1)}{1-\check{V}_{Bi}(t+1)}$ so that it takes negative and positive values, in compliance with our model's assumption about the range of preference values. This functional form has the desired behavior, such that values above 0 for B's prior preference lead to a growing positive effect on that preference, and those below 0 to a growing negative effect. The results we report in Figure 6 are robust to different values of γ and β.

Conformist Contagion

In the conformist contagion condition we seek to model a process where B's preference for i is mediated by B's taste for popularity and her perception of practice i's rarity. To do so, we define two additional parameters. First, we define ω_B as $B's$ taste for popularity, ranging from 0 to 1. We assign agents with a random taste for popularity, drawn from the inverse of a log normal distribution with a mean of $\log 0.15$ and standard deviation of $\log 2$. This ensures that the majority of agents are conformist (with half of agents having a taste for popularity at or greater than 0.85), and a minority nonconformist. Second, we define B's perception of practice i's rarity as function of how frequently she had observed other agents performing that practice. We therefore define for each agent a K-sized vector O_B which is initialized to 0, and where the value of cell i increases by 1 whenever B observes that practice enacted by others. Values in O decay as a function of a decay parameter λ. We can now define $\psi_{Bi} = 1 - \frac{O_{Bi}}{\max(O_B)}$ as B's perception of practice i's rarity.

Building on Flache and Macy (2011), we define conformist contagion as follows:

$$V_{Bi}(t+1) = V_{Bi}(t) + (2 \cdot |\omega_B - \psi_{Bi}| - 1)\gamma \tag{A9}$$

This mechanism of contagion ensures that B's preference changes as function of the distance between her taste for popularity and her perception of the practice's rarity, $|\omega_B - \psi_{Bi}|$. As that distance nears 1, that is, as the congruence between the practice's perceived rarity and the agent's taste for popularity grows, B increases her preference for practice i. As the distance nears 0, B decreases her preference for i. The results reported in Figure 6 are robust to different values for γ, as long as it is positive.

We also consider an alternative method of assigning tastes for popularity, where we dichotomously divide the population into conformists and nonconformists. We define the tastes for popularity for these two conditions as $\omega = 0.95$ (conformist) and $\omega = 0.05$ (nonconformist), and randomly assign 75% and 25% to each of these two conditions, respectively. Unlike the log-normal method which generates a skewed distribution of taste for popularity, this dichotomous method generates a bimodal distribution. The results reported in Panel A of Figure 6 are robust to this specification, suggesting that even when there is a clear division into conformists and nonconformists (such as in the case of early adopters who are distinctively and qualitatively different from mainstream audiences), there does not emerge a division into different cultural clusters.

In Panels B and C of Figure 6 we explore the effects of extending the baseline associative diffusion model to account for variation in conformity (as specified in eq. 9). Panel B reports the results of this model, where taste for popularity is generated using the log-normal method as detailed above. In Panel C, we run multiple simulations of the extended associative diffusion model where we vary the overall prevalence of conformity in each simulation. To do so, we use the dichotomous method for generating taste for popularity, and vary the proportion of conformists in each simulated run. For example, when the proportion of conformists is 0.6, 60% of agents have a high taste for popularity at $\omega = 0.95$ and the remaining 40% have a low taste for popularity at $\omega = 0.05$. As the diagram illustrates, cultural differentiation emerges as long as the proportion of conformists is greater than roughly 0.15.

Homophily

In the homophilous contagion condition we seek to model a process where B's change in preference for i is mediated by B's perceived homophily with A. Consistent with our assumption of superficial interaction, we assume that B only has partial information about A's other preferences. Specifically, B observes A perform only one additional practice, j. We define B's homophily with A as the perceived similarity between their preferences for j.

Diffusion models that take into account the effects of cultural similarity on adoption normally calculate this similarity in Euclidean space (e.g. Baldassarri and Bearman 2007). We build our implementation of homophily on DellaPosta et al. (2015). DellaPosta and colleagues' model calculates the similarity between agents B and A as the difference between two distances: the Euclidean distance between the agents and the expected distance between two random agents drawn from the population. The probability of social influence, or the likelihood that B will adopt A's preference, is proportional to the magnitude of that difference. In other words, the stronger the agents' similarity or dissimilarity (relative to what would be expected at random), the greater the likelihood of social influence. When the agents are dissimilar (i.e. their distance is greater than expected) adoption is negative. But because negative influence is rarer than positive influence, it randomly occurs only in 10% of cases.

We adapt this model to our setting, where preferences range from negative to positive values and where preferences are only partially observable. Although, unlike in DellaPosta et al' model, our agents do not observe others' full set of behaviors, we do assume that they are aware of the private preferences for the behaviors they observe. We therefore define the social influence of A on B as the inverse of the absolute distance between their preferences for practice j:

$$W_{Bj,Aj}(t) = 1 - ||V_{Bj}(t)| - |V_{Aj}(t)|| \tag{A10}$$

Because preferences are initially drawn from the range -1 to 1, $W_{Bj,Aj}$ ranges from 0 to 1. Eq. [A11] ensures that preferences remain within the -1 to 1 range. We define homophilous contagion as:

$$V_{Bi}(t+1) = (1 - W_{Bj,Aj})V_{Bi}(t) + W_{Bj,Aj}V_{Ai}(t) \tag{A11}$$

When A's and B's preferences for j are differently signed, i.e. one has a positive and the other a negative preference, we define homophilous contagion as:

$$V_{Bi}(t+1) = (1 - W_{Bj,Aj})V_{Bi}(t) - W_{Bj,Aj}V_{Ai}(t) \tag{A12}$$

and allow such influence to occur only in 10% of cases, as per DellaPosta et al. (2015).

This implementation ensures that, like in DellaPosta et al's model, equally similar and dissimilar agents have influence of the same magnitude, but in opposite directions. Panel A of Figure 7 reports results using this model.

To ensure that the results reported in Figure 7 are not driven by specific assumptions, we also examine an alternative specification for the homophilous contagion process, which builds on and extends Flache and Macy (2011). In this specification we assume that agents do not have access to interlocutors' private preferences. We define alternative homophilous contagion as follows:

$$V_{Bi}(t+1) = V_{Bi}(t) + [2(1-\Delta_{Bj}(t)) - 1]\Delta_{Bi}(t) \tag{A13}$$

where $\Delta_{Bi}(t) = \gamma - \hat{V}_{Bi}(t)$ is the distance between γ, which is what B infers as A's preference for an enacted practice, and B's own preference for that practice (transformed to the 0 to 1 range as explained above, to comply with Flache and Macy's (2011) model). This functional form ensures that as B's preference for j grows closer to her inference about A's preference, she updates her preference for i to be increasingly identical to her inference about A's preference. When her preference for j is significantly lower than γ, this update rule means that she decreases her preference for i. Results using this specification replicate the results reported in Figure 7.

Network Topologies

In addition to exploring the effects of alternative contagion mechanisms, we also explore three different network topologies: fully connected network, scale-free network and small-world network. We implement these different topologies as directed graphs (with no self-edges). In each simulation round one observer agent, B, is randomly selected with uniform probability. The actor agent, A, is selected with uniform probabillity from the subset of agents to whom B has an outgoing edge. In this subsection we explain how the three network topologies are generated.

The fully connected topology is a graph in which all potential edges are realized (with the exception of self-edges connecting a node to itself). Such a network implies that all edges have a uniform probability (equal to $\frac{1}{N(N-1)}$) of being selected. This is equivalent to selecting an observer and actor with uniform random probability.

A scale-free network is one in which node indegree follows a power law distribution such that the probability of nodes with k incoming edges, $P(k) \sim k^{-\alpha}$. We generate networks where α ranges from 2 to 3 and where each node has an outdegree of 6. To generate such a network, we randomly assign all nodes with a popularity score that follows a power law distribution. We then iterate over all nodes, and assign them with 6 random outgoing edges to other nodes with a probability proportional to these nodes' popularity. Such a process generates a network wherein each agent can observe only 6 other agents, but agents vary significantly in how many agents can observe them.

A small-world network topology is one in which nodes are segregated into clusters. Following Watts (1999), we implement a connected caveman topology with 5 clusters. To do so, we randomly divide the network into 5 equally sized and fully connected cliques. We then randomly rewire 10% of the edges in the network, by randomly selecting two edges and swapping their destination nodes. This generates edges that bridge between different cliques. The procedure we follow is similar to the procedure in DellaPosta et al. (2015).

An important feature of the network generation processes that we implement is that both the scale-free and small-world networks have the same overall number of edges, and the same node outdegrees. That is, in both types of networks each agent can observe the same number of other agents. These networks differ, however, in how these edges are distributed. In the scale-free topology a small number of agents account for the majority of indegrees. In the small-world topology indegrees are equally distributed.

Appendix Figures

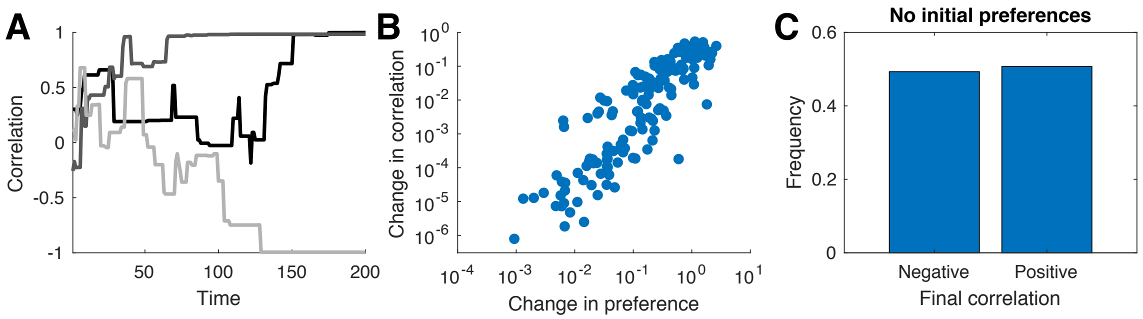

Figure A1: Path dependence in the two agent model. (A) Examples of the evolution of inter-agent preference correlation from three random simulation runs. (B) Changes in inter-agent preference correlation as a function of stochastic preference updates. (C) The proportion of negative and positive final inter-agent preference correlations when agents are initialized to have uniform 0 preferences.