From Meaning to Behavior: Mental Representation, the Patterning of Social Life, and Cultural Analysis

Lynette Shaw, Nina Lyn Cesare, and Michael Hughes Esposito

This paper looks at the relationship between mental representation processes and the emergence of profiles of human action and statement that are detectable via methods currently favored "big data" analyses. In particular, this paper begins with the role of mental networks of association in shaping not only humans' sense-making in situations but also in generating their repertories of observable behaviors. After delimiting the basic mechanics of this process, this work then considers the implications of it for classic sociological issues of polysemy, socially shared (i.e. collective) vs. idiosyncratic representations, and representational/cultural change. The paper then uses the examples of cluster analysis, multiple correspondence and principle component analysis, and topic modeling to consider how this link between behavior and mental representations can systematically increase the analytical leverage of our existing approaches to the empirical study of culture. In the final section, the paper then offers a demonstration of how this perspective can be used to motivate and guide a "data mining for culture" approach via the example of using Yelp reviews to identify shared patterns of consumption and speech in large-scale social data.

There are many different approaches that can be applied to the question of how social life comes to be ordered. In some cases, the emphasis may be placed on the self-organization of collectives that emerges unintentionally from individuals' rational pursuit of their own self-interest. In other situations, the focus may be more appropriately placed on the causal role that macro-level, structural forces that shape the opportunities of individuals within a society. Historically, the sociology of culture has taken a still different tactic by emphasizing the key role of meaning in individuals' experiences and actions and their relationships to the larger symbolic orders within which they live. While these various ways of accounting for different elements of the social order are all likely to remain important to the work of sociologists, recent models that have emerged at the intersection of cognition and culture (Dimaggio, 1997; Cerulo, 2002; Cerulo, 2010) are granting a new level of relevance to this third class of explanation.

The present work will seek to contribute to this growing subfield of culture and cognition research by offering a new model of how the associative structure of mental representations are causally linked to the existence of patterns in social life. Specifically, it will show how the cognitive associations that are underlie individuals' automatically imposed interpretations of their experiences can lead to the emergence of particular profiles of behavior and statement that characterize cultural groups. Furthermore, it will also demonstrate how these processes of mental representation lead to a new understanding of polysemy, deviations and conformance to shared profiles of behavior, and the change of shared representations and behavioral profiles through time. In the course of forging this link between the associative structures of mental representation and the patterning of social life, this paper will not only draw connections between existing cultural models in sociology and contemporary cognition research, but will also go further to consider the implications of this connection for the empirical study of culture. Specifically, it will demonstrate the natural convergence between how this cognitively driven patterning of social life manifests as emergent structure in social data and a suite of modern computational and statistical methodologies which have been built specifically to identify and track such structure in data.

In addition to its theoretical contribution, this paper give an example of how this approach can be used to motivate and guide cultural analyses involving large amounts of social media data. Social media spaces - defined as "web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system" (Boyd and Ellison, 2007) - provide unique, bounded social contexts in which researchers can watch patterns of individual and collective behavior emerge and evolve in real time. This study will demonstrate how cultural researchers can extract such patterns via a multi-faceted analysis of user reviews from the social media platform Yelp. The primary goal of this analysis will be to explore the ways in which Yelp reviews may be used to identify groups of users who exhibit a shared patterning of their consumption choices as well as their behaviors and statements within Yelp.

Meaning-making and Mental Representation

In order for the mind to get from the initial perception of a situation to a coherent understanding of it, a huge amount of interpretive work is required. Even those parts of experience which seem to be self-evident and obvious are, in fact, the products of a rapidly executed set of unconscious, cognitive processes whose purpose it is to take perceived stimuli and immediately situate it with respect to an individual's lived and learned history of experiences (Wyer, 2007; Kahneman, 2011; Payne & Cameron, 2013; Carlston, 2010; Morewedge & Kahneman, 2010). This linking of one's present moment to similar past experience then allows the mind, through a process known as "spreading activation" (Collins & Loftus, 1975; Damasio, 1989), to quickly access the set of other past experiences and understandings that have previously co-occurred with this sort of stimuli. This activation of related concepts or experience is what allows individuals to automatically and unconsciously go beyond the immediate stimulus and generate a set of inferences about what is occurring and what will happen next. Essentially, it is through a fast mapping of what is happening now back onto a set of associations that has been built during similar past moments which allows the mind to create meaning and understanding of the present moment in a seemingly immediate fashion.

Study of these types of interpretive processes and the mental structures they rely upon has fallen under a variety of headings including “mental models” (Holland, Holyoak, Nisbett, & Thagard, 1986), “categories” (Allport, 1954), “theories” (Murphy & Medin, 1985), and “schema” (Bartlett, 1932; Taylor & Crocker, 1981; D'Andrade, 1995). One of the most general terms that has been used for these types of cognitive structures and their related processes is that of mental representation (Payne & Cameron, 2013; Wyer, 2007). The important role that mental representation type structures play in social life has been identified by many (Carlston, 2010; Payne & Cameron, 2013). The critical place of these cognitive processes in our understanding of models of cultural phenomena has also been identified (Dimaggio, 1997; Cerulo, 2002; Martin, 2010) and the subfield of cognition and culture seems exceptionally well poised to offer a new set of insights into collective sense-making processes based on this understanding of meaning construction at the individual level.

Mental Representation and the Patterning of Social Life

As previously discussed, mental representations are built from networks of pre-established networks of associations that become activated when stimuli in the environment are matched up or likened to similar past experiences. This activation makes it possible for the mind to go beyond the immediate sensations of an experience in order to make association-based, inferential leaps about not only what is happening and what will happen, but also how to think about and respond to what is occurring (Bar, 2007). A simplified schematic of this process can be created as follows:

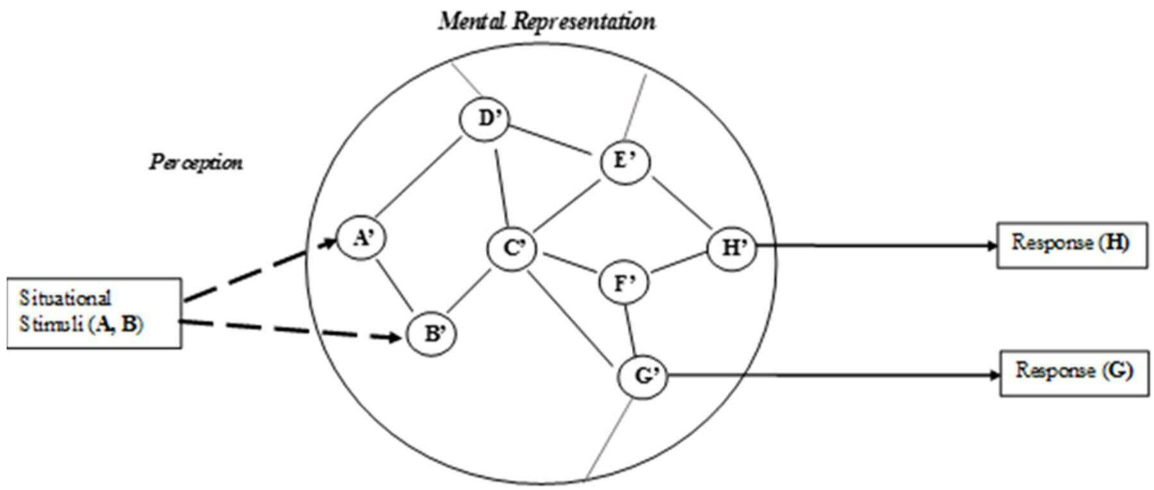

Figure 1 – *Interpretation and response formation via mental representation*

This figure can be interpreted as follows. First, a set of situational stimuli (**A** and **B** in Figure 1) are perceived by the individual and are then automatically and unconsciously related back to previously experienced stimuli that resemble what is presently being perceived (**A'** and **B'** above) . For example, we could think of an individual walking into a coffee shop she has never previously visited, seeing the counter and barista, hearing an espresso machine, and smelling coffee. These stimuli are linked back to prior experiences of coffee shop-related stimuli which generally resemble, though are not exactly the same as, what is being perceived in the present moment. This cuing of the individual's "coffee shop representation" then leads to the spreading activation of other associated concepts and experiences[1] that have historically co-

[1] This is a simplification. It is more likely that it brings to mind not specific past experiences but a generalized "prototype" that has been built out of the common elements of past experience with similar stimuli (Bar, 2007).

occurred with the sort of stimuli currently being experienced but which have not yet been directly experienced in the moment (e.g. **C'**, **D'**, **E'**, etc. in Figure 1). In this example, this process might "bring to mind" an understanding that the individuals in line near the counter are customers who are waiting to place an order and lead to the expectation that if she herself joins the line, then she will also be able to place an order for the sort of beverages or food that are usually served at coffee shops. These understandings and expectations are able to arise near instantaneously in the individual's awareness, not as a matter of conscious, deliberative consideration about the abstract function and purpose of coffee shops and their component parts, but because the mind unconsciously and immediately anticipates what is happening now and what will happen soon based on what things have co-occurred in past experience. It is through this set of associations that an individual not only forms her interpretation of the situation, but also is setup to enact the set of responses that go along with this interpretation (shown in Figure 1 by the transition from previously associated actions **G'** and **H'** to enacted versions of those actions in the present situation, **G** and **H**).

To be clear, this model does not obviate the place of conscious consideration, deliberative attention, or strategic action. To further extend the above example, the individual may not have to consciously figure out how she can place an order but she may very well consciously deliberate upon what she wants to drink and whether or not she will get a pastry. Knowing that this is an aspect of the situation to which she should consciously attend, however, is still buttressed by the automatically generated sets of expectations and understandings she relies upon to make sense of her experience. As such, we should not understand mental representation as completely determining individuals' behaviors, though in the scenario of a very habitual or rote activity that requires little if any conscious attention, this may effectively be the case (Lisman &

Sternberg, 2013). Instead, we would be better off looking at representations as essentially structuring individuals' interpretations and expectations of situations in such a way that it greatly narrows the set of actions and statements they are likely to respond with from an innumerable number of options down to a relatively small set of most probable behaviors (see Figure 2). In the case of this illustrative example, we may not be able to anticipate whether any given individual will order a cappuccino or a latte when she enters a coffee shop, but we can usually have a fair degree of certainty that this will be the sort of decision she is thinking about and not whether she will be reciting a passage from Twain or Dickens when she goes to order.

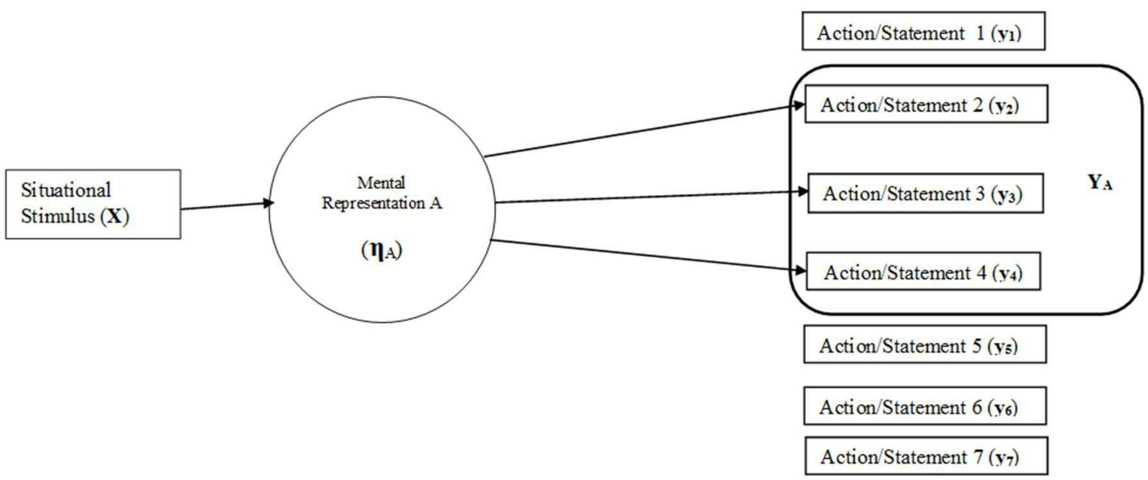

Figure 2 – *Linking of behavioral profiles to particular mental representation (via the same process shown in Figure 1)*

The canalizing of human behavior that occurs due to the imposition of mental representations is key to both unpacking some of the cognitive origins of the patterning of social life and to developing a new type of approach to empirically assessing the enactment of mental representations in daily life. According to this model, mental representations' reduction of possible behaviors into particular sets or profiles of actions and statements that "hang together" (represented as $\mathbf{Y_A}$ in Figure 2) can be treated as observable indicators of those unseen mental representations. Said differently, we should view patterning in observed behaviors as a signal of an underlying network of activated associations that the individual is applying. Recognition of the relationship between mental representation and patterned behavior is a great start to establishing an empirical approach to assessing individual's unseen sense-making processes. A few significant outstanding issues need to be addressed, however, in order to actually be able to start making use of this insight in social research.

Polysemy

The first issue, the polysemous nature of experience, is one that should be familiar to most sociologists interested in culture. Polysemy is the insight that any given event or object can house within it multiple meanings simultaneously. The determination of which one of the various meanings that come to the forefront at a given moment is thought to be dependent upon both the individuals experiencing the object or event and the contextual factors surrounding it. A direct analog to this phenomena exists within the context of mental representation imposition. For any given stimulus or set of stimuli, there is not one but many different associational networks that might potentially be activated (Feldman, 2006; Carlston, 2010). This is true both within individuals and across them. For any given person, the smell of coffee may be associated with coffee shops, but it also might be related to morning breakfasts at home or late nights working in the office. For a different person, say the barista, it might be associated with one's workplace or for a connoisseur, might be related to a whole body of assessments that can be made regarding quality and type of coffee. A deep exploration of the processes by which one of these mental representations comes to be applied over another one is, unfortunately, beyond the scope of this present work. It is sufficient for these purposes, however, to just mention that there seems to be a "best match" principle (Feldman, 2006; Bar, 2007) at play in mental representation imposition via which the representation that seems to be most consistent, based on current and past information, with the set of stimuli being experienced becomes preferentially activated and the other, potentially competing representations, become inhibited[2].

Much more relevant to the present discussion than the question of which representation wins out in a given moment are the implications of this variability in mental representation imposition for observable behavioral profiles. The first of these implications is that responses to events or objects cannot be understood as directly following from individuals' experience of them but must instead be understood as being mediated by any one of a number of possible, interpretative frameworks. Determining which representation that will be at play for a person in any given moment is not only a matter of their own personal history with a given type of stimulus but also a matter of the contextual factors that surround it just prior to and during its perception (Bargh, 1982; Bargh & Morsella, 2008; Higgins, 1996).

This situation of multiple interpretive frameworks potentially being applied is arguably the one of most interest to social science researchers. Many of our questions, such as for instance "Why do individuals support or oppose the death penalty?," may very well be at least partially answered by considering the different associations that are activated for different groups when

[2] A much fuller investigation of these processes and the set of socio-cultural dynamics which may emerge from them is available elsewhere (Shaw, in press).

such a concept or situation is presented to them. Activation of a set of associations involving the necessity of harsh punishments and trust in legal authorities would lead to a very different stance on the death penalty than a set of associations emphasizing the fallibility of human institutions and the misguidedness of physical violence (for a related exploration of how these different sorts of associations lead to different political standpoints, see (Lakoff, 1996)).

Variability in representation imposition is not only relevant to differences between groups, however, but also to understanding the different cultural expressions and enactments individuals play out at different times and in different contexts. When facing a conflict at work, for instance, whether or not other contextual factors preferentially activate a more "competitive" or a more "cooperative" representation of a situation can ultimately have significant effects on not only that individual's interpretation of what is happening, but also on her eventual interactions with others in her work environment (an excellent exploration on this variability in how individuals apply "culture" is available in Swidler's (2001) research on people's approaches to the subject of love and intimate relationships).

The following figure illustrates this sort of polysemy through the potential imposition of any one of a number of available mental representations:

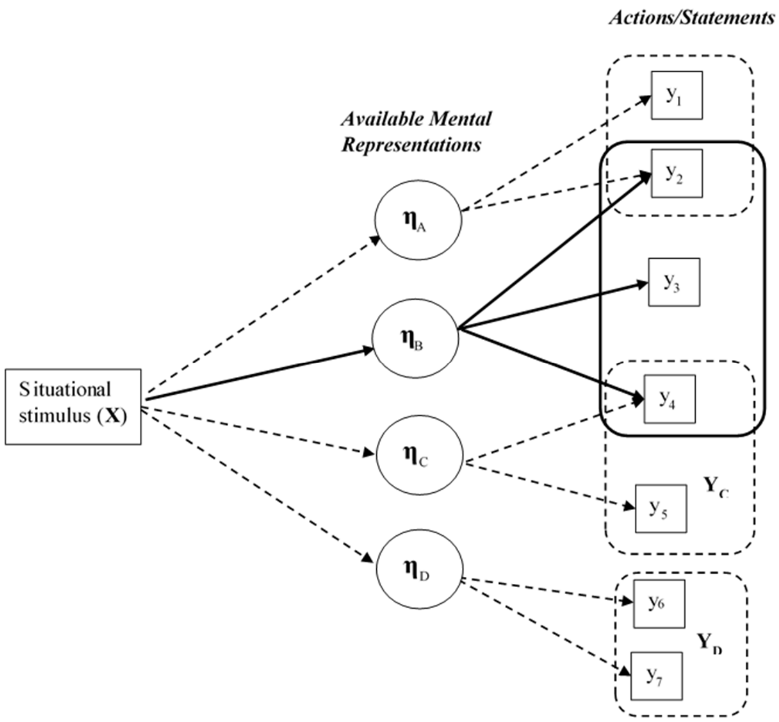

Figure 3 – *Polysemy in mental representation imposition and its relationship to behavioral profiles*

Here, we begin with an initial situational stimulus (**X**) that could potentially cue any of set of different mental representations ({$\boldsymbol{\eta}_A$, $\boldsymbol{\eta}_B$, $\boldsymbol{\eta}_C$, $\boldsymbol{\eta}_D$}). For the purposes of the present illustration, this set of mental representations could consist either of all the existing representations a given individual might be able to apply to the situation or the set of applicable representations available across a group of individuals. On the far right hand side of the illustration, we see the space of possible actions or statements that might be observed ({y_1, y_2 … y_7}). In accordance with the afore described relationship between representation imposition and profiles of observable behavior, we see how this given set of representations might be "read off" of different combinations of action and statement with which an individual might respond ({$\mathbf{Y}_A$, $\mathbf{Y}_B$, $\mathbf{Y}_C$, $\mathbf{Y}_D$}). As indicated by the bolded pathway, we see here an instance in which an individual ends up imposing the mental representation labeled $\boldsymbol{\eta}_B$ which is then indicated through their enactment of the profile of statement and action, $\mathbf{Y}_B$.

The main message of Figure 3 and the foregoing discussion is that as researchers, one of the best points of access we have into the meaning-making processes taking place inside individuals' heads is the *patterning* of their observable behaviors that arise as a result of those processes. This might seem like an obvious insight initially, but there are some important nuances to this point that are made more obvious by the above illustrations. Foremost is the realization that the way statements and actions *hang together* in the context of a behavioral profile is generally more important than any particular action or statement can be. Take for instance the hypothetical action-statements y_2 and y_4 in Figure 3. Both of these action-statements are associationally linked within representation $\boldsymbol{\eta}_B$ but are also part of $\boldsymbol{\eta}_A$ and $\boldsymbol{\eta}_C$, respectively, as well. Observing either y_2 or y_4 can narrow the set of potential representations that are suspected as being imposed, but unlike observing y_6 or y_7, it does not uniquely identify what representation is being used. In order to pinpoint the imposition of $\boldsymbol{\eta}_B$ in this stylized scenario, either y_3 would need to be observed *or* the co-occurrence of y_2 , y_3 and/or y_4.

This realization represents both a warning and an opportunity for researchers interested in empirical investigation of meaning structures. The warning comes from the realization that one must be cautious in assuming that a single statement or action can act as an indicator of an individual's representation of a situation. While such indicators are not impossible, their reliability depends on the uniqueness of their association with a particular framework of interpretation. The opportunity here, however, is a move away from the consideration of isolated behaviors in favor of a stronger emphasis on sets or *profiles* of action or statement. A given behavior might not be able to uniquely identify the imposition of a particular mental representation, but the manner in which it co-occurs with other behaviors may be able to do so. While a given behavioral observation may not make it clear as to what is going on in the mind of an individual, the manner in which that behavior hangs together with other observed behaviors might have a much greater ability to clarify the sense-making processes which are taking place.

Shared versus Idiosyncratic Representations

The associative structures which undergird mental representation emerge out of individuals' personal learning histories The linking of stimuli together in these networks arises from a sort of automatic, informal learning process in which stimuli that often co-occurs together are become cognitively associated with one another[3]. In addition to this foundational learning process, associations can also form via both more formal, conscious learning pathways and through social learning. In all cases, the formation of these associative networks depends fundamentally upon one's personal history of experience and learning. The implication of this for the present model is that realistically, no two individuals will have the *exact* same mental representation as each other. Each individual's history, no matter how similar, is still singular. As a result, no two people will ever be able to share the exact same set of associations as another.

Nevertheless, while mental representations can never be identical, they can most certainly have some very strong overlaps with each other. If one were able to look directly at the distribution of representations within a population, one would likely find a range of representations going from those which build upon associations many individuals have formed in

[3] This version of learning has been termed "Hebbeian Learning" and is captured by the axiom "neurons that fire together, wire together." (Hebb, 1949).

common to others that are based on almost entirely idiosyncratic associations. Fortunately for social science researchers, our interests are most likely to be centered on the systems of shared meaning that underlie social life and therefore, our primary concerns are likely to be in the sorts of representations which contain associative networks that are present in many people. Recognition that a particular mental representation of interest can be very similarly, but never identically, instantiated across individuals is important, however. Specifically, it leads to a further refinement of this model by identifying a persistent source of "noisiness" in the profiles of action and statement we observe. For the same reasons that we expect there to be individual variability in individual versions of a shared representation due to personal histories of learning and experience, we should also expect to see a congruent level of variability in individual enactments of the behavioral profiles that are associated with those representations. The following two figures illustrate the incorporation of this "noisiness" into this model for both the individual representation case and the "polysemy" situation discussed in the previous section.

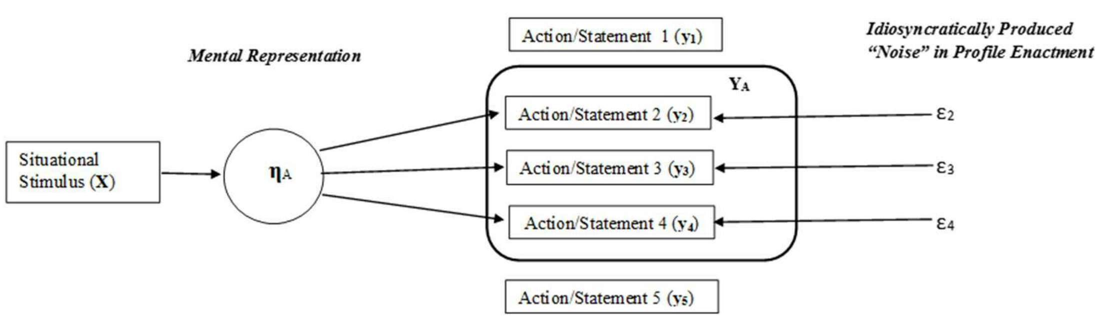

Figure 4 - Relationship of mental representation and behavioral profiles with idiosyncratically produced "noise" in enactment included

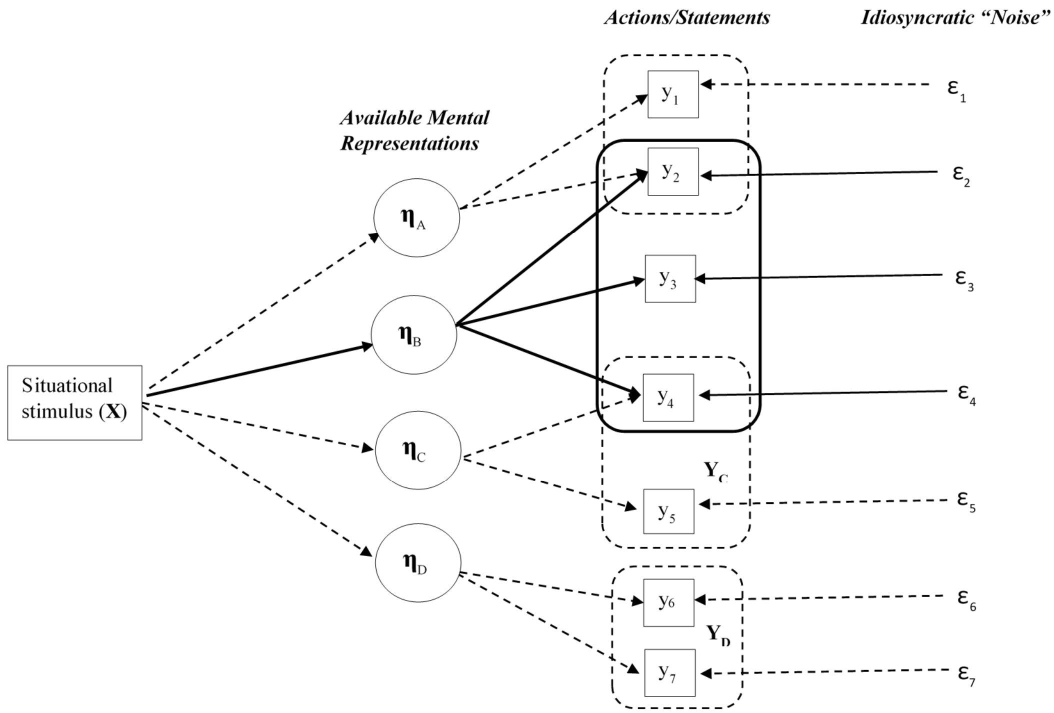

Figure 5 - Polysemy via mental representation with idiosyncratically produced "noise" in enactment included

Representational Change

The dependency of associational learning on personal experience entails a constant, background source of variability in conformance to shared representations and the related enactment of behavioral profiles. This idiosyncratic source of noise is distinguishable from another, more structural type of deviation that we should expect to find in the context of shared representational change. Again, the present treatment of the dynamics of representational change must remain relatively brief for the purposes of this work, but a deeper exploration of the mechanics responsible for representational adoption and change and its connection to larger scale cultural dynamics is available in related work (Shaw, in press).

As described in earlier discussions, a primary function of mental representation is to generate automatic inferences about what is and what will happen based on prior experiences with comparable situations. What happens, however, when these inferences are incorrect? What for instance occurs cognitively when a “coffee shop” representation incorrectly infers that there will be a counter at which to place an order but in the present situation, no counter is obvious? Parallel but much more significant examples of this scenario might be to ask what happens when one’s standing representation of an institutional order makes an inference that police will uphold a set of fundamental rights but the present actions of a police force seem to violate it or when an individual’s commonly held representation that a person who is able and willing to work will be able to achieve material security ends up failing in the face of a national economic crisis. Though the particular content of these automatically generated expectations varies widely across these illustrative examples, each one involves the disconfirmation of a representation that occurs when there is a breakdown in the ability of the established set of associations to successfully anticipate what is and will happen in the present situation.

On a cognitive level, disconfirming moments such as these can lead to varying degrees of response. For situations in which some but not most of the automatic, association-based inferences of a representation are invalidated, it is entirely possible that the minor disruption will essentially be smoothed over and the mind will continue to impose the representation in an automatic, unconscious[4] fashion (Morewedge & Kahneman, 2010; Lieberman, 2007). The larger the disruption, however, the more likely it is that the related parts of the representationally constructed understanding of the situation will move from being processed in an automatic, unconscious, even taken-for-granted way and will be pushed over into the realm of deliberative attention and processing (Lieberman, 2007). To relate this to the above examples, an individual might have to consciously ask herself "Where is the counter?" in the coffee shop scenario or in the police example, might have to start consciously reassessing what exactly the officer is doing and whether her rights will be upheld. If the disruption goes beyond a single incident and becomes a continuing source of disconfirmation, this momentary movement into consciously attending to representational content might become a more permanent state and even go so far as to initiate a period of attempting to change the affected sets of representations so that they are able to better account for the new order of things. Particularly important to this juncture is the process of "conceptual blending" (Fauconnier & Turner, 2002) which is in short, a process by which new potential understandings of situations are constructed via relating the present situation to. Relating these processes back to the example of a prolonged failure of one's representation concerning the promise of employment, as time goes on we should expect individuals facing this situation to shift into a state where they are attempting to put together a new understanding of how the obtainment material security works in the new economic circumstances. Furthermore, we should expect that the new representations and profiles of behavior which emerge from this

[4] Something interesting to consider in this regard is the degree different individuals are able to overlook varying sizes of disruption and the relationship such a characteristic might have to issues of confirmation bias (Morewedge & Kahneman, 2010).

activity to, due to conceptual blending processes, likely contain novel recombinations of the elements of existing representations.

These cognitive responses should often be reflected in individuals' observable behaviors, and specifically, it should be visible in changes that occur in the degree of their conformance to established profiles of action and statement. While the ability to overlook minor levels of disconfirmation should engender a certain robustness in individuals' conformance to established profiles of behavior, larger degrees of disconfirmation should relate to greater degrees of deviation away from the profile as individuals must creatively and deliberatively navigate aspects of the situation that were previously automatically and uncomplicatedly structured at the cognitive level. Initially, we might not expect a strong patterning in how individuals adapt to the new situation as each individual is having to figure out his or her own response to it on the fly. However, as the disruption persists and representations are created or modified to handle the seemingly now permanent altered state of affairs, we should expect a resurgence in conformance to new set of behavioral profiles. More specifically, we might even expect the new behavioral profiles to reflect some of the underlying processes of conceptual blending that have occurred in so far as we are able to identify particular recombinations of older profile elements within them.

Understanding process of representation change is significant not only at the individual level, but in thinking about larger cultural processes. When it comes to shared representations, we can begin to start defining the conditions under which we expect the prevailing representations of a group to begin to fail for individuals. A more in-depth treatment of these possible situations is available elsewhere (Shaw, in press), but in general, we can expect such disconfirmations to be more frequent during times of massive social or environmental upheaval (i.e. during “unsettled times” (Swidler, 2001) and “events” (Sewell, 1992)) and in those times and places in which individuals are interacting in contexts that are being shaped by different shared representation than the ones they have acquired (e.g. while traveling, when in a position of “brokering” between different groups, or as a minority who has a poor “fit” with the dominantly imposed representations of the culture). It is under these such conditions we would expect individuals to indicate their movement away from established shared representations and the adoption of new ones through first an uptick in the amount of observed deviation away from established behavioral patterns which is then followed by a decline in behavioral variability as new profiles cohere.

Implications of the Structuring of Social Life through Mental Representation for Empirical Research on Culture

Delving into the associative structure of mental representations results in a new way of looking at the patterning of social life as a partial outcome of many individual processes of meaning-making. Forging a connection between these cognitive processes and collective behavior not only furthers the sociology of culture theoretically but empirically as well. The development of methods for measuring meaning has been a long-standing endeavor for social scientists (for an excellent review, see (Mohr, 1998)). The model under development here does not serve as a replacement of these methodological approaches but instead offers a deeper, cognition-centered account of why these methods are relevant to the subject of culture. The payoff of laying out these microfoundations of meaning and behavior will be in opening up new venues for the empirical study of culture, including but not limited to arenas dealing with “big data," and in expanding the potential application of these established methodologies to new types of problems.

One of the most critical implications that follows from the relationship between mental representation in social contexts and the emergence of shared profiles of behavior and statement is the assertion that getting at collective sense-making (i.e. culture) requires researchers to go beyond individual behaviors and statements in order to look at the emergence, change, and dissipation of how such variables "hang together" and "fall apart" in social data. Said differently, these insights directly entail that investigation into macro-cultural processes requires a focus on finding and tracking *structure* in social data. Furthermore, given the highly contingent and historically specific nature of the associational learning processes that are posited as driving much of cultural development, this perspective also enforces the necessity of a kind of reflexivity or agnosticism in researchers that enables them to identify emergent structures in social data that may not align with their preconceived notions of what behaviors and statements should or should not be related to one another.

Fortunately for cultural research, the development of analytical tools which are specifically built to accomplish this exact task have been the subject of an enormous amount of interest and investment in the past decade. The rapid advances being made in these methodologies has followed in part from the unprecedented increases in computational processing power that has occurred over the past half-century. It has also been driven by the veritable explosion in the amount of data which exists in the world now, a so-called "Big Data" revolution which includes but is certainly in no way limited to the social data. The following subsections will briefly touch on the connection of some of these methodologies to the theoretical approach developed herein. The assertion of the relevance of these methods to cultural research will not be new, but the aim is to develop a deeper understanding of why they are appropriate and a more ambitious perspective on how they might be applied. The final section will then provide a preliminary exploration of how this theoretical and methodological insights help guide and motivate a "data mining for culture" approach through the example of using Yelp review data to look for cultural profiles in consumption and speech.

Cluster Analysis

The primary aim of clustering analysis is to identify the existence of "groups" within a set of observations. By taking a dataset of high dimensionality and algorithmically exploring the similarity of the observations to one another, these methods are often able to hone in on underlying patterns in such a way that reduces the complexity of the data into a simpler set of divisions between groups. Within the context of culture, we could think of applying a cluster analysis to the results of a survey which captured a variety of dimensions, say, a survey which asked about political views, personal values, transportation choices, and demographic information. Via this type of analysis, it would hypothetically be possible to take the whole set of observations in the dataset and see if certain clusters were found amongst them that related back identifiable social groups. Examples of possible clusters in this hypothetical survey might be those of young, liberal urbanites who preferred public transportation and middle-aged, conservative suburban residents who valued the place of religion in daily life.

The connection between the model developed here and cluster analysis, as well as related unsupervised learning methodologies, is the ability of these methods to pull out with minimal *a priori* specification how different observable characteristics have come to "hang together" and thereby reconstruct groups who, presumably, possess more or less similar representations of the aspects of life being considered. In effect, such analyses could give researchers a somewhat direct look at the representation-driven enactment of behavioral profiles which are of primary interest here and potentially go further to help identify what, if any, behaviors might act as particularly strong indicators for certain representations. This type of analysis could also identify those individuals who did not adhere notably to any of the common profiles and thereby facilitate further investigation into why such deviations occurred. Finally, clustering techniques may also be applied to longitudinal data in order to investigate patterns of coherence, dissolution, and reforming of behavioral clusters through time (with an expectation that such patterning would be reflective of transitions to and from "unsettled" and "settled" times).

Correspondence Analysis and PCA

Whereas cluster analysis and related methods are often used to look at the relationship of individual observations to one another within the space of multiple characteristics, other methods focus more directly on the way that the characteristics themselves tend to "hang together" across observations. Correspondence Analysis and the related method of Principal Component Analysis (PCA) already have a precedence in the study of culture, with the most well-known usage probably being Bourdieu's uses of Multiple Correspondence Analysis of survey data (e.g. (Bourdieu, 1984; Bourdieu, 1988)). Much as was the case for cluster analysis, these methods are able to leverage the co-occurrence of characteristics across observations in a dataset in order to examine both how different characteristics tend to appear together and identify the most important axes of differentiation between individuals.

The fairly obvious relevance of these methods to this model is in their ability to directly get at the observable profiles of action and statement that shared representations are expected to generate. While these methods can also be used to situate individuals in relation to each other through assigning them scores and thus placements along the axes which are found, a more compelling use of this method is likely to be in discovering the societal structures on top of which profiles of action and statement are constructed (see (Mohr J. W., 1998) for relevant discussion).

In Bourdieu's work, for instance, he was able to show how schema-driven profiles of behavior (i.e. habitus) are more deeply connected to the structures of economic and cultural distribution in a population (Bourdieu, 1984). In the context of the model proposed here, these axes would be interpreted as likely representing a set of deep and pervasive societal forces which strongly structure the mental representation formation (i.e. associative learning processes) of the individuals within it. In effect, these axes could be thought of as pointing to the set of environmental factors that play the most prominent role in organizing individuals' experiences and thus their histories of associative learning, in the realms of their lives being considered. Unlike a standard Bourdieusian approach, however, this model is not limited to primarily focusing on the distributions of economic and cultural capital in this regard. Instead, it supports the idea that any type of segregating structure which systematically and differentially shapes the lived and learned experiences of individuals will result in the emergence of different sets of shared representations, and thus profiles of behavior, across those groups. More specifically, the model indicates that these differences in shared representation and behavioral profiles will be most obvious in those areas of life that are most closely related to the systematic differences those individuals have experienced (e.g. behaviors related to assertiveness and career-orientation might differ primarily between men and women while those related to ways of interacting with formal institutions and their agents might be more notably different across races). As such, the axes that are found in these analyses would be expected to shift relative to the parts of a socio-cultural system being analyzed in a way that reflects the most salient dimensions of division within it.

Topic Modeling

The prior two examples focused on analyses of survey type data type. Another arena in which this model might be applied is that of automated content analysis and more exactly, to the application of topic models to large corpora of text. Topic modeling is a methodology which uses a statistical model known as Latent Dirichlet Allocation (LDA) to evaluate patterns of word usage across a body of 'documents' in order to reconstruct the 'topics' it contains (see (Mohr & Bogdanov, 2013) for a more thorough explanation). Cultural sociology has recently evinced a new interest in this method for getting at questions of culture and meaning in large bodies of text (see recent special issue of *Poetics* on the method (Mohr & Bogdanov, 2013)), and the method holds great promise in general for sociologists interested in taking on the analysis of big data.

For this model, topic modeling approaches represent a potential way of getting at behavioral profiles as they appear through the co-occurrence of words and phrases. Especially in cases where the substantive content can be kept fairly constant, the ability of topic modeling to extract patterns of word usage across texts has the potential to discover not just what is being discussed but also, the *way* it is being talked about. The patterns in how issues become "framed" (see Dimaggio, Nag, & Blei, 2013 for related work) connects back strongly to the construction of meaning and constitution of behavioral profiles via mental representation imposition. Not only does topic modeling allow researchers to map out the statement profiles that are at play across potentially vast milieus of interaction, but an extension of the topic modeling method known as Dynamic Topic Modeling makes it possible to pursue an analysis of the *changes* in these profiles, and thus shared representations, across time. The model presented here would have much to say not only on the expected shape of the new profiles which emerged (i.e. recombined elements of existing elements of representation per conceptual blending processes) but also would speak to the conditions under which dramatic changes are likely to occur (i.e. in circumstances of persistent representation disconfirmation).

The Continued Need for Qualitative Research

The above sections focused on some examples of how quantitative methods that emphasize patterns of co-occurrence can be linked to this model of mental representation imposition and behavioral profiles. While this approach is potentially able to supply these analyses of *how* behaviors “hang together” a greater level of explanatory power by furnishing them with a mechanism for *why* such patterns should arise, it still leaves some notable gaps in interpretation. While this model can point to the origins of meaning-making processes in mental representation and even help delineate how observed behavioral profiles map back to those representations, it is unable to go so far as to actually offer a robust reconstruction of what it is like to be inside the heads of the individuals being considered. Such reconstructions of lived experience ultimately require a richer understanding of the contexts within which individuals are situated than can be provided by statistically based analyses alone.

Fortunately, the acquisition of such rich contextual information is still possible through more intensive, qualitative methods like interviews, historical research, and ethnographic work.

Given the central place in this model of both polysemy and the role of individuals' unique histories of experience and learning in their processes of sense-making, it would be highly inconsistent to not acknowledge the necessity of gathering a wide base of specific knowledge in order to reliably describe the experience of these patterns of social life from the perspective of those enacting them. As such, the ideal body of work that could proceed from this linking of mental representation to social patterns would be one that is be fundamentally mixed-method in nature. Identification of what statements and actions are hanging together or being decoupled at a given time might be able to provide a powerful roadmap to the symbolic interplays that are going on between individuals' and the larger social world within which they are situated. To go the next step into actually describing the landscapes which are being traversed, however, ultimately requires researchers to move deeper into the particular circumstances of those individuals in order to understand what the observed sets of actions and statements actually mean to the people enacting them.

Data Mining for Culture: the Preliminary Explorations of Yelp Reviews

Social media provides unsolicited insight into how users think and behave, how they interact with others, and overtime how these interactions coalesce into shared meanings. As stated by Latour (2007), "The precise forces that mold our subjectivities and the precise characters that furnish our imaginations are all open to inquiries by the social sciences. It is as if the inner workings of private worlds have been pried open because their inputs and outputs have become thoroughly traceable." This perspective, it may be argued, is invaluable to analyses of unintended, cognitive structuring of social life. Though it is unrealistic to draw a strict boundary between online and offline interaction, social media spaces do serve as unique, semi-bounded contexts in which researchers may watch collective interpretations of a phenomenon emerge and evolve.

Description of Data

To analyze processes of collective sense-making, we use the Yelp challenge dataset - a dataset offered publicly for research purposes that contains a large volume of data on users, businesses and reviews extracted from Yelp. The full dataset contains data on 1.6 million reviews pertaining to 61,000 businesses, generated by 366,000 users from 10 cities across 4 countries. Due to the complexity of the computational processes being considered for this preliminary set of analyses, however, we randomly sample 10,000 users from this dataset and created from these users a toy dataset containing all of their reviews and the businesses to which these reviews are linked. In total, 61,184 businesses are included in this toy dataset, with the average user generating approximately 4 reviews. These data include user metadata (including IDs of friends, the duration of their time on Yelp, the number of reviews they have posted), review content and metadata (including the number of stars given), and business metadata for the establishments reviewed (including the category of the business and a description of the business ambiance). These measures provide an unsolicited look at the potential cultural spheres within Yelp as a community, as well as potential insights into how the occupation of these spheres relates to concrete outcomes, including how the user composes their reviews from a linguistic standpoint.

Yelp permits businesses to adopt a selection of category indicators to convey information about a given establishment. These categories are structured hierarchically, and businesses may have one or more categories associated with their profile. Although Yelp allows users to review a variety of business types - from doctors' offices to hotels to auto mechanics - the vast majority of the reviews left by Yelp users within this data are restaurant reviews. In order to extract a usable amount of detail from these data yet retain a small enough number of categories to avoid excessive data sparsity, we somewhat narrow the full list of 1,000 potential business categories to more informative list of about 875. Most often, we choose to retain the more specific children category designations (i.e. "Absinthe Bar" and "Sports Bar" categories as opposed to the general "Bars" designation), though in some cases the more general categorization is retained (e.g. "Italian" versus "Sicilian" or "Abruzzese").

Results

Following Bourdieu's classic application of the method (Bourdieu, 1984), we began with a multiple correspondence analysis, a technique used to find underlying structure within categorical data points within a multidimensional space, to examine the ways in which individual business preferences "hang together" as indicated by patterning in which businesses individuals chooses to review. Similar to principal components analysis - a dimensionality reduction procedure which clusters points based on their quantitative distance from one another within a Euclidian plane - multiple correspondence creates a matrix composed of binary indicators based on categorical characteristics and performs on this matrix a basic correspondence analysis.

One obvious initial challenge of using multiple correspondence analysis - or similar forms of dimensionality reduction - to obtain information on the underlying structure of our data relates to the sparsity of the data. While breaking the parent category "restaurant" into a set of meaningful subcategories provides a wealth of valuable information to our data, the binary value matrix generated from these indicators is incredibly sparse; only 2% of the cells in the data-set are filled with values other than 0. As a result of this, computing distances/dissimilarities are uninformative for a large portion of the sample. To illustrate, if a user only rated one restaurant, s\he is then 100% dissimilar from people that never rated that sort of place and would belong to their own cluster. This process, repeated over the course of the data, would give us an unwieldy number of unusable, isolated clusters. To manage this, we have preliminarily parsed down our data to include only categories with less than 200 values of 1 in them. This reduced the sparsity of the data to approximately 13% - still sparse, but sufficiently full to apply the method. The results of a multiple correspondence analysis using these pared-down data are displayed in Figure 6.

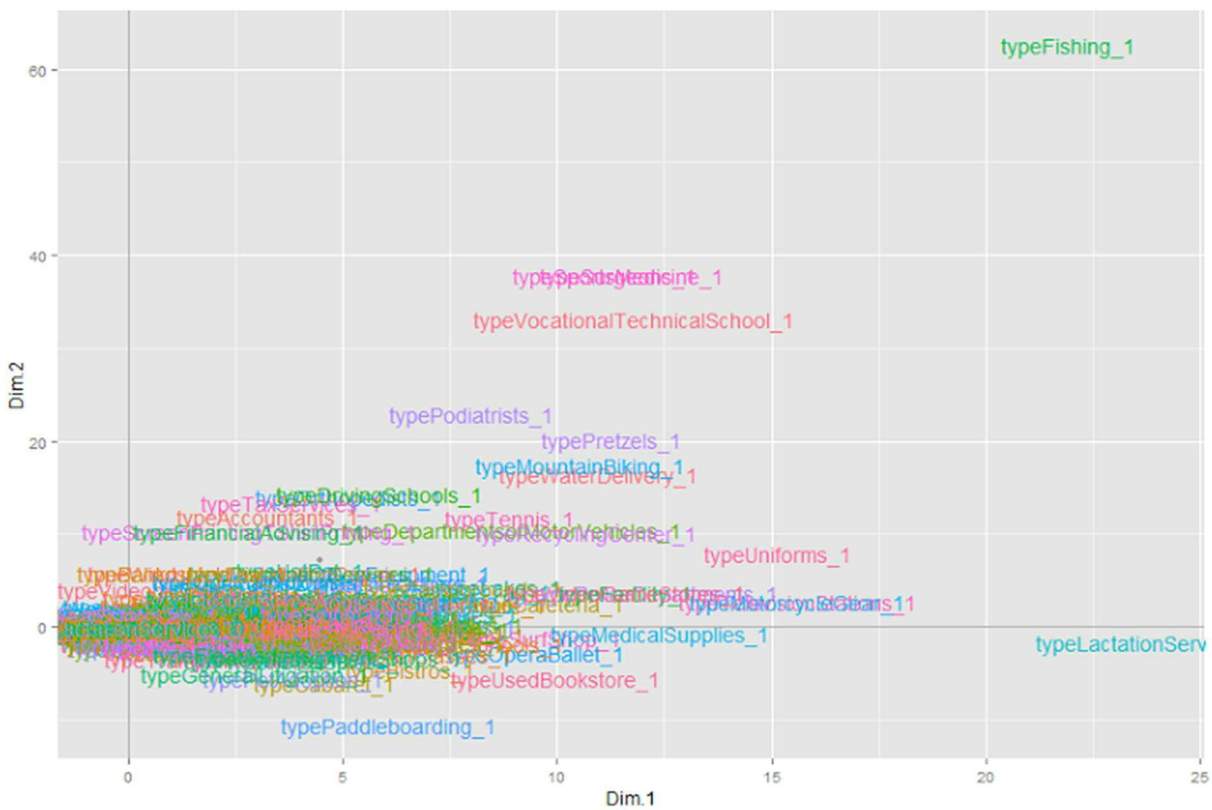

Figure 6 - *Multiple Correspondence Analysis of Yelp Users*

These results indicate that this method is not able to extract differentiation between users within our data based solely on their patterns of business reviews. The vast majority of raters are clustered into one identifiable category, surrounded by a handful of isolates. While this may be due in part to the fact that Yelp users self-select into the Yelp community, it is possible that the decision to post a review on Yelp may also be influenced by their offline connections, and that these same connections may influence their business preferences, resulting in a large body of culturally homogenous raters. It could also be that this dimensionality reduction methods does not lend itself well to data as sparse as that which is generated via Yelp. This setback could be potentially indicative of a significant limitation of using this method with digital data, which is often characterized by this degree of sparsity, for cultural analysis.

Given this limitation, our next step in this early exploration of the data was to use a graph-clustering technique, utilizing Jaccard's index as a distance measure, to cluster our data. To graph-cluster starts by constructing a distance matrix of size NxN with each element representing the dissimilarity between individual xi and ji. The next step in this process is to create an actual graph out of these data points using the the k-nearest neighbor technique. Essentially, this method selects the the kth closest point to point x, records its distance, and then connects it to other points within this range. To find clusters within these connected points, we chose to use a random walk technique or "walk-trap". This process involves sending a single point through the network, and recording where it spends the most time bouncing between points. Areas where the point gets caught for extended periods of time are clusters. Figure 7 displays the clusters as they emerge using this technique.

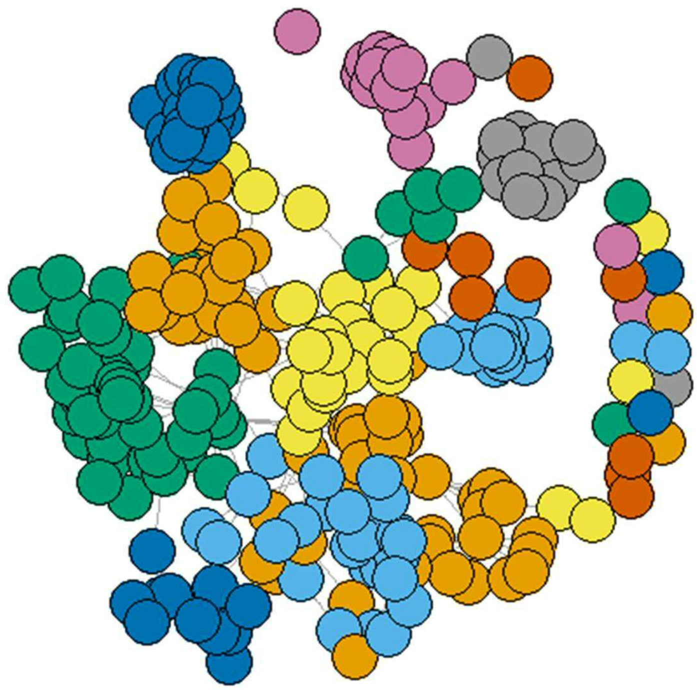

Figure 7 - *Structure of within-sample clusters using graph-clustering technique.*

Using this clustering technique, we see some meaningful structure emerge from within the data. Table 1 displays the traits that appeared in over 20% of the clusters displayed in Appendix A. Nearly all of the business categories displayed in these results are restaurants, indicating that the majority of variance within the data stems from preference in restaurants. There are some apparent patterns in the data - such as the tendency for those who review casinos and resorts or nail salons and day spas or media such as books and videos- to occupy the same cluster. From this first pass of investigation, it is also clear from the dominance of restaurant reviews that 1) there is a strong lopsidedness in Yelp usage toward the reviewing food and 2) that tastes in food across group seem to be generally eclectic, with American food being found across most such clusters.

We note, however, that there is nonetheless significant overlap within the traits of each cluster, and that a large portion of the data seems to fall within cluster 10. This could indicate that even reducing data sparsity and utilizing methods more appropriate for detecting patterns than cluster analyses such as MCA may not fully capture nuances contained within Yelp data. It could also be that the decision to become or not become a Yelp user is in itself a component of individuals' shared understanding of the world, and that this understanding is schematically tied to business preferences as well. Further unpacking of these early results, along with additional analyses using different methodological angles and potentially different subsets and organizations of the data, will likely be required to tease out these underlying drivers.

Discussion

Our preliminary results illustrate some degree of meaningful cultural patterns emerging within Yelp as a social space. We see, for instance, evidence that those who review event venues such as resorts and casinos emerge as a unique category of user. Likewise, those who utilize personal grooming services - such as nail salons and spas - emerge as a unique category as well. Nonetheless, the distinctions between individuals within the groups that emerged using graph-clustering techniques were not particularly severe. This could be due to the fact that our data are very sparse. It is clear that traditional dimensionality techniques that may effectively replicate the patterns of cultural distinction described by Bourdieu may not apply well to digital media. However, it appears that even methods more well-suited to work with sparse data reveal a significant amount of homogeneity in business preference. This could indicate that the decision to ‘Yelp’ or not ‘Yelp’ is a component of an individual's schematic understanding of the world that is also tied to his or her business preferences.

There are several next steps still ahead for this still early analysis of this rich dataset, with which we expect to both clarify our findings so far as well as hopefully discover and explore additional dimensions of potential structure in the data. As of yet, we have not incorporated review ratings of different business categories into our analysis, but we are interested in seeing if users of various sorts express a tendency to rate certain categories of business differently. Another line of investigation will also incorporate the ways in which users speak about their experiences with business, as indicated by the structure and content of their reviews. For this line of analysis, we will be focusing on metrics such as review length, review sentiment, the use of common within-cluster terms, and reading level of review, with the aim of identifying if these features track with other behaviors such as business category review patterns and networks of “friends” within the Yelp site. We also will be considering some additional methodological approaches, most namely associational rule learning, to see if they are better suited to extracting patterns in business category review behavior or if they also reaffirm our initial findings.

Regardless of the substantive content of our present findings, we argue that digital data presents new opportunities for cultural researchers to understand how collective mental representations emerge and evolve within social spaces via recognizing how these sense-making processes translate into concrete behavioral profiles. The ability to collect large volumes of unsolicited personal information in real time marks the introduction of a new era of social research. While the structure of these data - both in terms of how they are sampled and their content - presents new methodological challenges, researchers have the opportunity to view and understand collective sense-making in a way never before possible.

Works Cited

Allport, G. W. (1954). *The Nature of Prejudice.* Cambridge, Massachusetts: Addison-Wesley Publishing Company, Inc.

Bar, M. (2007). The Proactive Brain: Using Analogies and Associations to Generate Predictions. *TRENDS in Cognitive Science*.

Bargh, J. (1982). Attention and Automaticity in the Processing of Self-Relevant Information. *Journal of Personality and Social Psychology*, 425-436.

Bargh, J., & Morsella, E. (2008). The Unconcious Mind. *Perspectives on Psychological Science*, 73-79.

Bartlett, F. (1932). *Remembering.* Cambridge, UK: Cambridge University Press.

Bourdieu, P. (1984). *Distinction.* Cambridge: Harvard University Press.

Bourdieu, P. (1988). *Homo Academicus.* Stanford: Stanford University Press.

Boyd, D. & Ellison, N. (2007). Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication.*

Carlston, D. (2010). Models of Implicit and Explicit Mental Representation. In B. Gawronski, & B. K. Payne, *Handbook of Implicit Social Cognition* (pp. 38-61). New York: The Guilford Press.

Cerulo, K. (2002). *Culture in Mind.* New York: Routledge.

Cerulo, K. (2010). Mining the Intersections of Cognitive Sociology and Neuroscience. *Poetics*, 115-132.

Collins, A. M., & Loftus, E. F. (1975). A Spreading-Activation Theory of Semantic Processing. *Psychological Review*, 407-28.

Damasio, A. (1989). Time-locked Multiregional Retroactivation: A Systems-level Proposal for the Neural Substrates of Recall and Recognition. *Cognition*, 25-62.

D'Andrade, R. (1995). *The Development of Cognitive Anthropology.* New York: Cambridge University Press.

Dimaggio, P. (1997). Culture and Cognition. *Annual Review of Sociology*, 263-287.

Dimaggio, P., Nag, M., & Blei, D. (2013). Exploiting Affinities Between Topic Modeling and the Sociological Perspective on Culture: Application to Newspaper Coverage of U.S. Government Art Funds. *Poetics*, 570 - 606.

Fauconnier, G., & Turner, M. (2002). *The Way We Think.* New York: Perseus Books Group.

Feldman, J. (2006). *From Molecule to Metaphor: A Neural Theory of Language.* Cambridge, Mass: MIT Press.

Hebb, D. O. (1949). *The Organization of Behavior.* New York: Wiley & Sons.

Higgins, E. (1996). Knowledge activation: Accessibility, applicability and salience. In E. Higgins, & A. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 133-168). New York: Guilford Press.

Holland, J., Holyoak, K., Nisbett, R., & Thagard, P. (1986). *Induction: Processes of Inference, Learning, and Discovery.* Cambridge: MIT Press.

Kahneman, D. (2011). *Thinking, Fast and Slow.* New York: Farrar, Straus and Giroux.

Latour, B. (2007). "Click Era Spawns a Data Rich World." Times Higher Education. Retrieved from:http://www.timeshighereducation.co.uk/208526.article

Lakoff, G. (1996). *Moral Politics: What Conservatives Know That Liberals Don't.* Chicago: University of Chicago Press.

Lieberman, M. D. (2007). Social Cognitive Neuroscience: A Review of Core Processes. *Annual Review of Psychology*, 259-89.

Lisman, J., & Sternberg, E. (2013). Habit and Nonhabit Systems for Unconcious and Concious Behavior: Implications for Multitasking. *Journal of Cognitive Neuroscience*, 273-283.

Martin, J. L. (2010). Life's a Beach but You're an Ant, and Other Unwelcome News for the Sociology of Culture. *Poetics*, 229-244.

Mohr, J. W. (1998). Measuring Meaning Structures. *Annual Review of Sociology*, 345-70.

Mohr, J. W., & Bogdanov, P. (2013). Introduction - Topic Models: What They Are and Why They Matter. *Poetics*, 545 - 569.

Mohr, J., & Bogdanov, P. (Eds.). (2013). Topic Models and the Cultural Sciences. *Poetics, 41(6)*.

Morewedge, C., & Kahneman, D. (2010). Associative Processes in Intuitive Judgement. *Trends in Cognitive Science*, 435-440.

Murphy, G. L., & Medin, D. L. (1985). The Role of Theories in Conceptual Coherence. *Psychological Review*, 289-316.

Payne, B. K., & Cameron, C. D. (2013). Implicit Social Cognition and Mental Representation. In D. Carlston, *Oxford Handbook of Social Cognition.* Oxford University Press.

Sewell, W. (1992). A theory of structure: duality, agency, and transformation. *American Journal of Sociology*, 1-29.

Shaw, L. (in press). Mechanics and Dynamics of Social Construction: Modeling the Emergence of Culture from Individual Mental Representation. *Poetics*.

Swidler, A. (2001). *Talk of Love: How Culture Matters.* Chicago: University of Chicago Press.

Taylor, S. E., & Crocker, J. (1981). Schematic Bases for Social Information Processing. In E. T. Higgins, C. P. Herman, & M. P. Zanna (Ed.), *Social Cognition: The Ontario Symposium. 1*, pp. 89-134. Hillsdale, NJ: Erlbaum.

Wyer, R. (2007). Principles of Mental Representation. In A. W. Kruglanski, & E. T. Higgins (Eds.), *Social Psychology Handbook of Basic Principles* (pp. 285-307). New York: The Guildford Press.

Appendix A: Cluster Descriptives

Table 7: Cluster Characteristics for Graph-Clustering Results

Cluster membership	Business Type	Frequency
1	American (new)	14
	Lounges	13
	American (traditional)	11
2	American (new)	44
	Breakfast/Brunch	60
	Burgers	38
	American (traditional)	51
3	American (new)	27
	Japanese	40
	Sushi bars	38
4	American (new)	21
	Pizza	31
	Sandwiches	21
	Italian	28
	American (traditional)	20
5	American (new)	115
	Breakfast/Brunch	78
	Italian	71
	American (traditional)	106
7	American (new)	10

	Day spas	14
	NailSalons	11
8	American (new)	35
	Breakfast/Brunch	33
	Steakhouses	52
	Italian	34
	American (traditional)	41
9	American (new)	70
	Japanese	45
	Breakfast/Brunch	56
	Chinese	47
	Pizza	44
	Mexican	46
	Sushi bars	40
	Burgers	43
	Sandwiches	40
	Italian	43
	American (traditional)	68
10	American (new)	49
	Japanese	23
	Breakfast/Brunch	40
	Seafood	23
	Pizza	43
	Bakeries	27
	Mexican	44

	Sushi bars	26
	Steakhouses	26
	Burgers	40
	Lounges	25
	Sandwiches	38
	Italian	43
	American (traditional)	44
	Thai	22
	Asian fusion	22
	FastFood	22
	Cafes	28
	Coffee/Tea	27
11	Burgers	5
	Books	5
	Mags	5
	MusicVideo	5
12	Buffets	4
	American (new)	4
	Seafood	4
	Chinese	4
	Pizza	5
	Burgers	5
	Sandwiches	5
	American (traditional)	6

	Beer	5
	Wine/Spirits	5
13	Casinos	27
	Resorts	20
14	American (new)	5
1	Pizza	6
	Burgers	5
	AutoRepair	5
15	American (new)	5
	Pizza	6
	Home services	6