Full Title: Quantifying the Information Gained from Extra-network Stimuli: An Experiment on the Emergence of Social Conventions

Short Title: The Informational Value of Extra-network Signals.

Disciplinary Classifications: Network Science, Sociology, Information Theory, Psychology

Author: Jon Atwell
atwell@northwestern.edu

Affiliations:

1. Northwestern Institute on Complex Systems, Northwestern University, Evanston IL, 60208
2. Institute for Policy Research, Northwestern University, Evanston IL, 60208

ORCID: 0000-0002-9706-2640

Preprint Location: osf.io/preprints/socarxiv/zka28/

Keywords: social cognition | network science | information theory | behavioral experiment | social conventions

Critical Scientific Statement

Humans' information processing capacities are vast, yet quantitative research on social processes often treats information from social network ties as being the exclusive relevant source. In many social settings, information about the behaviors of socially distant actors can be found in our physical and digital surroundings, and here I present experimental results showing that the availability of such information can alter the outcome of coordination games with arbitrary solutions, a class of games commonly used to model the self-organization of social forms such as conventions, norms, and beliefs. The modeled signal source is comparatively information rich, often carrying more relevant information than additional weak ties. This suggests research on social dynamics should consider such sources as potentially causally relevant; even small exposures to them can tip dynamic processes to different outcomes. This result is especially pertinent for online communities, where users collectively produce culture and extra-network signals are often algorithmically-limited, but need not be.

Abstract

Which stimuli found in social settings are incorporated into individuals' socially-oriented cognitive processes? Research on social self-organization focuses on explicit and direct communication via network ties, yet it is plausible that *extra-network* stimuli carry valuable information. Here, I assess the informational value of such stimuli for the self-organization of conventions, a common large-group coordination problem. Participant exposure to a small number of additional stimuli can tip groups to coordinated outcomes through a rapid decrease in the diversity of the potential solutions. The information gained through these exposures is greater on average than that of a variety of alternatives, including additional weak ties. These results suggest actors ought to be attentive to extra-network stimuli and that research assessing the role of social network topology in determining the outcome of social influence and information-driven dynamic processes might often have misspecified causal mechanisms.

Introduction

Of the profusion of stimuli humans process, what types of stimuli play a role in the production and transmission of durable social patterns and cultural forms such as conventions, norms, styles and cultural beliefs? Current research treats explicit communication from known interlocutors—often consciously processed—as the exclusive relevant type (1-7). This diffusion-over-networks approach fits well the common-sense psychology of how we attribute mental states to others (8). Research in social cognition, however, offers a much richer picture of social cognition and information processing (9-14); a wealth of cues, many of them non-linguistic, make their way into our cognitive processes and affect behavior. The mode of cognitive processing of such stimuli may alternate between fast and automatic, and slow and deliberative (15). Are stimuli of this type redundant or otherwise irrelevant inputs for social dynamics and self-organization, or might they tipping processes toward particular outcomes?

The answer to this question has important implications for how we conceptualize shared mental constructs such as conventions, norms, and, even more broadly, cultural schema. If the successful production and transmission of these social constructs depend on attentiveness—conscious or unconscious—to stimuli originating from a broader range of actors than just acquaintances, here referred to as *extra-network stimuli*, the sharedness of the constructs is necessarily more expansive and diffuse. Conversely, the absence of such stimuli due to spatial segregation or the design of digital platforms could undermine the information processing necessary for the creation of shared constructs.

Assessing the importance of extra-network stimuli is challenging because the stimuli are hard to track, particularly when they are processed implicitly by actors, which prevents reliable self-reporting (16). Research programs such as implicit association testing (17) and eye-tracking (18) offer fruitful pathways forward for studying implicit cognition, but rather than similarly attempt to identify information consumption or representations at the individual level, this article instead argues that different types of stimuli are coupled to different information *source types* and then

experimentally tests whether conjunctive exposure to different source types alters the outcome of a large-group coordination game. The affirmative results reported below suggest a need for more empirical research on the functional significance of such sources and signals.

Information and Source Types

A distinction is commonly made between local and global sources of information (19-21). The former includes anything learned directly from people or objects around the receiver and the latter is a one-to-many transmission, often with information about the world outside the receiver's milieu. In an age of social media and algorithmically-structured interactions, this distinction establishes two extremes of a continuum. Information from both source types, and anywhere in between, influences individuals' beliefs, opinions, tastes, and other cultural schema. Orthogonal to this common distinction, however, is a source's representativeness of the underlying features of the domain it covers. Global sources can be more or less biased in the mundane sense of conveying a well-balanced sample of goings-on. Can the same be said for local sources? The fundamental importance of social networks is that they circumscribe interaction and communication to a subset of the relevant population, suggesting local sources are inherently biased. If, however, social cognition makes use of *extra*-network stimuli—local stimuli other than those obtained via direct, conscious communication—local sources need not be biased. The formalisms of information theory help to explain why this might be the case.

Information theory treats communication as the realization of a stochastic process. It is generally assumed that signals come from an *ergodic* source, meaning the time averages for the random variables of the system are the same for each instantiation of the process, a more processual law of large numbers (22). This quality implies that given enough observations of a signal source, the information contained therein will reliably reflect the underlying random variables. Much of human interaction is decidedly *non-ergodic*, however. Social networks make it likely that the messages conveyed are not representative of the population at large and, furthermore, the underlying state of the system is often changing (i.e. the source is *non-stationary*). The lack of source ergodicity alone prevents a series of observations from ever revealing the true state of the system and a group of isolated individuals with distinct, non-ergodic sources of information about the system will necessarily have differing knowledge of it. This problem is exacerbated when the system itself is non-stationary (i.e. evolving).

A core insight of network science has been that information aggregation on an underlying network can address the problem of source non-ergodicity; some network connectivity patterns facilitate the diffusion information about the state of the stationary system to all individuals (7). However, when the behavior of individuals depends on their knowledge of the behavior of others, as is the case with any social process of self-organization, the system is non-stationary or evolutionary. Successful information aggregation in evolutionary settings again becomes precarious; diffusion of information is often inhibited by the tendency of tightly connected communities in the network to settle on different behaviors and produce more information that reinforces the non-representativeness of local knowledge of the system (23). That types of sources of information might successfully combat this tendency?

Information aggregation processes generally benefit from information from a more representative sources. A truly ergodic source is not necessary (ergodicity assumes stationarity, which is often lacking anyway), but something more representative might be found in the abundance of indirect, extra-network social cues in our physical and digital lives. A variety of social science perspectives see materiality and our local environment as fundamental parts of our social umwelt. Objects afford actions and meanings; we can read social position off physical, linguistic, and sartorial clues; we struggle to not hear others' conversations; and above all we orient our public presentations toward a sense of how we are perceived by an abstract, generalized other (24-34). The informational value of these elements is assumed to be functionally insignificant in contemporary casual-mechanisms focused research. This is a parsimonious and justified choice if the source of these signals is the same, or simply has the same properties, as the network source. It is likely not the same, however, because social space is layered into physical (and digital) space (i.e. is non-injective), creating many opportunities for exposures to signals from outside one's ego network but still within a culturally-relevant group.

As a first step toward understanding the potential informational value of these signals, this article abstracts away from the actual myriad forms of the signals to first focus on the effect of exposures to a signal source that does not have the biases of network sources. More specifically, it tests whether adding signals from a completely unbiased source can alter outcomes in a large-group pure coordination game. Contra the *structured* source that is the flow of signals over a social network, this idealized source is *unstructured* insofar as there is no pointwise correlational structure between the set of all possible individual signals. The structure of the network-based source ensures there are such dependencies between signals. Very few, if any, social processes are unstructured at any scale, but a purely unstructured source serves as an important baseline; if the addition of a fully-unstructured source has no effect, it is almost certain that other, more biased sources would either. Such a source is equivalent to homogeneous mixing without actual interaction.

While the results presented below make clear the value of unstructured sources, it is not clear *a priori* that combining signals from structured and unstructured sources would facilitate different outcomes. Individuals derive benefit from successful interactions and therefore have no (direct) incentive to heed a signal from elsewhere, especially without knowing whether it was linked to successful behavior in its interactional context. But assuming individuals do nonetheless heed these signals, it might not change behaviors. Or the changes could be detrimental to the process of self-organization, encouraging too much exploration and not enough exploitation. After describing the design of the experiment and reporting the basic results, I explore in several ways why the addition of the unstructured source has the effect it does.

Experimental Design

To assess the importance of unstructured signals for self-organization, I make use of a recent experimental design of Centola and Baronchelli (35) (henceforth CB) exploring the relation between network topology and the emergence of social conventions in groups. The game design

was derived from theoretical (36-39) and formal models (40, 41) of the emergence of conventions and has since been extended (42). Broadly defined, conventions are arbitrary but agreed-upon social practices that fulfill a function in social interactions. Alternatives could work just as well or even slightly better, but what is most important is that individuals have shared expectations about meanings and behavior. For example, titles like Mr., Mrs., and Professor can be used to show respect, deference, or to establish relative social rank in many societies. Alternatives include avoiding prolonged eye-contact or, in some languages, using formal pronouns. The meanings conveyed through these conventions have lasting effects on interactions. Language and culture are deeply conventional (38), but so are more procedural practices like which side of the road to drive on and or when to pay for a meal at a restaurant. Defined formally, a convention is a solution to a pure coordination game (36).

The results of the original experiment show that group-wide conventions emerged when the population mixes at random (i.e. homogenous mixing), but not for the random and lattice graphs tested. Groups embedded in these networks exhibit well-entrenched alternatives at the end of game play. This result makes clear that homogeneous mixing alone circulates sufficient relevant information for what is ultimately a complex contagion with many possible states (43) to reach the whole population. It is, however, only a starting point, and the question remains of what, if any, additional sources might allow networked groups to successfully self-organize conventions.

The game shows participants a headshot of an individual and incentivizes them to coordinate on a name for the individual. Participants are partnered with one of their network neighbors at random each round. At the start of each of round of the game, participants are given a timed period to individually submit a name and upon submission are shown the name their partner submitted. Partners who match individually receive a small monetary reward for the round. Partners who do not match are penalized, as long as their current cumulative rewards are sufficient. Participants are then assigned new partners for the next round. The three interaction topologies treatments are a random graph of constant degree four, a circular lattice of degree four, and a fully-connected graph, which is equivalent to homogeneous mixing for pairwise interactions.

The partners do not know each other's identities before or after the interaction, as this would introduce second-order social processes and skills that would obfuscate the dynamics of the core task and undermine the value of the experiment; for the low degree networks, participants would likely quickly solve the unique two-person coordination games by memorizing identities and the corresponding names, leading to N^2 pairwise solutions. The social value of conventions, however, is in the fact that they facilitate successful interactions across large groups of both known alters and strangers alike by removing the cognitive load of having to memorizing highly specific settings for particular behaviors. Withholding identities induces the more generalized group coordination setting.

To explore the value of unstructured signals within this design, we treat the whole group as if it occupies the same social (or even physical) space in spite of the social distances inherent in the

social networks in which participants are embedded. Situating socially distant actors in a common social space ensures that they are exposed to some amount of unstructured information about the behavior of others within the group at large. These exposures take the form of each participant being shown additional names, each played by any other participant that round. The number of names--zero, one or two--constitutes the second treatment variable. The name or names are shown on the round's result page along with the name played by their partner. No information regarding who played the name or whether it matched that of the respective partner is given. Names are drawn at random with replacement and can include null values if the corresponding player failed to submit a valid name. (See Software and Subject Management Details for details.) Crucially, there is no direct benefit for matching any or all of these additional names and, furthermore, without the knowledge of the corresponding partner's behavior, the value of that name for coordination's sake is not known. It is merely a trace of socially-relevant activity that observers can process as they see fit.

Based on the findings in CB that the size of the group did not have an effect on the outcome, all trials in this design were conducted using 24 participants. In addition to the three network treatment factors in the original, I include a small-world network seeded with circular lattice of degree four. Small-world networks share the important characteristics of local clustering and short characteristic path lengths with real world social networks (44), but, as an interpolation between lattice and random networks, would not have added anything to the analysis in the original experiment given that the random topology did not self-organize a convention. Because of the potential for the unstructured signal treatment in the present experiment to change that outcome, small-world network topologies are included.

The unstructured signal exposure factor has three levels. Participants may be exposed to zero, one, or two additional names. The inclusion of the no-name condition replicates the CB experiment. Trials for this treatment were done before the other levels to confirm the commensurability of the game interface in spite of minor design differences (see Model Details). Because the fully-connected network produces global conventions without the addition of unstructured signals, I omit trials for the one and two additional names treatment levels. Most other treatment conditions have four trials, but replicated ones have two and the lattice network with a single additional name also has two because of the low likelihood of success given the characteristics of the topology.

Results

Table 2 shows outcomes for the whole experimental design. In addition to the results in CB fully replicating, it is clear that unstructured signals in the form of exposure to additional names affect the outcome. The effect is conditional, however, on features of the network topology treatment—likely the average longest path length. It is not obvious that exposure to unstructured signals could break the symmetries blocking the coordinated outcome because i) participants might well ignore information not immediately relevant to game play and financial incentives; ii) the wide variety of names seen through these exposures may be too hard for individuals to process into actionable information in a small amount of time; and iii) the additional variation in names could

impede the emergence of locally-consistent behavior even if individuals properly incorporate it into their expectations. All that is clear *a priori* is that for many cases, each additional name yields additional information. The informational value of the unstructured signals and some alternative mechanisms are assessed below, but first we describe the temporal dynamics of the trials.

Fig. 1 details the dynamics of select runs by plotting the number and concentration—as defined by the normalized Herfindahl-Hirschmann index—of names in circulation, and the percentage of pairwise interactions that are successful. The first and second columns strongly suggest that driving up the concentration of name usage, particularly in the early rounds, is the mechanism by which exposure to additional names aids in successful self-organization. This pattern also holds true for fully connected networks with no additional names and offers further evidence that seeing a more diverse set of names from across the population counterintuitively increases the concentration of name usage. Exposure to this wider range of names can lead to a lower local success rate, but, counter to the explanation given in CB for the success of trials with fully connected networks, early coordination failure does not itself lead to later success; random-graph treatments with no additional names also suffer from early local failure without long-term success, but, more importantly, trials with additional names can experience local success early on and still converge to a global convention *as long as the concentration of names is also rapidly increasing*. This change in concentration is of ultimate importance because the total number of names in circulation appears to depend primarily on the interaction topology and can decrease substantially without much change in the concentration, as a few entrenched alternatives with equal frequencies corresponds to a low concentration score.

The presence of both outcomes within the same treatment cell, however, means that factors other than the number of additional names and network topology affect the outcome. Close inspection of individual participant's name-usage patterns reveals that some deploy high commitment strategies that are slow to respond to the changing distribution and possibly disrupt the progression toward order. Whether this is the case is hard to assess because the path toward high or low concentration is established well within the first ten rounds, a period when it is hard to adjudicate whether the high commitment strategies are in fact hindering any nascent drive toward high concentration and an eventual group-wide convention.

Taken together, these observations strongly suggest that signals of an unstructured variety can affect at least some classes of processes of social self-organization by increasing the rate at which possible states are eliminated from the system. By reflecting the diversity within the population, unstructured signals carry important information about the state of the broader system. In this experimental design, however, the signals are also simply additional information. In order to understand if these signals are in fact qualitatively different, the next section considers the value of each additional name and compares them to other potential source types.

Exploring the Information Content of the Unstructured Source

One of the key insights of information theory is that the amount of information gained by receiving a message or signal depends on what the receiver already knows. That means it is possible to receive an information-less message, one that does not alter the receiver's knowledge. While a useful baseline, most signals do carry information, and the additional names in this design most often do. Thus, in order to know whether the *source type* of the signal rather than simply the number of signals has functional significance, those signals need to be compared to possible alternative sources of additional signals. If signals from unstructured sources tend to yield the same or less information than other sources, then any effects are more practically driven by the number of signals more than the type.

To assess this relative informational value of the unstructured signals within the name game experiment, I compare them to several alternative sources: additional names from network neighbors, additional names from new ties selected at random, additional names from new *weak*—maximally structural distant—ties, and the additional names from the maximally informative alters, defined post hoc. These comparisons are made at each round of the trial using the Kullback-Leibler divergence (45) from the distribution of names seen by individuals to the true distribution of names in the population. The experimental design collected data only for the case of the unstructured signals, so the other comparisons are made to synthetic distributions comprised of each participant's distribution of names from normal interactions and additional names from the relevant source. Details of the divergence measure, including the required creation of a common support and the specification of the new sources, appear in the Simulation section.

The comparison cases are all from structured sources with a range of potential informational values. The first case—randomly selected network neighbors with whom the participant is not currently paired—is not likely to provide much additional information, as it is drawn from the same subset of the population the participant already interacts with. The names played by new, randomly-selected neighbors are likely to provide more information than the first comparison, but because individuals' name-use should quickly become more predictable, these signals are likely to be informative primarily in the early rounds. Following the literature on weak ties and structural holes (46, 47), the names played by the structurally most distant participants—modeled here as those with the longest shortest path to the ego—should have relatively high informational value. The final comparison is to the names played by the participants whose histories of names are most informative of the true distribution of names in the population.

It is important to note that this approach cannot assess the effect of the actual exposure to the comparison signal types. The round-by-round true distribution is changing, as it did in the actual trial, based on participants' exposure to the unstructured source and not in response to the synthetic exposures. This means the differences in divergences presented below are pointwise and not cumulative, a shortcoming that will likely lead to noisy estimates of the differences. Instead of a pointwise comparison, one should rightly imagine the total informational divergence between the comparison cases as the sum of the divergence across all rounds, but that

comparison is not possible with the current data. Nonetheless, the pointwise comparisons are a useful approximation of the information gains from various types of exposures.

Fig 2 shows the population average and standard variation of participants' differences in Kullback-Leibler divergences between the unstructured additional name treatment that was actually used in the experiment and the synthetic comparison source. Where the difference is positive, the distribution of names seen by the unstructured signals yields more information than the comparison source. To contextualize that difference, a comparison of the divergences from the true distribution to participants' distributions both with and without the unstructured signal to the true population is plotted on the right axes. The latter establishes the total amount of information participants without the unstructured signal would need in order to know the true distribution. One minus the latter divided by the former thus assesses what percentage of total uncertainty that was made up by the addition of the unstructured signals. The value of the participant with the highest ratio is recorded for each round of the game. This ratio tends toward one as participants accumulate a history of names, the group self-organizes, and the entropy decreases. (See the Methods section for the definitions.)

The population mean of the difference in divergences is greater than zero for nearly all games and rounds for the comparisons to the random neighbors, random others, and weak ties. The information gain of the unstructured signals is only less than that of the ties to the maximally informative participants, assessed post hoc. In the first round, differences are minimal in virtue of the low concentration among the alternatives leading to almost any additional name being equally informative. From there the comparative informativeness of unstructured signals builds, except when compared to the most informative. In trials in which a convention does emerge, this advantage appears to decline over time, but that is a direct result of the emerging order and the increasing concentration of names reducing the total amount of information necessary to describe the system. In cases with no emergent convention, the advantage is maintained, although it is not enough to overcome the now entrenched alternatives.

These comparisons do not make clear that the relationship between the number of additional names participants are exposed to and total information gained is not likely to be linear; the first several additional names are likely to carry the most information and subsequent ones less. To highlight this feature of unstructured signals, Fig 3 plots the average information gain of repeated, simulated exposures to a given number of names drawn from the empirical distribution. The second derivative of the number of additional names is negative for all surfaces, corresponding to the fact that, on average, each additional name is less informative than the last. Consistent with the comparisons in Fig 2, the value of the unstructured signals is diminished as the system becomes more ordered. (See Simulations for details.)

These simulations clarify the informational content of the unstructured source, but not the importance of it to the behaviors of the participants. A full model of the participants' behavior is outside the scope of this article, but simple tallies of the times the participants played and matched on a name i) they first saw through the unstructured source (*before*), and ii) they saw

only through the unstructured source (*only*) strongly suggests their behaviors actively incorporated the information. Table 2 presents the grand average of the trials in the cells of the fraction of matches fitting the *before* and *only* definitions. The *before* fractions strongly suggest participants are responding to seeing names first from the unstructured source and then from the structured source. The *only* fractions, however, provide clear evidence that participants are responsive to the unstructured source. The dynamic consequences of this responsiveness can be seen in Table 1, but there is no clear effect on the outcome in Table 2.

Discussion

This study shows that the addition of unstructured, extra-network sources of information can alter the outcomes of the self-organization of conventions. Stimuli from this source type convey more information most other structured sources, including weak ties. Of the alternative sources tested, the only one to outperform the unstructured source contained the maximum amount of information, defined post-hoc. While this does not ensure that actors are attuned to such sources *in vivo*, it is clear that humans have the cognitive capacity to process a great deal of stimuli, and the results reported here show that the incorporation of data from this source can decisively alter the outcomes in processes of self-organization involving collective computation (48, 49). This potentiality is likely to be present in any setting where the dynamics bring the system near critical or tipping points, where small differences in local behavior cascade to substantially alter future system-wide dynamics (50). In the context of this experiment, that period of criticality is early in the game, where the concentration of the distribution of names is set on a path toward either high- or low-concentration within the first ten rounds of the game (see Fig. 2). This means the symmetries that emerge in the control group are broken in a way that is unlikely to be noticeable to the participants. There is no period of entrenched, competing alternatives with one eventually winning out. Rather, competition is forestalled while the eventual winning name propagates through the network.

This study is limited by the fact that it focused on a single game where access to an unstructured source would plausibly make a difference. Assessing the consequences of the availability of unstructured sources of information in more diverse settings, particularly those not starting from a state of complete disorder, will be a harder and more valuable research endeavor. Nonetheless, the evidence of the importance of unstructured signals fits with a range of thinking in the social sciences. Social theorists have long believed that social actors consciously observe the broader social environment for relevant cues (25, 51-54) and have become interested in the role of fast, implicit modes of cognition (55-57). The implications of such behavior are potentially vast given the rise of networked and algorithmically-driven societies. Codified social networks and the opaque algorithms of social media tend to limit exposure to less structured sources, potentially undercutting the emergence of broadly accepted cultural forms and meanings. Recent evidence suggests cross-cutting exposures can actually increase political polarization (58), but this effect is measured well after the entrenchment of alternatives, exactly the outcome the results here suggest can be prevented if unstructured signals are present all along.

Methods and Materials

Data. The analysis code and the data from all completed trials are available at [TO COME]. Some trials lost participants mid-game and were excluded from further analysis. Visualizations of the trials are available [TO COME]. The web interface was developed using the oTree python package [59].

Divergence Difference Measure. Divergence differences reported on the left axes in Fig 2 are defined as

$$DD_{i,t}(\Phi, \Theta, \Lambda) = D_{KL}(\Phi || \Theta) - D_{KL}(\Phi || \Lambda)$$

where $D_{KL}(P || Q)$ is the Kullback-Leibler divergence from Q to P, Φ is the true population distribution, and Θ and Λ are the comparison distributions. For left axes in Fig 2, Θ is participant i 's observed distribution *with the additional synthetic names* and Λ is the observed distribution with the true additional unstructured names. Because the KL divergence increases the less similar two distributions are, $DD_{i,t}$ is positive when the additional names yield information. We use the Kullback-Leibler divergence instead of the symmetric Jensen-Shannon [44] measure because there is a clear true distribution. However, because the true distribution and the distribution of names seen by each participant do not have a common support, one is first created by adjusting the participants' distributions to have non-zero weights (.001) on each missing category and then normalizing the other weights. Additionally, the distributions span several rounds of game play, because participants remember prior rounds, but names also disappear from the true distribution over time. Both facts suggest a need for finite memory, which is implemented by including the distribution of names for previous k rounds for each distribution. The first k rounds include all previous rounds. Because results are approximately equally for $k=4, 8, 12$, and 16 , only values for $k=8$ are reported.

Divergence Ratio Measure. The divergence ratio reported on the right axes in Fig 2 is defined as

$$DR_{i,g} = \text{Max}(\{R_{i,t,g}\}) \text{for } R_{i,t,g} = 1 - \frac{D_{KL}(\Phi_{i,t,g} || \Theta_{i,t,g})}{D_{KL}(\Phi_{i,t,g} || \Lambda_{i,t,g})}$$

where i indexes the participant, t indexes the round, g indexes the trial (game), $\Phi_{i,t,g}$ is the distribution of names actually seen, and $\Lambda_{i,t,g}$ is the distribution of names excluding those from the unstructured source. The denominator thus accounts for the total information needed to know the true distribution and ratio represents how much of that information was transmitted by the names from the unstructured source.

Simulations. To clarify the informational value of the ℓ additional names from the unstructured source, we compare that source to other types. This requires constructing synthetic distributions by starting with the base distribution of the names (i.e. excluding those from the unstructured source.) for each participant and adding to that the names the participant would have seen if she had been exposed to the alternative source. The three sources are constructed in the following way: *Random neighbors* randomly selects ℓ names each round from among those played by the participant's neighbors with whom the participant is not currently paired with. *Random other* selects ℓ non-neighbors at the start of the game and reports their names played each round. The *weak ties* source identifies at the start of the game the ℓ participants for which the participant has the longest shortest path to and reports their names played each round. When there is a tie among path lengths, a new alter is randomly chosen. Because the addition of a weak tie changes the shortest paths, when there is more than one additional name the path lengths are recalculated with each additional tie. Finally, the *maximum information* source first tracks the total information gained by exposure to the names of all tuples of length ℓ of other participants. The tuple with the largest information gain is considered the maximum information source.

A second simulation explores the value of each additional name from the unstructured source. Because the unstructured source randomly selects names, there is variation in the information gain accorded to each sample. To account for this, for each additional number of names and each round, a synthetic distribution is constructed 200 times. The average for each individual is recorded. The grand average across participants is recorded for each game. For each cell of the table of graphs, the average of the games within it are reported.

References

1. Mason W, Conrey F, Smith E (2007) Situating social influence processes: Dynamic, multidirectional flows of influence with social networks. *Pers Soc Psychol Rev* 11(3):279-300.
2. Borgatti SP, Mehra A, Brass D, Labianca G (2009) Network Analysis in the Social Sciences. *Science* 323:892-895.
3. Castellano C, Fortunato S, Loreto V (2009) Statistical physics of social dynamics. *Rev Mod Phys* 81(2):591–646.
4. Chistakis N, Fowler J (2009) *Connected: The Surprising Power of Social Networks and How They Shape Our Lives*. (Little, Brown and Company, New York).
5. Galeotti A, Goyal S, Jackson M, Vega-Redondo R, Yariv L (2010) Network Games. *Rev Econ Stud* 77:218-244.
6. Bakshy E, Rosenn I, Marlow C, Adamic L (2012) The role of social networks in information diffusion. *Proceedings of the 21st international conference on World Wide Web* (ACM, New York, NY), pp 519-528.
7. Centola D (2018) *How Behavior Spreads: The Science of Complex Contagions*. (Princeton University Press, Princeton).
8. Heider F (1958) *The psychology of interpersonal relations*. (Wiley, New York).
9. Call J, Carpenter M (2002) Three Sources of Information in Social Learning. *Imitation in Animals and Artifacts*, eds Dautenhahn K, Nehaniv CL. (MIT Press, Cambridge) pp. 211-228.
10. Rietveld E, Kiverstein J (2014) “A Rich Landscape of Affordances.” *Ecol Psychol* 26(4):325–52.
11. Uleman JS, Saribay SA, Gonzalez CM (2008) Spontaneous Inferences, Implicit Impressions, and Implicit Theories. *Annu Rev Psychol* 59:329-360.
12. Fiske S, Taylor S (2013) *Social Cognition: From Brains to Culture*, 2nd Ed. (SAGE Publications, Los Angeles).
13. Hoppitt W, Laland K (2013) *Social Learning: An Introduction to Mechanisms, Methods, and Models*. (Princeton University Press, Princeton).
14. Happé F, Cook J, Bird G (2017) The Structure of Social Cognition: In(ter)dependence of Sociocognitive Processes. *Annu Rev Psychol* 68:243-267.
15. Fiske ST, Lin MH, Neuberg, SL (1999) The continuum model: Ten years later. *Dual-process theories in social psychology*, eds Chaiken S, Trope Y. (Guilford Press, New York).
16. Greenwald AG, Banaji MR (1995) Implicit Social Cognition: Attitudes, Self-Esteeming and Stereotypes. *Psychol Rev* 102(1):4-27.
17. Greenwald AG, McGhee DE, Schwartz JLK (1998) Measuring Individual Differences in Implicit Cognition: The Implicit Association Test. *J Pers Soc Psychol* 74(6):1564-1480.

18. Tylén K, Allen M, Hunter BK, Roepstorff A. (2012) Interaction vs. Observation: distinctive modes of social cognition in human brain and behavior? A combined fMRI and eye-tracking study. *Front Hum Neurosci* 6(331): 1-10.
19. Katz E, Lazarsfeld PF (1955) *Personal Influence* (Free Press, New York).
20. Bass FM (1969) A New Product Growth for Model Consumer Durables. *Manage Sci* 15(5):215-27.
21. Onnela JP, Reed-Tsochas F (2010) Spontaneous emergence of social influence in online systems. *Proc Natl Acad Sci USA* 107(43):18375-18380.
22. McDonnell T, Bail C, Tavory I (2017) A Theory of Resonance. *Sociological Theory* 35(1):1-14.
23. Baronchelli, Andrea (2018) The emergence of consensus: A primer. *R. Soc. Open Sci* 5:172189.
24. Shields P (1998) The Interactions Between Ergodic Theory and Information Theory. *IEEE Trans Inf Theory* 44(6):2079-2093.
25. Simmel G (1903) The Metropolis and Mental Life. *The Sociology of Georg Simmel*. (New York: Free Press) pp. 1-12.
26. Mead GH (1934) *Mind, Self and Society*. ed Morris CW. (University of Chicago Press, Chicago).
27. Goffman E (1959). *The Presentation of Self in Everyday Life*. (Anchor Books, New York).
28. Barker RG (1968) *Ecological Psychology: Concepts and Methods for Studying the Environment of Human Behavior*. (Stanford University Press, Palo Alto).
29. Gibson JJ, (1977) The Theory of Affordances. *Perceiving, Acting and Knowing: Toward an Ecological Psychology*, eds Shaw R, Bransford J. (Lawrence Erlbaum Associates, Hillsdale, New Jersey).
30. Bourdieu P (1984) *Distinction: A Social Critique of the Judgement and Taste*. (Harvard University Press, Cambridge).
31. Hutchins E (1995) *Cognition in the Wild*. (MIT Press, Cambridge).
32. Law J and Mol A (1995) Notes on Materiality and Sociality. *Sociol Rev* 43(2):274-94.
33. Latour B (2005) *Reassembling the Social: An Introduction to Actor-Network-Theory*. (Oxford University Press, Oxford),
34. Tilly C, Keane W, Küchler S, Rowlands M, Spyer R, eds. 2006. *Handbook of Material Culture*. (SAGE, Los Angeles).
35. Browning CR, Calder CA, Soller B, Jackson AL, Dirlam J (2017) Ecological Networks and Neighborhood. *Amer J Sociol* 122(6):1939-88.
36. Centola D, Baronchelli A (2015) The spontaneous emergence of conventions: An experimental study of cultural evolution. *Proc Natl Acad Sci USA* 112(7):1989-1994.
37. Lewis D (1969) *Convention: A Philosophical Study*. (Harvard University Press, Cambridge).

38. Young HP (1993) The Evolution of Conventions. *Econometrica* 61(1):57–84.
39. Garrod S, Doherty G (1994) Conversation, co-ordination and convention: An empirical investigation of how groups establish linguistic conventions. *Cognition* 53(3):181–215.
40. Marmor A (2009) *Social Convention: From Language to Law*. (Princeton University Press, Princeton)
41. Baronchelli A, Felici M, Loreto V, Caglioti E, Steels L (2006) Sharp transition towards shared vocabularies in multi-agent systems. *J Stat Mech* P06014.
42. Baronchelli A, Loreto V, Steels L (2008) In-depth analysis of the Naming Game dynamics: The homogeneous mixing case. *Int J Mod Phys C* 19(05):785–812.
43. Centola D, Becker J, Brackbill D, Baronchelli A (2018) Experimental evidence for tipping points in social convention. *Science* 360(6393):1116–1119.
44. Centola D, Macy M (2007) Complex Contagions and the Weakness of Long Ties. *Am J Sociol* 113(3):702–734.
45. Watts D (1999) Networks, Dynamics, and the Small-World Phenomenon. *Am J Sociol* 105(2):493–527.
46. Lin J (1991) Divergence Measures Based on the Shannon Entropy. *IEEE Trans Inf Theory* 37(1):145–151.
47. Granovetter M (1973) The Strength of Weak Ties. *Am J Sociol* 78(6): 1360–1380.
48. Burt R (1992) *Structural Holes*. (Harvard University Press, Cambridge).
49. Flack J (2010) Multiple time-scales and the developmental dynamics of social systems. *Philos Trans R Soc Lond B Biol Sci* 367(1597):1802–1810.
50. Foster J (2018) Culture and computation: Steps to a Probably Approximately Correct theory of culture. 68:144–154.
51. Daniels B, Krakauer D, Flack J (2017) Control of finite critical behaviour in small-scale social systems. *Nat Commun* 8(14301):1–8.
52. Simmel G (1955) *Conflict and the Web of Group-Affiliation*. (The Free Press, New York).
53. Lorrain F, White HC (1971) Structural equivalence of individuals in social networks. *J Math Sociol* 1(1):49–80.
54. Blau P (1977) *Inequality and Heterogeneity: A Primitive Theory of Social Structure*. (The Free Press, New York).
55. Bourdieu P (1998) *Practical Reason: On the Theory of Action*. (Stanford University Press, Stanford)
56. DiMaggio P (1997) Culture and Cognition. *Annu Rev Sociol* 23:263–287.
57. Vaisey S (2010) Justification and Motivation: A Dual-Process Model of Culture in Action. *Am J Sociol* 114(6): 1675–1715.
58. Cerulo K (2010) Mining the Intersection of cognitive sociology and neuroscience. *Poetics* 38(2):115–132.

59. Bail C, Argyle L, Brown T, Bumpus JP, Chen H, Fallin Hunzaker MB, Lee J, Mann M, Merhout F, Volfovsky A. (2018) Exposure to opposing views on social media can increase political polarization. *Proc Natl Acad Sci USA* DOI:10.1073/pnas.18004840115.
60. Chen D-L, Schonger M, Wickens C (2016) oTree-An Open-Source Platform for Laboratory, Online, and Field Experiments. *J Behav Exp Finance* 9:88–97.
61. Casler K, Bickel L, Hackett E (2013) Separate but Equal? A Comparison of Participants and Data Gathered via Amazon’s MTurk, Social Media, and Face-to-Face Behavioral Testing. *Computers in Human Behavior* 29(6): 2156–2160.
62. Crump M-J-C, McDonnell J-V, Gureckis, T-M (2013) Evaluating Amazon’s Mechanical Turk as a Tool for Experimental Behavioral Research. *PLoS One* 8(3):e57410.
63. Hauser D-J, Schwarz N (2016) Attentive Turkers: MTurk Participants Perform Better on Online Attention Checks than Do Subject Pool Participants. *Behav Res Methods* 48(1):400–407.
64. Mason W, Suri S, (2012) Conducting Behavioral Research on Amazon’s Mechanical Turk. *Behav Res Methods* 44(1):1–23.
65. Paolacci G, Chandler J, Ipeirotis, P (2010) Running Experiments on Amazon Mechanical Turk.” *Judgm Decis Mak* 5(5):411–419.
66. Rand D-G (2012) The Promise of Mechanical Turk: How Online Labor Markets Can Help Theorists Run Behavioral Experiments. *J Theor Biol* 299:172–179.
67. Shank D-B (2016) Using Crowdsourcing Websites for Sociological Research: The Case of Amazon Mechanical Turk. *Am Sociol* 47(1):47–55.
68. Chen D-L, Schonger M, Wickens C (2016) oTree-An Open-Source Platform for Laboratory, Online, and Field Experiments. *J Behav Exp Finance* 9:88–97.
69. Ipeirotis P (2010). Analyzing the amazon mechanical turk marketplace. *XRDS: Crossroads*, 17(2):16–21.
70. Smith V-L (1976) Experimental Economics : Induced Value Theory. *Am Econ Rev* 66(2):274–279.
71. Hagberg A-A, Schult D-A, Swart P-J (2008) Exploring Network Structure, Dynamics, and Function Using NetworkX. *Proceedings of the 7th Python in Science Conference (SciPy 2008)*:11–15.

Figures and Tables

Fig 1. Game Dynamics. Rows identify the number of additional names. Column one (A-C) shows by round the number of names in circulation, defined as having been played (participants occasionally submit new names after the first round) and possibly repeated in the present or a future rounds. (Names not occurring in the present round can reappear because participants may alternate names between rounds.) Column two plots by round the normalized HHI concentration index of the distribution of names in circulation. Column three (D-F) plots the fraction of partners who successfully coordinated, as a 3-value moving average. The network topology and outcome of each trial is coded according to the key in panel C.

Fig 2. Divergence Differences. Each panel plots the population average and standard deviation of participants' divergence difference (see Methods section) and the divergence ratio (see Methods section), for all games in the treatment identified by the row. The first column (panels A-F) plots the differences and ratios when the comparison case is Random Neighbors. The second column (G-L) compares to Random Others, the third (M-R) to Weak Ties, and the fourth (S-X) to Max Info Others.

Fig 3.

Information Gain by Number of Additional Names. Each panel plots divergence differences (see Methods section) as a surface over the space of rounds and the additional number of names. The first column (A-C) contains only the successful (i.e. a single convention emerged) trials for the network topology identified by the row and with one or more additional names. The second column (D-F) contains the failed runs and the third (G-I) contains the trials with no additional names.

Table 1. Experiment Results. Each box in the table represents one trial of the experiment for the given treatment combination. If the box is green, a global convention emerged, if red, it did not. Some conditions had fewer trials because i) they were replicating the results in C&B, or ii) in the case of the full network, very likely to produce global conventions, or iii) in the case of the one-name lattice treatment, very unlikely to produce a global convention. Network topology treatments are arranged from left to right by increasing average longest shortest path lengths.

Add'l Name Count	Network Topology			
	Full	Random	Small W	Lattice
0 names
1 name
2 names

= Convention Emerges
 = No Convention

Table 2. Unstructured Source Influence on Name Choice. Each cell contains the grand-average for the applicable runs of the fraction of times participants matched on a name they had i) first seen via the unstructured source, and ii) had only seen via the unstructured source.

Additional Name Count		Outcome	
		Convention	No Convention
1 name	Before	.42	.31
	Only	.09	.10
2 names	Before	.54	.54
	Only	.13	.16

Supporting Information

Subject Recruitment. All subjects were recruited through Amazon's Mechanical Turk marketplace (Mturk). This platform enables the recruitment of workers to complete online tasks and has been validated as a source of subjects for a wide range of behavioral experiments [1-7]. To build the subject pool, any worker with a reasonably healthy history of work on Mturk was offered compensation to complete a short training module and take a comprehension test. The task paid well for the amount of time it took, so it was easy to recruit workers into the subject pool. This initial interaction familiarized subjects with the interface and game structure, but also satisfied Mturk's requirements for uniquely identifying and communicating with the worker in the future. This allowed the creation of a well-defined subject pool and control over which workers were permitted to join live sessions of the experiment. The day before a scheduled trial, a number of subject pool members were invited to play. Following the recommendation of Mason and Suri [4], roughly four times as many pool members were invited than necessary to run the experiment. Those invitees were contacted shortly before the scheduled game as a reminder and again at the start of each session in order to provide a link to the interface.

The subject pool included slightly more than 300 members at any given time, as members were removed periodically for being unresponsive to invites and then replaced. Pool members were allowed to complete up to five sessions; the average number of completed sessions was 2.95, with an average of 20 days between sessions. For each game a list of invitees was created from this pool using the following criteria; No participants from the last session can be included; Those who were recently invited but did not start a session receive the highest priority; those who have yet to be invited get the next highest priority; those who have been invited but have not played get the next highest priority; for the remaining pool members, priority is less than the previous categories and inversely related to how recently they completed a session. If ever a set of members of equal priority was larger than what is necessary to create the list of invitees, the correct number of members was added by sampling uniform at random from that set. The above procedure was repeated until no more than four participants who completed the same previous session were included.

Game Play Details. The game interface was created using the oTree, an open source python platform for experiments [8]. As participants arrive for a session, they are shown a brief description of the task and the IRB statement of risk (Exempt-status). If they accept the task, they are shown the game instructions, which include a small comprehension test. Next, they must agree to not use means outside of the interface to attempt to coordinate. This establishes grounds for dismissal, but the real controls against such attempts are built into the software (see Software and Subject Management Details). Once they agree, they are taken to a waiting page until enough participants arrive. Subjects are paid one cent for every five seconds of waiting and are shown how long they have been waiting and their bonus for that time. This greatly increases the retention of subjects and their satisfaction with the task and compensation, which is crucial for maintaining a responsive pool.

Once enough participants have arrived, the game begins. Participants are shown a headshot of an individual and are asked to submit a name for her. Participants have 15 seconds to submit it and failure to do so in that time results their submission being recorded as null. Fig S1 is an example of this submission page. After everyone has submitted, the results are displayed. This differs from the design in the original experiment, which allowed for an asynchronous progression through rounds. While it seems unlikely to make a difference, synchronizing the rounds was necessary in this design because the second treatment factor required sampling from all names played in a round. To maintain commensurability, rounds were synchronized for all treatment levels, a choice that ultimately did not have an effect.

Partners who successfully coordinated earn \$0.10. Partners who did not coordinate are penalized \$0.05, unless their current cumulative total is already \$0.00. The results page always shows the names each partner played, whether they matched, the amount of the reward or penalty, their current cumulative rewards and their history of successes or failures through the previous rounds (but not the names associated with those successes or failures). In the treatments where it was applicable, additional names were displayed. Fig S2 is an example of the results page. This page times out after 10 seconds. The progression through the submission, waiting, and results pages constitute of round of play. There are 25 such rounds in each session. After the 25 rounds, the final page reported the participant's final earnings, broken down into base pay, waiting bonus and game bonuses, and offered the opportunity to leave feedback.

Software and Subject Management Details. In order to ensure the subjects' experience corresponded to the ideal experiment and that the experiment is internally valid, there are a number of important details to the design of the software and management of subjects. The features described below were added after reviewing data from trial runs of the experiment and additional research on the experience from the perspective of the workers done through direct (and compensated) correspondence with some and by reviewing forums that host robust communities related to working on Mturk (e.g. MturkNation, TurkerHub, Reddit).

In general, workers on Mturk are interested in participating in academic research but not at rates lower than they typically earn. In fact, because a new academic researcher lacks a good reputation (Turkopticon is an example of a system to create and disseminate reputation scores for all job posters) and presents a risk to workers, a premium on the typical rate can be necessary to ensure data quality, especially if the task takes more than a couple minutes. The modal worker is not doing tasks for entertainment, but rather to supplement income [5, 9]. This makes them sensitive to the effective hourly wage, and, although there is variability in their reservation wages, most workers seem to have the federal minimum wage in mind. Because most choose to accept a task based on a rough calculation of the effective hourly wage, any misrepresentations of estimated earnings or time to completion are likely to provoke ire.

Importantly, because workers think in terms of an effective hourly wage, the compensation structure of the experiment needs to correspond to the workers' overall incentive structure in the Mturk marketplace. In particular, a significant portion of the overall payment must be guaranteed for the completion of the task. Workers always have the opportunity to quit the current task and start a different one and therefore are often aware of a changing opportunity cost. If the task

progresses slowly or bonus earnings appear lower than expected or advertised, a small guaranteed payment might lead them to exit the task in spite of already having committed time to it. It is important then that the subject's current expected payment is roughly at or above the subject's reservation wage for that amount of time. In practice this means the guaranteed payment should be large relative to potential bonus earnings and idle time should be additionally compensated. The former incentive structure generally corresponds to rational behavior in regards to opportunity costs, but furthermore does not appear to undermine the ability of within-game incentives to "induce preferences" in the sense of Smith [10]; once the guaranteed payment meets the reservation wage, in-game bonuses become an exciting opportunity to exceed the worker's earnings goal and workers try in earnest to maximize bonus earnings.

The compensation of idle time should in principle not be necessary if that time is included in the estimate of the time necessary to complete the whole task. However, workers generally view idle time as distinct from time spent on the task, likely because the experience of waiting makes them more aware of opportunity costs. Casual experimentation in the design of this experiment led to an idle-time compensation structure that delivered an effective hourly wage close to the prevailing reservation wage, broken into small increments. Paying one cent for every five seconds of waiting ensured a wage of \$7.20/hour and that any amount of "unpaid" time was very small. Additionally, live tracking of the elapsed time and accrued bonus on the wait page created a gratifying experience for workers of being able to watch earnings grow.

A complication of this overall compensation structure is that it is more challenging to implement a "show-up" payment often used with traditional subject pools. Ensuring that the required 24 subjects are available requires over-recruiting, carefully monitoring arrivals and immediately removing the task from the Mturk listings once enough accept. Still, this often results in subjects in excess of 24. The Mturk platform does not allow people to be turned away once the task has been accepted, but one can ask workers to "return" the task. Unfortunately this precludes being able to compensate them through the normal means, requires them to trust the task poster to follow up, and generally risks upsetting them, possibly causing reputational damage. Furthermore, given that workers have been personally invited and have been watching for an email to start the task, sending them off with a smaller show-up fee might harm their responsiveness to future invitations. While it increased costs, the long-term solution used here was for subjects in excess of 24 to play a version of the game against bots designed to make moderately intelligent choices. None of those subjects ever gave any indication they knew they were not doing the real task (although interestingly, some playing live participants commented that they thought they were playing bots). Subjects who played against bots were paid what they earned and were later dropped from the analysis.

Once subjects joined a session (after having read a brief description of the task and the IRB statement), they were shown the game instructions. This page had a comprehension task built in; rather than clicking a standard button to advance to the next page, subjects were directed at the end of the instructions to click a button hidden as the text "match names with playing partners" where it first appeared in the instructions. Knowing to click these words and scanning for them reinforced the basic task of the game and ensured active engagement. Roughly 10% of would-be

subjects never made it passed this stage, although it was not possible to tell if they left the game for other reasons. Any subject who did not pass this comprehension test was unable to join the game and was forced to return the task.

Those subjects who passed the instructions and comprehension task were then asking to agree to not attempt use means outside of the game's interface to coordinate. This is a real concern for a population that has a robust set of forums and chat rooms dedicated to its community. I describe the primary software and management mechanisms used to defeat such attempts below. By asking subjects to agree to not use external means of coordination, a very weak form of control, I hoped to make subjects think twice about engaging in such behavior, but more importantly, create defensible grounds for removal from the subject pool if there was evidence of attempts to coordinate.

If subjects agreed to not use external means of coordination (all did), they were taken to a page to wait until enough subjects arrived. This page showed them the number of minutes and seconds they have been waiting and their compensation for that wait. Once at least 24 subjects made it to the wait page, the game began. The first page of each round displayed a headshot of an individual and a field to enter a name. An example of this page appears in Fig S1. With only two exceptions, the picture was always a younger woman because pretesting showed that other demographics had some focal points; photographs of middle-aged and older men without distinctive characteristics often quickly elicited a handful of competing alternatives (e.g. Bob, Mike, Bill). The same was true of older women (e.g. Janet, Susan, Patty). Pictures of younger women (roughly 25-45) elicited a broad range of first submissions. Participants had 15 seconds to submit a name and were taken to a wait page until all participants submitted names. On the results page, participants were shown the name they submitted, the name their current partner played, and, when applicable, the names from the unstructured source. They were also explicitly told whether the names matched, their payoff for the round, their cumulative earnings and their history of matches and failures (but not the related names). This page timed out after 10 seconds, although participants could advance past it to a wait page. An example of this page appears in Fig S2.

I took several measures to ensure there was no external collusion among the subjects. Together these measures address multiple potential means of collusion and work together to all-but-ensure it did not happen. As subjects first arrive, the related IP addresses are screened to make sure there is only one subject per IP-address. The first subject to arrive from any address is allowed to remain, but all others are blocked from continuing, informed why, and asked to return the task. It is not uncommon for Mturk workers to work in the same household or workspace and while they might honestly avoid collusion, I erred on the side of caution and permitted only one. The more problematic case is a worker with multiple Mturk accounts. This is a violation of the Mturk Terms of Service agreement, but by their own admission on forums, some workers use multiple accounts. Those willing to ignore such rules might also try to use virtual private networks to use different IP addresses for each account, but they would have had to do that before arriving. Furthermore, in virtue of how invitation groups are chosen, the probability of two or more of their worker IDs being present in the same group is very low. This is because the experimenter

controls which worker accounts receive an invitation to a game and have the necessary “qualification” for it. (A qualification is a virtual token within the Mturk platform that can be used to control which workers can do which tasks.) Without an invitation and the requisite qualification, workers are unable to join a session. I used qualifications to ensure that no more than 4 workers ever played the same session together. The final precaution I took to protect against in vivo–collusion was to make sure there was no relationship between arrival time and the location in the network; while small-scale efforts at collusion could be successful if those colluding are network neighbors, the nature of the game renders such efforts ineffective or possibility counterproductive if would-be colluders are not network neighbors.

Within the game itself, there are several features to prevent collusion. The first is actually just a basic design of the game; participants have only 15 seconds to submit a name. Failure to submit costs a participant rewards and can lead to expulsion from the game, a fact participants are reminded of every time they fail to submit. Given this time constraint participants typically submit immediately (~5 seconds). Even those who do not submit immediately have very little time to attempt to communicate with other participants about emerging patterns in the names. Timed submissions do nothing to protect against premeditated attempts at collusion, however. Given that workers know the date and time of the game in advance, they might seek each other out on forums in advance and agree to use the same name. I have found no evidence of this on the forums I was able to gain access to (some have very high bars for admission), but I nonetheless added more comprehensive features to defeat such efforts. The primary means is a screening and blocking of names that exhibit surges in frequency. A name that appears for the first time with more than two instances is “blacklisted” for that round. It is not farfetched for the name “Sarah” to be the first submission for multiple participants, but, erring on the side of caution, my software barred any name that first appeared with three or more instances. When a name was barred, those submitting the name were told the name is blacklisted, but not why. Partners of those submitting the barred name were informed that their partner had not submitted a name, not that it had been barred. Multiple names could be barred in any given round, but the barred status is confined to that round only. If one or two instances of a previously barred name are submitted in a subsequent round, it is accepted without comment. This is because participants frequently exhibit the suboptimal behavior of introducing new names well into the game (this behavior was also observed in the original experiment) and there is no reason for barring these submissions. This barring mechanism was triggered a total of 20 times in the 825 total rounds of play. All instances occurred in the first round of play. 12 of instances were three participants submitting the same name. Five instances included multiple names being barred. The largest incident was the submission of the same name by 9 participants. The barring mechanism effectively removed the name from circulation, as participants immediately abandoned it, meaning that a barred name never became a global convention. Interestingly, only 5 of the 20 trials for which the barring mechanism was trigger ultimately resulted in a global convention, suggesting something about the barring mechanism or the behavior it targets might have harmed the prospects for future success. It is clear that, whether premeditated or coincidence, these instances did not contribute the emergence of conventions, although they might have inhibited it

in some instances, leading to the underestimated of the effect of the presence of the unstructured source.

The screening of names for the purpose of barring is only applied to names that have not yet appeared in the session. Once a name has been successfully introduced, it might still be implicated in collusion efforts happening outside of the game interface. If the name “Sarah” is submitted by a single participant in the first round, but then by 10 participants in the second round, collusion could be the reason. That is not guaranteed to be the case, however. After a participant submits the name, their partner sees it and is very likely to play it. In treatments where random names are also shown, any number of other players may have seen the name on their results page and accordingly played it. Given this structure, preventing collusion requires judging whether a name is spreading around the network too quickly. To do this, I track each participant’s exposure to names, meaning simply any name they have seen or successfully played (i.e. was not barred). Once a participant has been exposed to a name, I assume it is completely reasonable for them to play it and therefore the spread of that name to this participant did not happen outside the confines of the game interface. One could have a more demanding definition of exposure that considers the likelihood that participants will forget names seen many rounds ago, but implementing a suitable rule for the game would require making specific assumptions about how participants process the information that might not be justified.

To reconcile the idea of proper exposure with the fact the participants contribute names they have not been exposed to, only one player per round is permitted to play a name they were not properly exposed to. If two or more participants submit a name they were not properly exposed to, the name is barred for all such participants and their submission is recorded as a null. Others playing the same name but who have been properly exposed do not have their submission barred. While not a foolproof system for guarding against impermissible coordination, it would take an impressive amount of organizing by the participants to defeat the system and it still grants participants significant latitude in name choice. This type of screening happened only 6 times in the 792 applicable rounds of play and none of the affected names went on to become a global convention.

The two mechanisms of barring name described above offer significant protection against any organized effort on the part of participants to game the system. Another possible way to game the system would be to submit natural focal point words, such as “woman” or “name”. Such efforts would have the virtue of being of not needing to be centralized. This potential problem is addressed simply by creating a list of words the input field will not accept, thereby preventing the participant from submitting and advancing to the next page. Whereas barring surges in names requires the names to have been submitted and tallied, focal point words can be screened out before submission and participants can still submit a valid name. Roughly twenty focal point words, mostly nouns, including a number of obscenities, were screened-out. Furthermore, any names that appeared in the trainings module were added for the obvious reason that participants could be primed to focus on them.

A few additional design features work to ensure the integrity of the experiments. The pictures of the individuals to be named by participants were changed every two games. No worker in the

subject pool can play two games in a row, so the changing the picture every two games ensured no worker ever saw the same picture. The usual means of downloading images was also disabled for the game just in case workers considered attempting to use the image as a means of coordination on forums.

As the above description of the software suggests, using Mturk as a subject pool presents some challenges and still has some of issues any subject pool can have, but most of these issues can be addressed through proper subject pool management and software design. The attempts to game the system were successfully defeated and, if anything, led to the underestimation of the effect of unstructured information sources. In principle a very sophisticated effort built around knowledge of my defeat devices could go undetected by my software, but even in the highly improbable event such an effort succeeded regularly, one would still need to account for the consistent pattern in the results; the information dynamics play a crucial role in determining group-level outcomes. Finally, related to the consistency of the results, the fact that subject pool members often participated in several games allowed them to learn from their experience and possibility improve their play overtime. Because there was no systematic relationship between treatment variables and when the trial fell in the sequence, the possibility of improved game play only bolsters the central pattern in the results; when a group is able to find a global convention does not depend on the skill of the participants but rather the information dynamics.

Network Design Details. All networks were generated using the python package NetworkX [11]. The small world networks had a rewiring probability of .2 and were guaranteed to be connected. These graphs were used to create pairs of neighbors for the 25 rounds of each trial. Participants need a partner for each round, so the set of pairings for each round was chosen random from among complete pairings. Because the sequence of pairings could have its own effect, the sequences were reused for the different treatment levels of unstructured information. For example, the 12 trials with small world networks used only four network-pairing sequence combinations, reusing each of the four once in each information treatment. There was no obvious effect of pairing sequences so no further analysis was pursued.

Time left to complete this page: 0:08

Name Game

Choose a **first name** for this person. You are trying to match with one of your playing partners.

\$0.00

Last reward

\$0.00

Total Earnings

Your Choice:

Next

Round 1 : Playing

Round 2 :

Round 3 :

Round 4 :

Round 5 :

Round 6 :

Round 7 :

Round 8 :

Round 9 :

Round 10 :

Round 11 :

Round 12 :

Fig S1. Name Game Submission Page. The page displays the time remaining to submit, the total earnings, the last reward (or penalty), the current round and round histories.

Time left to complete this page: 0:08

Name Game

-\$0.05

Last reward

\$0.05

Total Earnings

Random players: Liz

You:

Sarah

Partner:

Jennifer

No match

Next

Round 1 : No Match

Round 2 : Match

Round 3 : No Match

Round 4 :

Round 5 :

Round 6 :

Round 7 :

Round 8 :

Round 9 :

Round 10 :

Round 11 :

Round 12 :

Fig S2. Name Game Results Page. The display includes the submission of the participant, partner and a random player and the history of success.