Homo Socialis and *Homo Sapiens*

Robert L. Goldstone[1]*

[1]*Department of Psychological and Brain Sciences, Indiana University,*
rgoldsto@indiana.edu

ABSTRACT

The assumption that individuals are behaving rationally can, at times, usefully constrain predictions of individual and collective behavior. However, success in predicting human and group behavior will often require relaxing this assumption of rationality, instead employing evolutionary, neural, and cognitive constraints. One particularly important form of neural and cognitive constraint is that interacting individuals each possess a network of concepts, and communities are accordingly social networks of neural networks. The structured nature of human conceptual systems suggests that communicating is better modeled as a process of *aligning* conceptual systems rather than simply *transmitting* atomic beliefs. Communicating individuals can establish norms, conceptual structures, and rule systems that did not preexist prior to the communication process. For this reason, the dichotomy between rule-based and centralized groups versus self-organized and decentralized groups is false – one of the major activities that self-organized and decentralized groups engage in is the establishment of rules, laws, norms, leaders, and institutional hierarchies that will then govern their subsequent interactions.

Explaining how patterns of collective behavior emerge from interactions among individuals with diverse, sometimes opposing, goals is a societally crucial and particularly timely pursuit. It is timely because humans are more tightly connected to one another now than ever before. From 1984 to 2014, there has been more than a million-fold increase in the number of devices that can reach the global digital network. Although web technology is new and transformative,

from a broader perspective, it is also just a recent manifestation of humanity's perpetual drive to become more intermeshed. Earlier manifestations of this drive include the printing press, global transportation networks, telecommunication systems, and the academy. These social networks have catalyzed the formation of otherwise unattainable social patterns. Understanding the origins and possible destinations of these social patterns is both scientifically and pragmatically consequential.

In the article "*Homo Socialis*: An Analytical Core for Sociological Theory," Gintis and Helbing construct a foundation for understanding collective behavior patterns. This foundation is grounded in rational agents that follow Savage's axioms of completeness, transitivity, and independence from irrelevant alternatives. This assumption of rational individual agents is supplemented by a group-level assumption that the agents communicate and coordinate with each other across social networks. The framework adopts a complex systems perspective on social structures and patterns, and in this regard offers an explanatorily satisfying account for how social patterns arise and evolve. Much in the same way that physics explains macroscopic properties like temperature in terms of interactions between molecules moving at different speeds, the complex systems approach to social phenomena explains macroscopic phenomena such as the spread of gossip in a community, the establishment of a common currency, and the spontaneous formation of social norms in terms of interactions between agents. One of this approach's appeals is that macroscopic phenomena need not be assumed. Their patterns can naturally fall out of the interactions among hypothesized agents whose behavior is aptly constrained.

While one's naïve presumption may be that humans are limitless, for the purposes of explaining and predicting behavior, it is a virtue for a theory if the behavior of its agents is constrained. Constraints make the theory testable and capable of generating genuine predictions. If the behavior of the agents were unconstrained within the theory and they could behave in virtually any fashion, then many different collective-level patterns could be derived, but at the cost that none of the patterns would be strongly predicted. Good theories expose themselves to the possibility of being disconfirmed. They make predictions that could be empirically wrong, and if they are not wrong, then their success is bona fide. Constraints properly force a theory to "stick its neck out" and risk disconfirmation.

1   Rational, Evolutionary, Neural, and Cognitive Constraints

In the theoretical foundation adopted by Gintis and Helbing, individuals are constrained to follow the dictates of rationality. This is a prevalent constraint in economics and psychology theories, and has the prominent advantage of generating surprisingly strong predictions from intuitive and rather minimal

assumptions. Gintis and Helbing describe a compelling example of the power of this assumption. A rational actor assumption supplemented by straightforward assumptions about an economy containing multiple resources unequally distributed among individuals is sufficient to predict the establishment of a common standard of currency based on one of the resources. The emergence of a common currency from the collectively traded resources is a persuasive example of a group-level pattern that emerges from individuals who are motivated by their own concerns but none of whom are trying to attain the group-level pattern.

Constraining agent behavior by rationality is an effective way to generate strong predictions for patterns that can be observed at the collective level, but is by no means the only one. Other possible sources of constraint are *evolution*, *neuroscience*, and individual *cognition*. Evolution imposes constraints on individual human behavior because rational behavior, no matter how advantageous, will not be observed unless there is an evolutionary pathway that can achieve it. In fact, humans and other animals are adapted to specific problems such as mate finding, food gathering, and threat detection, rather than general-purpose rational inference or optimal decision making. Evolutionary considerations can explain why humans generally fail at rational reasoning tasks that involve logical operations such as modus tollens (If P then Q, Not Q, therefore Not P) when they are given an arbitrary, decontextualized problem, but succeed when give an adaptively relevant instantiation of the same logic problem involving the detection of cheaters (Barkow *et al.*, 1992). Generally put, decision making mechanisms cannot be guaranteed to perform well outside the situations in which they evolved (Trimmer and Houston, 2014). By considering how close a target context is to an organisms' evolutionary relevant context, we can predict how close people will come to optimal decision making. A good example of the difference between ecologically stable strategies and the kind of rational agent specified by Savage's axioms is Trimmer *et al.*'s (2012) application of genetic programming to evolve agents to learn decisions under varying environmental conditions and subject to mutation. Their simulated, evolved agents tended to evolve the empirically well supported Rescorla-Wagner rule

$$\text{new\_estimate\_of\_value} = \text{old\_estimate} + \alpha \, (\text{actual\_value} - \text{old\_estimate})$$

to guide their behavior rather than adopting optimal decision rules, because the Rescorla-Wagner model was robust to suboptimal parameter settings whereas the optimal decision rule often performed very poorly with small parameter mutations. A consideration of evolutionary constraints suggests the persistence and ascendency of rules of behavior and learning that are not optimal as long as the rules are robust to noise and reasonably apt across a range of environments.

Neural constraints feature prominently in accounts of human decision making. As with evolutionary constraints, if there is not a neural architecture that can subserve rational decision making, then individual rationality will not be observed. Although there are some cogent proposals for neural implementations of rational decision processes (Courville *et al.*, 2006), there is also widespread evidence for decision making phenomena that would seem peculiar and inexplicable without a consideration of their neural foundation. For example, people are better at emitting a behavioral response in anticipation of reward, and better at withholding a response in anticipation of punishment, compared to the opposite pairings (Guitart-Masip *et al.*, 2012). Exactly this dissociation is predicted by the underlying differential anatomy of the substantia nigra and inferior frontal gyrus, and an understanding of how they figure in instrumental learning.

A third form of constraint is cognitive. When humans make decisions, they are constrained by limitations in their perceptual sensitivity, finite working memories, bottlenecks in attention, inefficiencies in integrating information across channels, and restricted abilities to think ahead and plan for uncertain outcomes. Previously learned information that is relevant for a current task is frequently ignored simply because of the computational difficulties in retrieving the information (Ross, 1987). Limitations in perception, attention, memory, and reasoning arise because often the choice between making quick, adequate decisions requiring few computational resources versus protracted, optimal decisions requiring many resources favors the former. As Homo sapiens, we live our lives in real-time, and our imperfect constraints for selecting objects to attend, choices to select, and paths to pursue suffice often enough to meet most of our situational needs (Gigerenzer and Todd, 1999; Simon, 1956).

Together, these three forms of constraints offer empirically well supported predictions of human behavior. Constraints of rationality can also account for some of the empirical evidence for evolutionary, neurological, and cognitive constraints on decision making. For example, failures of humans to exhibit transitivity of preferences can be reconciled with rational choice if we posit that preferences are changing from moment to moment. However, the specific and systematic nature of demonstrated intransitivities is more satisfyingly explained if we invoke perceptual and cognitive limitations. People tend to choose A over B, B over C, and C over A when there is a dimension along which A has a large advantage over B (and two dimensions along which B is only slightly superior), another dimension along which B has a large advantage over A (and slight inferiority on two other dimensions), and a third dimension along which C has a large advantage over A (Tversky, 1969). People pay more attention to dimensions along which there is a large difference among the choices (Mellers and Biagini, 1994), and cognitive constraints like this make well constrained and highly specific predictions for individual decisions. Constraining agents not only by rational considerations, but by evolvability,

neurological implementation, and cognitive function is likely to pay large dividends when we are trying to predict the behavior of groups comprised by actual Homo sapiens.

Human choice behavior does not always satisfy the axioms of rational decision theory, such as the independence of irrelevant alternatives and transitivity of preferences (see Shafir and LeBoeuf, 2002 for a review). One strategy for dealing with these apparent violations of rationality is to explain them away. Perhaps the choices can be construed as still being rational in some sense, or are rare enough so as to be practically insignificant. A different strategy is to develop positive, mechanistic accounts of the processes that people employ when making real-world decisions. Decisions that are mysterious if we do not take cognitive, neural, and evolutionary processes into account can be dramatically clarified when we do. For example, Camerer *et al.* (1997) analyzed taxi cab drivers' log books of fares taken and hours worked to conclude that taxi drivers tend to stop working early on days that are relatively lucrative in terms of fare income per hour. This apparently illogical tendency persists over years of cab driving experience even though the drivers could be earning 7.8% more if they stopped work at the same time every day, or 15.6% more if they adopted the sensible strategy of working more on lucrative days. This systematic daily stopping behavior could potentially be explained as rational under a reconstrual of rationality framed in terms of maximizing local pleasure per day or minimizing cognitive effort. A more productive approach, though, is to follow Camerer *et al.*'s suggestion that people adopt an aspiration level for their daily wage and when that target is reached, they stop working for the day. One reason for preferring psychological accounts like this, or one in which professional golfers are critiqued as being irrationally loss averse (Pope and Schweitzer, 2011), is that the decision makers themselves often agree that their decisions were irrational when apparent suboptimalities are brought to their attention, and then modify their subsequent decisions.

These evolutionary, neural, and cognitive constraints are presented not to argue against the fertility of a rational agent assumption. There will probably not be a singular answer to the question of what the best level of analysis for predicting individual and collective behavior is. These additional constraints are proposed here because they are likely to be at least as productive for useful predictions of individual human behavior as is an assumption of rationality (see also Jones and Love, 2011, for a similar conclusion regarding the predictive virtues of process-oriented and rational accounts of behavior).

2 Human Social Processes

Gintis and Helbing augment the standard individual-oriented approach to rational agents by a careful consideration of the social networks in which they

are embedded. Agents' judgments and actions are crucially shaped by their interactions within this social network. Social structures from this perspective are complex systems that emerge from individual interactions. To supplement their powerful framework, I would like to consider some human-level factors that govern social interactions and shape the eventual emergent patterns of collective behavior.

2.1   Social Networks of Conceptual Networks

The majority of formal models of individuals interacting in social networks incorporate rather impoverished representations of agents. Often, each individual is represented by a single, albeit time-varying, number (e.g. "Probability of cooperation = 0.8"). In more complicated models, agents are represented by a vector of values across a set of dimensions. This is the basis for Gintis and Helbing's social trait vector representation, although they supplement this vector with filters that determine whether network neighbors will influence each other based on their traits. Actual human knowledge has a much richer structure than a vector. For example, an evolutionary theorist has concepts about natural selection, sexual reproduction, and genetic variability within a population, but these concepts are not independent elements, but rather support and contextualize one another. Concepts gain their meaning by their relations to other concepts (Goldstone & Rogosky, 2004). If each person is to be modeled as a conceptual network, then a social group is to be modeled as network of networks. From a modeling perspective, the intellectual interest is in the study of how these two levels of networks interact. Communicating is not simply transmitting individual concepts. Many agent-based models notwithstanding, social interaction is usually not aptly modeled by switching a single value of an agent's belief vector to another agent's value. Rather, communication involves aligning the conceptual systems of agents (Barr, 2004). One implication of this alignment process is that as concepts migrate across people, they will be systematically altered to fit their owners' individual conceptual network.

   One promising application of this "networks of networks" premise is modeling belief propagation in social networks. It has often times been observed that people are not equally influenced by their network neighbors, but rather are selectively influenced by those neighbors that are similar to themselves on any of a number of dimensions (McPherson *et al.*, 2001). This effect has often been incorporated into models of belief propagation, such as Axelrod's (1997) Culture Model in which the probability of neighboring agents interacting is assumed to be proportional to the agents' similarity. However, even if dissimilar individuals do interact, they may not successfully change one another's opinions because their conceptual networks do not align well. For example, in Kunda and Thagard's (1996) parallel constraint neural network model

of opinion formation, a fact that is not supported by a person's preexisting network of opinions and beliefs will be either ignored or altered. Conversely, when people are exposed to advocacy on a specific issue, their opinions on related issues can be affected even though they resist changing their opinion on the issue itself (Wood *et al.*, 1994). Construing groups as social networks of individuals possessing elaborate conceptual networks helps explain why a person may change their mind or behavior only when they (a) receive multiple mutually-supporting messages from several neighbors (Centola, 2010), (b) are presented with a coherent story, not just isolated facts (Slovic, 1995), and (c) eventually change their overall affective response through the accumulation of many factual beliefs which may not, by themselves, influence behavior at all (Edwards, 1990).

2.2 *Choreography That Is Written for, and by, the Dance*

The notion of a correlated equilibrium features prominently in Gintis and Helbing discussion of mechanisms that allow groups to achieve stable, efficient outcomes that are in the best interests of the groups' members. Correlated equilibria are achievable if a "choreographer" is able to send a signal to the agents that inform them how to behave. The agents have an incentive to heed the advice, because it will maximize their outcome. Traffic lights are a canonical example of such a choreographer. Every driver is motivated to follow the lights' instructions to stop and go. Even though each driver would like to only go, each realizes that this will likely result in costly accidents that are avoided by following the signals. Government institutions, laws, and regulations can frequently be construed as choreographers.

An important special case of choreography is a correlating device that is created by the interacting agents – choreography that is written by the dance itself. The term "choreographer" may connote that the agents' instructions are written out ahead of time, and are external to the agents. Classic government institutions, considered in a local context, fit this description. However, many of the most scientifically interesting and effective cases of social choreography are action instructions that are written by the interactions of agents. Traffic lights that react to vehicular flow patterns, rather than following fixed timing patterns, can substantially expedite travel (Helbing *et al.*, 2007). Small teams of people in a graphical communication experiment evolve ad hoc sign systems that are more systematic and learnable than those created by people working in isolation (Fay *et al.*, 2008). When motivated to take advantage of paths laid down by predecessors, but also modify those paths to fit personal needs, people create path networks that are more optimal than the ones they create when they are trying to form optimal paths networks (Goldstone *et al.*, 2006). Originally unorganized groups will propose, vote upon, and live under rule, monitoring, and sanction systems that they construct themselves (Janssen

et al., 2008; Samuelson and Messick, 1995). In general, groups generate norms, rule systems, and communication conventions that are often more robust, efficient, and fair than those handed down from above.

The power of coordinating choreography that is written during the social dance reveals an important and neglected group dynamic. Often times a contrast is drawn between the emergent patterns of self-organized groups and groups that are driven top-down by a leader, rule system, or governmental structure (Resnick, 1994). What this rhetorical antithesis misses is that some of the activities that self-organized groups engage in are the election of leaders, the ratification of constitutions, and the formation of institutional hierarchies. Many groups that follow rules are typically self-organized, and the rule systems themselves are self-organized. The rules are the tangible products of courts, parliaments, congresses, and governments at city, county, state, country, and world levels. For example, in the absence of an existing governmental structure that effectively regulated lobster harvesting, the harvesters themselves created this structure (Acheson, 2003). Rules and their less explicit cousin, norms, are complex systems in their own right, no less so than bee hives or traffic jams. They do not exist on their own, but rather depend upon supporting structures for their continuation. Rules require legal and governmental systems to be created, changed and eliminated (Ostrom *et al.*, 2003). They require monitor systems (e.g. police) to assure that they are being followed. They require sanctioning systems (e.g. jails) to assure that discovered rule violations are punished. In this manner, groups that face scarce resources are often importantly not *simple* decentralized systems but rather decentralized systems that spontaneously create rule systems, which are themselves decentralized.

3   Smart Humans Supporting Smart Groups

While highly sympathetic with Gintis and Helbing's venture of putting a science of social organization on a proper formal foundation, my above comments aim to ground the foundations of *Homo socialis* more firmly in an understanding of *Homo sapiens*. To that end, I have argued for important, distinctly human constraints on people's behavior that cannot be deduced directly from dictates of rationality but requires appreciating the evolutionary, neural, and cognitive context in which people exist. Likewise, at the group level, there are distinctly human group processes that shape how agents that are *Homo sapiens* will interact. The richly interconnected nature of human conceptual networks entails that communication is more aptly modeled by a graph alignment process than simply copying vector elements. Finally, given that *Homo sapiens* are undeniably classifiable as "obligatory gregarious" (Cacioppo and Patrick, 2008), we strive to coordinate by self-organizing our own endogenous choreography, rather than depending on external, omniscient, and fixed choreographers.

A perennial challenge for modeling collective behavior is to achieve a good balance between incorporating a rich and elaborated model of individual behavior and achieving a constrained and elegant model of social phenomena. Given the complexities of specifying interactions among agents, the modeler might reasonably opt for a purposefully idealized account of the internal workings within each agent. This is a justifiable decision, but there is nonetheless a unique explanatory contribution achievable by models consisting of agents with richer internal lives. Unlike bees and ants, human sometimes have an appreciation for the higher-order patterns that they are creating, and this alters the resulting collective dynamics. When the individuals that comprise a collective are capable of developing a concept of the collective, then the collective's identity and goal-directedness may be intensified. When individuals can entertain thoughts like "My efforts may make other people also volunteer their time to The Cause too" and "We should develop technology that allows people to see, use, and extend what other people have contributed" then the groups formed by these individuals look increasingly like self-steering systems.

For cases in which agents are aware of the group patterns that they collectively form, there may be synergy rather than competition among different levels of organization for explanatory power. Often in science, the more "unit-like" one level of explanation is, the less unit-like higher and lower levels are. However, when the agents at one level can understand, identify with, and seek out ways to perpetuate, the higher-level group they comprise, then more coherent, "smarter" agents can create more coherent, "smarter" groups. Evidence from open source software sharing, sports fans, political movements, families, and churches indicates that people are often motivated to ensure the health and growth of their groups (Haidt *et al.*, 2008). Humans have instituted infrastructures such as peer-to-peer Internet protocols, chat rooms, Twitter, coffee houses, patent systems, professional organizations, and religious organizations to deepen the bonds that connect us to one another and promote exchange of information, beliefs, and innovations. Compelling explanations of these and other social phenomena will likely require rich models of individual human reasoning and decision making, as well as rich interactions among individuals. If constrained in the ways suggested, the added individual and collective complexity can aptly capturing uniquely human collective patterns without becoming overly flexible.

References

Acheson, J. M. 2003. *Capturing the Commons: Devising Institutions to Manage the Maine Lobster Industry*. University Press of New England: Lebanon.

Axelrod, R. 1997. "The dissemination of culture: A model with local convergence and global polarization". *Journal of Conflict Resolution*. 41: 203–226.

Barkow, J., L. Cosmides, and J. Tooby. 1992. *The adapted mind: Evolutionary psychology and the generation of culture*. New York: Oxford University Press. pp. 163–228.

Barr, D. J. 2004. "Establishing conventional communication systems: Is common knowledge necessary". *Cognitive Science*. 28: 937–962.

Cacioppo, J. T. and W. Patrick. 2008. *Loneliness: Human nature and the need for social connection*. New York: W.W.H. Norton & Co.

Camerer, C., L. Babcock, G. Loewenstein, and R. Thaler. 1997. "Labor supply of New York City cabdrivers: One day at a time". *The Quarterly Journal of Economics*. 112(2): 407–441.

Centola, D. 2010. "The spread of behavior in an online social network experiment". *Science*. 329: 1194.

Courville, A. C., N. D. Daw, and D. S. Touretzky. 2006. "Bayesian theories of conditioning in a changing world". *Trends in Cognitive Sciences*. 10: 294–300.

Edwards, K. 1990. "The interplay of affect and cognition in attitude formation and change". *Journal of Personality and Social Psychology*. 59(2): 202–216.

Fay, N., S. Garrod, and L. Roberts. 2008. "The fitness and functionality of culturally evolved communication systems". *Philosophical Transactions of the Royal Society B*. 363: 3553–3561.

Gigerenzer, G. and P. M. Todd. 1999. *Simple Heuristics That Make Us Smart*. New York: Oxford University Press.

Goldstone, R. L., A. Jones, and M. E. Roberts. 2006. "Group path formation. IEEE". *Transactions on System, Man, and Cybernetics, Part A*. 36: 611–620.

Guitart-Masip, M., Q. J. Huys, L. Fuentemilla, P. Dayan, E. Duzel, and R. J. Dolan. 2012. "Go and no-go learning in reward and punishment: Interactions between affect and effect". *Neuroimage*. 162: 154–166.

Haidt, J., P. Seder, and S. Kesebir. 2008. "Hive psychology, happiness, and public policy". *Journal of Legal Studies*. 37: S133–156.

Helbing, Dirk, J. Siegmeier, and S. Lämmer. 2007. "Self-organized network flows". *Networks and Heterogeneous Media*. 2(2): 193–210.

Janssen, M. A., R. L. Goldstone, F. Menczer, and E. Ostrom. 2008. "Effect of rule choice in dynamic interactive spatial commons". *International Journal of the Commons*. 2: 288–312.

Jones, M. and B. C. Love. 2011. "Bayesian Fundamentalism or Enlightenment? On the Explanatory Status and Theoretical Contributions of Bayesian Models of Cognition". *Behavioral and Brain Sciences*. 34: 169–231.

Kunda, Z. and P. Thagard. 1996. "Forming impressions from stereotypes, traits, and behaviors: A parallel-constraint-satisfaction theory". *Psychological Review*. 103: 284–308.

McPherson, M., L. Smith-Lovin, and J. Cook. 2001. "Birds of a feather: Homophily in social networks". *Annual Review of Sociology*. 27: 2001.

Mellers, B. A. and K. Biagini. 1994. "Similarity and choice". *Psychological Review*. 101: 505–518.

Ostrom, E., T. Dietz, and P. Stern. 2003. "The struggle to govern the common". *Science*. 302: 1907–1912.

Pope, D. G. and M. E. Schweitzer. 2011. "Is Tiger Woods loss averse? Persistent bias in the face of experience, competition, and high stakes". *American Economic Review*. 101: 129–157.

Resnick, M. 1994. *Turtles, Termites, and Traffic Jams*. Cambridge, MA: MIT Press.

Ross, B. H. 1987. "This is like that: the use of earlier problems and the separation of similarity effects". *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 13: 629–639.

Samuelson, C. D. and D. M. Messick. 1995. "When do people want to change the rules for allocating shared resources?" In: *Social dilemmas: Perspectives on individuals and groups*. Ed. by D. A. Schroeder. Westport, CT: Praeger. 143–162.

Savage, L. J. 1954. *The Foundations of Statistics*. New York: JohnWiley & Sons.

Shafir, E. and R. A. LeBoeuf. 2002. "Rationality". *Annual Review of Psychology*. 53: 491–517.

Simon, H. March 1956. "Rational Choice and the Structure of the Environment". *Psychological Review*. 63(2): 129–38.

Slovic, P. 1995. "The construction of preference". *American Psychologist*. 50(5): 364–371.

Trimmer, P. C., J. M. McNamara, A. I. Houston, and J. A. R. Marshall. 2012. "Does natural selection favour the Rescorla-Wagner rule?" *Journal of Theoretical Biology*. 302: 39–52.

Trimmer, P. C. and A. I. Houston. 2014. "An evolutionary perspective on information processing". *Topics in Cognitive Science*. 6: 321–330.

Tversky, A. 1969. "Intransitivity of preferences by Amos Tversky". *Psychological Review*. 76: 31–48.

Wood, W. L. S., J. A. Ouellette, S. Busceme, and T. Blackstone. 1994. "Minority influence: A metaanalytic review of social influence processes". *Psychological Bulletin*. 115: 323–345.