conscious determination to act in a certain way. Among all possible behaviors or concepts that constitute a population of solutions for a specific problem, how to select one or a few winning solutions amid complex agent-environment interactions to optimize adaptation is indeed subject to evolution. When evolutionary process is understood as a Darwin machine with operations of variation, selection, and heredity (as Wilson et al. understand it), what can a Darwin machine do for the science of intentional change?

To answer this question, one can look to the artificial intelligence concept of Evolutionary Algorithms (EAs), which are developed to solve optimization and search problems. EAs are composed of algorithms for reproduction, variation generation, and selection procedures, just like Darwin machines. To run EAs, one needs to specify an initial population (i.e., potential solutions to the problems in question) plus the means to select winning solutions that can be inherited with possible recombination or mutation in the next generation. A fitness function is needed in EAs to determine the fitness score, by summing up values across different factors on a common currency to index how close a given solution is to achieving the aims. For example, the best-looking face can be found by running an EA that has a variety of faces that evolve from an initial generation of population to the next by recombining features from the faces selected by humans. Although the solution (the best-looking face) can be found, the fitness function remains unknown.

Therefore, the science of intentional change that depends on evolution processes will require knowledge of the fitness functions. Indeed, Ostrom's eight design principles that were emphasized in Wilson et al. are examples of the knowledge required to formulate a fitness function, which was not obtained through any evolutionary process. If a Darwin machine cannot operate without fitness function, and the fitness function (e.g., Ostrom's principles) is identified without running the evolutionary algorithm (Darwin machine), then the science of intentional change must focus on the source and properties of the fitness function.

Furthermore, evolutionary theory at its best provides a stochastic approach to study changes, which can be either intentional or unintentional, as opposed to an analytical approach to delineate causal links (mechanistic pathway) that give rise to the changes (intervention). The stochastic and analytical approaches differ in their prediction and explanatory powers. Even when provided with sufficient initial conditions (candidate solutions and the constraints in the environment) and a fitness function, EAs as a stochastic process can provide knowledge of what solution works better than others nondeterministically (therefore with limited explanation power), and the solution cannot be known until computation of numerous iterations is completed (therefore with limited prediction power). On the contrary, an analytical process should be able to predict the outcome and explain the causal links leading to the outcome; for example, applying a hypothesis-testing experiment to test Ostrom's principles with an experiment group versus a control group.

Indeed, multiple aspects of the science of intentional change have been successfully studied in psychology and neuroscience with analytical approaches. One can conceptualize that intentional change involves goal-directed behaviors based on the incentive values of various goals and their related solutions that are encoded and maintained in domain-specific long-term memory systems. Only through analytical approaches were molecular mechanisms of synaptic transmission developed from basic invertebrate neuromuscular preparations (Swain et al. 1991) mammalian brain memory formation and change in hippocampus (Redondo & Morris 2011) and even identified techniques of planting a false memory animals (Ramirez et al. 2013). Brain imaging studies of decision making with multidomain information, a general form of intentional change, have identified the neurocircuits underlying temporal discounting of rewards (Kable & Glimcher 2007) and the common currency of incentive values integrated from social, emotional, and cognitive domains (Ho et al. 2012) – a form of fitness function. In behavioral intervention studies, key mechanisms underlying cognitive behavioral intervention to change an addicted behavior (e.g., smoking) have been identified, such as the self-referential process (Chua et al. 2011; Strecher et al. 2008) and deliberate processing (Ho & Chua 2013).

Notably, a socially inclusive stance, which can manifest in forms of altruism (Swain et al. 2012), in-group identification (Wheeler et al. 2007), and other forms demonstrated in many examples mentioned in Wilson et al., seems to play a key role in promoting positive changes at multiple levels. It may be possible to form a testable hypothesis that recognizing and respecting self and others' perspectives impartially is a central mechanism in promoting intentional behavioral and cultural change. Then, a series of analytical experiments could be carried out to test this hypothesis systematically, as opposed to be randomly conducted to create a sufficiently large population, as prescribed by a Darwin machine.

Interestingly, a hypothesis that one's social "fitness function" can be shaped to be either partial (self-defensive) or impartial (inclusive of others) is consistent with the landmark work in developmental psychology that focuses on parent-infant attachment (Bowlby 1969; 1973). After studying associations between maternal deprivation and juvenile delinquency, John Bowlby postulated his attachment theory based on an innate need to form close affect-laden bonds, primarily between mother and infant. Among studies in brain circuits underlying attachment, for example, Kim and colleagues (2010) showed that mothers who reported higher maternal care in childhood showed larger gray matter volumes and greater functional responses in some of the same brain regions implicated in appropriate responsivity to infant stimuli in human mothers (Swain & Lorberbaum 2008; Swain 2011; Swain et al. 2012; 2014). Thus, by studying the brain basis of the interactive baby-signal/parent-response (Swain et al. 2004) in the parent-infant dyad (Mayes et al. 2005), we may discover candidate brain mechanisms for a psychological fitness function in humans for intentional change.

ACKNOWLEDGMENTS

The authors are supported by grants from the National Alliance for Research on Schizophrenia and Depression (James Swain); the Klingenstein Third Generation Foundation (James Swain); NIMHD/NICHD RC2MD004767-01 and the Michigan Institute for Clinical Health Research and the National Center for Advancing Translational Sciences UL1TR000433 (James Swain and Shaun Ho); and the University of Michigan, Robert Wood Johnson Health and Society Scholar Award (James Swain and Shaun Ho).

The perils of a science of intentional change

doi:10.1017/S0140525X13003348

Muhammad Ali Khalidi

Department of Philosophy and Cognitive Science Program, York University, Toronto, ON M3J 1P3, Canada.

khalidi@yorku.ca http://www.yorku.ca/khalidi/

Abstract: The attempt to construct an applied science of social change raises certain concerns, both theoretical and ethical. The theoretical concerns relate to the feasibility of predicting human behavior with sufficient reliability to ground a science that aspires to the management of social processes. The ethical concerns relate to the moral hazards involved in the modification of human social arrangements, given the unreliability of predicting human action.

Whether intended or not, there is an illuminating ambiguity in the subtitle of the target article. The phrase "toward a science of intentional change" can be interpreted in at least two ways. First, there is the science that studies changes in human intentional or representational systems, such as language and culture. This science would investigate the ways in which the human capacity to represent the world has evolved, perhaps using insights from the evolution of other representational and communicative systems found in other species. This leads to a theoretical question about the nature of human cognition, posed through the lens of the theory of evolution. How can human intentional systems both be adapted to solve certain cognitive problems and yet be flexible enough to occupy a variety of cognitive niches?

Second, there is the science that would attempt to achieve changes in human society *intentionally*. This science would be not just explanatory but an applied science such as engineering, deliberately aiming to modify human social arrangements in order to achieve certain outcomes. This leads to a more practical question. How can the human environment be purposely altered in order to encourage cooperation and eliminate destructive behavior? There is also perhaps a third reading of the title, which straddles the first two, and concerns the science that would seek to alter human society by purposely changing our intentional or representational systems. How can human intentional systems be modified in such a way as to re-engineer our social arrangements for the sake of better outcomes? In this commentary, I will try to raise concerns about the answers that Wilson et al. give to each of the first two questions in turn, concerns that also pertain to the third question.

It is tempting to answer the first question in a glib fashion, simply by saying something about striking a balance between adaptiveness and flexibility. Indeed, the authors themselves, in using the analogy of the immune system, acknowledge that there is no reason that adaptiveness and flexibility cannot coexist. It is clearly a matter of achieving the right combination of innate responses (so as not to have to reinvent the proverbial wheel for every variant on a familiar situation) and learning (so as not to come up with an inappropriate programmed response to a situation bearing a mere superficial resemblance to a previously experienced one).

A variety of answers to this question have been given by a number of cognitive scientists working within a broadly evolutionary framework (see, e.g., Buller 2005; Carey & Spelke 1994; Cummins & Cummins 1999; Mallon & Stich 2000). There is a great deal more work to be done on this topic when it comes to specific human cognitive capacities, as the balance is likely to be different when it comes to different human abilities. However, any attempt of this kind seems incompatible with what has been called the "massive modularity hypothesis," which posits "hundreds or thousands" of cognitive modules (Tooby & Cosmides 1995), each specifically designed for a narrowly defined cognitive task. On such an evolutionary model, there is little room for a compromise between adaptation and flexibility, simply because the model emphasizes adaptive cognitive modules to the exclusion of cognitive plasticity. Wilson et al. do not seem to acknowledge that this version of evolutionary psychology is not compatible with what we know about the flexible behavior of human beings.

When it comes to the second question I have two concerns, one theoretical and the other ethical, both of which I think deserve more attention by the authors. The theoretical concern has to do with the feasibility of predicting human behavior reliably enough as to warrant constructing a science of social change. One of the lessons of the cognitive revolution is that human behavior cannot always be predicted, though it can often be successfully explained in hindsight. Not only is the prediction of human behavior not feasible when one restricts oneself only to citing environmental variables; even if one posits internal cognitive states, these states do not always enable one to predict behavior (Andrews 2012). The unreliability of prediction when it comes to complex natural systems, whether meteorological systems, biological ecosystems, or human societies, means that it is risky to intervene to produce certain desirable outcomes. The practice of cloud-seeding in meteorology is just one example of the way in which the attempt to interfere in the workings of a complex natural system can have unforeseen consequences. Similar considerations apply to biological ecosystems: It would be dicey to alter a population's environment in order to get a lineage to evolve in a certain direction. Likewise, an applied science of intentional social change is liable to be on shaky ground, as the specificities of each human community and social context are likely to render prediction quite unreliable.

Given the precariousness of predicting the effect of social interventions, the moral hazards of such attempts at social engineering loom especially large. There have no doubt been various successes when it comes, say, to modifying classroom settings in such a way as to improve learning outcomes; but generalizing from these success stories to human society at large is a risky endeavor. The advantages of enhancing human cooperative behaviors, reducing violence, and other desirable outcomes need to be weighed seriously against the ethical costs of interventions involving social control that may have unforeseen consequences. Among the principles that the authors endorse when it comes to the modification of human behavior is that of "consensus decision making," which holds that people prefer "to do what *we* want, not what *they* want." But if so, then attempts to become "wise managers" of social behaviors are unlikely to be welcome in general, and are liable to backfire.

Incorporating coordination dynamics into an evolutionarily grounded science of intentional change

doi:10.1017/S0140525X13003191

Viviane Kostrubiec[a] and J. A. Scott Kelso[b,c]
[a]PRISSMH EA EA 4561, University of Toulouse, 31062 Toulouse Cedex 9, France; [b]Center for Complex Systems and Brain Sciences, Florida Atlantic University, Boca Raton, FL 33421; [c]Intelligent Systems Research Centre, University of Ulster, Magee Campus, Derry BT487JL, Northern Ireland.
viviane.kostrubiec@univ-tlse3.fr kelso@ccs.fau.edu

Abstract: We suggest the authors' endeavor toward a science of intentional change may benefit from recent advances in informationally meaningful self-organizing dynamical systems. Coordination Dynamics, having contributed to an understanding of behavior on several time scales – adaptation, learning, and development – and on different levels of analysis, from the neural to the social, may complement, if not enhance, the authors' insights.

Inspired by the notion of a "Darwin machine," Wilson et al. aim to reconcile diametrically distinct evolutionary processes, such as innate versus adaptive and domain-general versus task specific, in a move toward a science of behavioral and cultural change. We applaud this step, though we think that the authors' rapprochement between Darwin machines and "multi-agent cooperative systems" requires some elaboration. What seems to be missing are the concepts, methods, and tools of self-organizing dynamical systems tailored specifically to the coordinated activities of living things – how they move, adapt, learn, develop, and so on (Beek et al. 1995; Calvin & Jirsa 2010; Haken et al. 1985; Kelso 1995; Kelso & Haken 1995; Schöner & Kelso 1988; Turvey & Carello 2012; Warren 2006; Zanone & Kelso 1992). Among others, Coordination Dynamics (CD) has long been inspired by the works of Howard Pattee, who understood the significance of biological coordination, particularly the complementary nature of symbolic and dynamic descriptions (Kelso & Engstrøm, 2006; Pattee & Raczaszek-Leonardi 2012).

Instead of opposing genetically fixed and adaptive processes, Coordination Dynamics sees them as dual processes evolving on different time scales. Apparently "fixed" processes are not immutable; they are stable or slowly evolving. In complex systems, processes evolving on slower time scales have been shown to constrain faster ones (Haken 1983). This opens the possibility to inquire under which conditions fast-evolving processes escape such slowly evolving (viz. inherited) constraints and reorganize