

result of modularity and plasticity, the organism can respond to new situations that recur with a novel trait, which then is able to spread throughout a population via selection for the ability required to produce the trait. In this view, evolution begins with a recurrent developmental change brought about either by a mutation or (more commonly) by environmental induction. Natural selection then consolidates the trait by modifying genes influencing the regulation of the trait.

Rather than relying primarily on mutations to structural genes within the DNA, evolution more often simply rearranges developmental regulatory genes to create novel structures, often conserving a similar program or module in a host of organisms. Rather than always viewing the organism as passively shaped by the environment, it is often the behavior of the organism that actively *creates* the environmental conditions under which morphological traits are then selected (Wcislo 1989). This is especially true in humans. In this view, the processes of genetic assimilation are set in motion by the behavior of the organism. As a consequence, behavior often takes the lead in evolution, in as much as genetic-based morphological changes often follow the path initiated by behavioral innovations.

Given the evolutionary logic that predominates today in evolutionary psychology (i.e., that recurrent environmental situations have led to modules specifically designed to respond to them), another question that remains is whether there are other types of plasticity capable of dealing adaptively with novel environments. The narrow view of evolutionary psychology criticized by the authors wrongly denies the overwhelming importance of domain-general psychological mechanisms in humans for imagining, creating, and adapting to novel environments. As the authors effectively argue, instead of pitting themselves against their caricature of the social sciences as a monolithic standard social science model, evolutionary psychologists need to embrace aspects of mainstream psychology that are critical in constructing a science designed to manage change. Domain-general mechanisms of classical conditioning, operant conditioning, and social learning enable organisms to take advantage of important contingencies that were not recurrent over evolutionary time (MacDonald 2013). In this sense, human cognitive ability—prototypically human general intelligence and problem solving—enables novel solutions to various human goals. Evolved motivational systems facilitate the evolution of any cognitive mechanism, no matter how opportunistic, flexible, or domain-general, that is able to solve the problem or achieve the goal.

Domain-general mechanisms: What they are, how they evolved, and how they interact with modular, domain-specific mechanisms to enable cohesive human groups

doi:10.1017/S0140525X1300321X

Kevin MacDonald

Department of Psychology, California State University—Long Beach, Long Beach, CA 90840.

kmacd@csulb.edu <http://www.csulb.edu/~kmacd/>

Abstract: Domain-general mechanisms are evolutionarily ancient, resulting from the evolution of affective cues signaling the attainment of evolutionary goals. Explicit processing is a particularly important set of domain-general mechanisms for constructing human groups—enabling ideologies specifying future goal states and rationalizing group aims, enabling knowledge of others' reputations essential to cooperation, understanding the rights and obligations of group membership, monitoring group members, and providing appropriate punishments to those who deviate from group aims.

The target article proposes that domain-general psychological mechanisms are essential to creating cohesive, effective human

groups. I agree, but elaborate on what domain-general mechanisms are, how they can evolve, and how they may control domain-specific mechanisms.

The target article avoids dealing with the difficulties involved in proposing domain-general psychological mechanisms that have been raised by evolutionary psychologists, relying on an analogy with the immune system that does not shed light on the evolution of domain-generalities.

Domain-generalities are evolutionarily ancient. Although there are a variety of evolved, special-purpose learning mechanisms, learning is also characterized by domain-general mechanisms that are able to achieve evolutionary goals by making novel and serendipitous associations with environmental cues. Such mechanisms are domain-general because they are able to respond adaptively to ephemeral, nonrecurrent environmental regularities that are detectable by the organism's sense organs (i.e., they are not restricted to statistical regularities over evolutionary time, as required by evolutionary psychologists; e.g., Barrett & Kurzban 2012; Tooby & Cosmides 1992; see MacDonald 2013). Such mechanisms can evolve because of the evolution of affective cues (prototypically pain and pleasure) that signal the attainment or nonattainment of evolutionary goals (e.g., satiation of hunger, the pleasure of sexual intercourse) (MacDonald 1991; 2013; MacDonald & Hershberger 2005). This then allows humans and animals to alter their behavior in response to ephemeral, nonrecurrent environmental contingencies.

Affective motivational mechanisms imply a set of adaptive problems to be solved but whose solution is underspecified. Such systems enable the evolution of any cognitive mechanism, no matter how opportunistic, flexible, or domain-general, that is able to solve the problem. Humans evolved the domain-general symbolic systems and reframing processes emphasized in the target article, as well as mechanisms underlying general intelligence, particularly the executive functions of working memory and a central executive able to direct attention and manipulate information that it receives from inputs from specialized, domain-specific mechanisms (e.g., spatial and symbolically coded information) (Chiappe & MacDonald 2005; Geary 2004).

The target article reviews several validated programs for change, but no attention is given to exactly what evolved mechanisms are involved and how domain-specific and domain-general they are. I argue that an evolutionary science of change must carefully tease out domain-specific, modular mechanisms from domain-general mechanisms, and in the case of the latter, must be clear on exactly how they are domain-general and how this promotes change.

An important contrast is between explicit and implicit processing. Implicit processing is characteristic of modules emphasized—often to the exclusion of domain-general mechanisms (e.g., Tooby & Cosmides 1992)—by evolutionary psychologists. Modules are evolutionarily ancient and their operation is fast, unconscious, automatic, and domain-specific (designed to solve specific problems). On the other hand, explicit processing is relatively recent and processing is conscious, relatively slow, effortful, and domain-general (Stanoovich 2004). Explicit and implicit processing are intimately related. Particularly important is the effortful control of implicit processing related to social and emotional behavior, including control over evolved modules designed to solve problems of survival and reproduction that were recurrent over evolutionary time (MacDonald 2008; 2009; 2010). The inputs to effortful control mechanisms include a wide range of nonrecurrent information—information resulting not from evolutionary regularities but from explicit appraisals of costs and benefits related to the contemporary world of DNA testing and video recording.

The control of evolved modules by domain-general explicit processing is a critical aspect of behavioral change. MacDonald (2008) describes the psychology and neurobiology of how effortful control can result in behavioral change in the case of aggression, ethnocentrism, emotional behavior, and drug use.

Domain-general mechanisms can indeed produce a wide variety of phenotypes for selection to act upon (the analogy of

the immune system used in the target article), but it is not just a matter of randomly producing thousands of hopeful monsters. Explicit processing involves taking in information from a variety of modular systems (e.g., perceptual information, affective desires such as sexual desire), coordinating and integrating it, and making plans of action that may involve effortful control of modular desires (Geary 2005; MacDonald 2008). But it may also include envisioning possible future states (e.g., utopian visions of the future central to Marxist ideology made possible by symbolic processing) and rationalizing group aims (MacDonald 2009; 2010), and finally deciding to act.

Explicit processing is likely unique to humans, or at least is so highly elaborated among humans that there is a qualitative difference between humans and animals. Explicit processing is required to enable at least some of the qualities essential for successful groups noted in the target article, including the following:

Group identity. Some animals have a strong sense of group identity and group boundaries (e.g., van der Dennen 1999) that is presumably modular. However, the rights and obligations of human group membership are typically explicitly articulated—they are often formally written and may be subject to judicial oversight (e.g., being in the military, a union, or a religious sect). The vast differences among human groups in rights and obligations and the fact that rights and obligations may change rapidly in the contemporary world strongly suggest that they are processed explicitly rather than exclusively via evolved mechanisms designed to track regularities of group living over evolutionary time.

Proportional costs and benefits; monitoring. Explicit processing allows people to build explicit representations of others' reputations (e.g., for shirking communal work) and the costs and benefits of actions, thus enabling human cooperation (MacDonald 2008). Whereas modular mechanisms have built-in assessments of costs and benefits (e.g., Buss & Shackelford 1997), explicit representations of others' reputations are able to track rapidly changing, novel environmental contexts (e.g., developing new technology to assess others' reputations) and are able to form explicit memories and written records of past interactions with others, thus enabling indirect reciprocity (e.g., Semmann et al. 2005; Smith 2005).

Punishment; fast and fair conflict resolution. Assessing the fairness of conflict resolution is often complex, requiring explicit knowledge of contemporary contexts incomprehensible by mechanisms attuned to regularities of the evolutionary past (e.g., assessing whether a group member is embezzling funds via computer fraud). On the basis of such explicit assessments, punishment can be finely graded, from gentle reprimands to expulsion, to ensure the viability of the group.

Intentional change, intrinsic motivations, and goal generation

doi:10.1017/S0140525X13003221

Riccardo Manzotti and Paolo Moderato

Institute of Communication, Consumption, and Behavior "G. P. Fabris," IULM University, 20142 Milano, Italy.

riccardo.manzotti@iulm.it paolo.moderato@iulm.it
www.consciousness.it

Abstract: Wilson et al. draw our attention to the problem of a science of intentional change. We stress the connection between their approach and existing paradigms for learning and goal generation that have been developed in machine learning, artificial intelligence, and psychology. These paradigms outline the structural principles of a domain-general and teleologically open agent.

Wilson et al. aim “to sketch a basic science of intentional change centered on evolution.” They warn that a critical issue is the lack of

a conceptual unification among the involved disciplines. In the same spirit, we believe it is worthwhile drawing attention to a pair of unmentioned and related research areas that have addressed the issue of change in machine learning, artificial intelligence (AI), and developmental studies—namely, intrinsic motivations (IM) and hierarchical open-ended architectures (HOA). Both areas are closely related to the evolutionary psychology (EP) versus standard social science model (SSSM) debate and endorse the SSSM-related “open-ended capacity for change” in contrast to the “elaborate innateness” of EP.

To recap, the main connections between IM/HOA and intentional change are the following. First, thanks to their design-oriented approach, IM/HOA help us to understand the key notions of intentional change and open-endedness. Second, they clarify the details of the Darwin machine whose stability is crucial for “positive intentional change.” Third, likewise to Wilson et al., both IM and HOA address the issue of learning new motivations.

The notion of intrinsic motivation aims to model doing something because it is inherently interesting or enjoyable (Deci & Ryan 1985). In developmental psychology, there has been considerable interest in IM as a way to add new goals (Barto 2013; Mirolli & Baldassarre 2013). Is intrinsically motivated behavior the result of innate metacognitive rules such as maximizing novelty, surprise, curiosity, and exploration (Berlyne 1966; Dayan & Balleine 2002; Dember & Earl 1957; Dickinson & Balleine 2002)? Or the outcome of innate motivational open-endedness (Barto et al. 2004; Manzotti & Tagliascio 2005)? Both options are currently being scrutinized (Georgeon et al. 2012; Manzotti 2010; Oudeyer & Kaplan 2007). The objective of these approaches is to apply traditional learning paradigms (respondent and operant conditioning) to explain how an intentional cognitive agent may produce a new goal. Hence, there is a critical difference between being motivated (either intrinsically or extrinsically) and being able to develop new goals. The notion of IM encourages considering intentional change from a learning standpoint. Motivational theories address learning and decision making.

In both machine learning and AI, reinforcement learning (RL) has modeled with success motivational processes. RL “addresses how predictive values can be learned, it is naturally relevant to the study of motivation” (Barto 2013, p. 19). Yet, RL seldom considered the issues of generation and evolution of reward functions (Duda et al. 2001; Sutton & Barto 1998). As a result, RL has not focused enough on the Kantorian key notions of evolution and devolution of stimulus functions (Kantor 1958). Usually, RL assumes that the reward function is given inside a module—the critic—whose nature is left unspecified. In RL, the theoretical point of departure is the study of IM. In fact, IM extend the original paradigm since intentional change may be seen as the result of teleological open-endedness (Baldassarre & Mirolli 2013; Barto et al. 2004). Furthermore, convincing experimental evidence about the neural underpinnings of IM has been put forward (Gottfried et al. 2003; O’Doherty et al. 2001). In IM, a revised and open notion of RL’s critic is the abstraction representing where and how new goals become part of the agent teleological structure. The notion of IM helps to complete a proper basic scientific foundation for an applied science of intentional change.

We have therefore reached the second area of research we want to draw attention to. To achieve IMs, an agent must be able to represent unexpected external stimuli and states of affairs. Hierarchical open-ended architectures (HOA) are a promising option (Dileep 2008; Kurzweil 2012; Manzotti et al. 2012; Sendhoff et al. 2009). HOAs stand for architectures designed to represent and to interact with a potentially unlimited hierarchy of external stimuli. In fact, if a system could not develop a hierarchically organized manifold of concepts, how could it develop new motivations? Motivational openness must be flanked by conceptual and perceptual openness. A system must be able to extend its stimulus repertoire to use it as a reward. Borrowing von Uexküll’s *Umwelt*, a process of intentional change modifies the