Chapter 4
Social Norms from the Perspective of Embodied Cognition

Chris Goldspink

Cognition is not a phenomenon that can be successfully studied while marginalizing the roles of body, world and action.

(Andy Clark, 1999)

Introduction

Like others within this title, I attempt to come to terms with the way in which human social norms emerge from, but are irreducible to, processes at the level of the individual. The particular contribution of this chapter is to suggest an emergentist account of norms which draws on the developing theory of enactive cognition. I use this to consider the characteristics needed for a system capable of simulating human-like norms in a computational environment.

Our understanding of emergence has been greatly expanded through computer simulation, but to date this has cast a light primarily on emergence within physical systems. Attempts to apply lessons from this work to social systems have largely proceeded by attempting to make simple agents more 'intelligent'. The model of intelligence used is generally that of first-generation artificial intelligence—known as cognitivism or representationalism—where the agent is equipped with some

C. Goldspink (✉)
Victoria Institute for Education, Diversity and Lifelong learning, University of Victoria, Melbourne, Australia

Graduate School of Education, Faculty of Social Science and Law, University of Bristol, Bristol, UK

Systems Centre for Learning and Leadership, University of Bristol, Bristol, UK

School of Humanities and Social Science, University of Newcastle, Newcastle, Australia
e-mail: cgoldspink@inceptlabs.com.au

M. Xenitidou and B. Edmonds (eds.), *The Complexity of Social Norms*, Computational Social Sciences, DOI 10.1007/978-3-319-05308-0_4,

limited ability to represent particular characteristics of its environment in a rudimentary computational 'mind' (Franklin, 1998). This has led to many interesting simulations, but these fall well short of allowing us to simulate many of the more complex aspects of human social interaction, including that of norm formation, maintenance and change (Sawyer, 2001, 2005).

Picking up on the statement from Andy Clark cited above (1999),[1] the central argument of the chapter is that much of what is interesting about the mechanisms of norms does not happen between passive agents nor agents with abstracted 'minds' but rather in bodies with brains operating within highly contingent environments, which they themselves contribute to generating and can also potentially change. It is towards understanding this type of phenomena that theories of embodied and enactive cognition are directed. Furthermore, recent advances in robotics have shown how these theories can be applied to practical experiments which in turn support the ongoing development of an emergentist understanding of social behaviour.

I begin with a recount of the central problems enactivist theories are directed at solving. Most fundamental of these is inadequacy of the theorization of the interplay between micro and macro social phenomena. I provide a brief restatement of the contentious issues within emergentism, connecting these to the problem of understanding norms. This discussion includes reprising levels based on the account of emergence which considers the defining features of human social and cognitive agents relevant to understanding normative behaviour. To ground the theoretical discussion I then sketch a typical normative scenario and use it to identify critical cognitive capabilities which appear to play a role in norms. I then provide an overview of the developing enactive account of these capabilities and their relevance to understanding norm emergence. I conclude by comparing alternative simulation paradigms, thereby summarizing where we are with respect to being able to simulate the mechanisms identified as relevant to norm emergence.

The Micro–Macro Problem and Its Implications for Understanding Norms

An adequate theoretical account of norms should pose plausible answers to questions such as the following:

- Where do norms come from?
- How are they maintained?
- What leads them to change or to disappear?

Existing theoretical accounts often fall well short of this in that they lack a sufficiently detailed account of the mechanisms at work. This is surprising as Sripada

[1] Clark is an advocate of embodied cognition; in this chapter I argue more for a more radical extension of the embodied standpoint—that of enaction. The enactive view has it that an agent's cognitive capabilities in effect give rise to distinct worlds—as Varela once expressed it by 'laying down a path by walking'.

and Stich (2006) state: '*No concept is invoked more often by social scientists in the explanations of human behaviour than* "*norm*"'. Indeed, the concept has been incorporated into a wide range of alternative and often competing theories of social behaviour. This lack of agreement about what norms are, and how they operate, has led to the suggestion that it is a generic concept (Gibbs, 1965) with no explanatory value.

The normative literature can be divided into at least two fundamentally distinct perspectives: the social philosophical tradition and the view from the philosophy of law.

In the social philosophical tradition (Lewis, 1969) norms are seen as a particular class of emergent pattern which spontaneously emerge in a population. From this perspective, a 'norm' is identified when a pattern of social behaviour is observed which is apparently prescriptive/proscriptive—people behave 'as if' they were following a rule. This is a bottom-up or a micro-to-macro account.

By contrast, the view offered by the philosophy of law posits norms as a source of social order. This standpoint assumes the prior existence of (powerful) social institutions which are the source of rules. When generally followed, these rules lead to the social pattern we call norms. This is a top-down or a macro-to-micro account.

Clearly the answers to the questions of where norms come from, how they are maintained and what leads them to change mounted from the perspective of these two alternatives differ markedly. However it is possible that there may be a point of synthesis which could unite these apparently opposed viewpoints. This is what an emergentist account of norms attempts, and it is linked to wider concerns about the relationship between micro and macro phenomena within the social sciences. It is worth revisiting where the debate about the micro–macro problem currently stands before attempting an extension of that debate with specific relevance to understanding norms.

Simply stated the problem is that we have no adequate way of accounting for the relationship between the (bottom up) actions of individuals and resulting social structures and the (top down) constraint those structures place on individual agency. This is not for want of trying.

The problem is central to many nineteenth- and twentieth-century social theories. Examples include Marxian dialectical materialism (Engels, 1934) built upon by, among others, Vygotsky (1962) and Leont'ev (1978); the social constructionism of Berger and Luckman (1972); Giddens' structuration theory (1984) and the recent work of critical realists (Archer, 1998; Archer, Bhaskar, Ciollier, Lawson, & Norrie, 1998; Bhaskar, 1997, 1998). These alternative theories are frequently founded on differing assumptions, extending from the essentially objectivist/rationalist approach of Coleman (1994), through the critical theories of Habermas and the radical constructivism of Luhmann (1990, 1995).

Many of these accounts conclude that structure and agency come together in *activity* or in *body-hood*—the specific psycho-motor state at the instant of enaction. Both Vygotsky and Giddens, for example, focus on action as the point of intersection between human agency and social structures, and it is implicit in Bourdieu's *habitus* also.

Essentially all of these accounts have the limitation that they fail to provide an account of the mechanisms which link the micro and macro conditions in a way which can be tested empirically or be made operational such as through multi-agent simulation. I will return to this challenge in the final section of this chapter.

In the recent past, our understanding of the mechanisms that connect micro and macro has been significantly advanced by the systems sciences, and in particular complex systems theory, as well as by developments in social simulation, evolutionary robotics and artificial life. Much of this has been done under the banner of emergentism, and this is the viewpoint that I will be developing here.

What I am essentially arguing in this chapter is that to provide an adequate account of norms we need an adequate account of *social* emergence. The key challenge is that mechanisms of emergence are likely different when we consider natural systems and social systems. We recently provided a brief account of the history of the concept of emergence and its contribution to current thinking about the interplay between micro and macro social phenomena and suggested a form of emergence particular to social phenomena which we called reflexive emergence (Goldspink & Kay, 2007; Goldspink & Kay, 2008). Reflexive emergence is associated with agents, such as humans, whose cognitive capabilities make them self-aware and strategic in their actions. This work is built on earlier contributions to understanding different orders of emergence—the argument that different cognitive capabilities support qualitatively distinct forms of emergence.

Orders of Cognition Give Rise to Orders of Emergence

Gilbert (2002), for example, has distinguished between what he called first- and second-order emergence. First-order emergence includes macro structures which arise from local interactions between agents such as particles, fluids and reflex action. This corresponds to the focus of interest to natural scientists and much of the research into complex systems which has its origins in the natural sciences. Second-order emergence is argued to arise '*where agents recognise emergent phenomena*, *such as societies*, *clubs*, *formal organizations*, *institutions*, *localities and so on where the fact that you are a member or a non-member*, *changes the rules of interaction between you and other agents*' (Gilbert, 2002: 6).

In a similar vein, Castelfranchi has distinguished what he refers to as cognitive emergence which '… occurs where agents become aware, through a given "conceptualization" of a certain "objective" pre-cognitive (unknown and non deliberated) phenomenon that is influencing their results and outcomes, and then, indirectly, their actions' (1998: 27). Castelfranchi thus conceives of a feedback path from macro pattern to micro behaviour and specifies a cognitive mechanism. He argues that this mechanism has a significant effect on emergence and gives rise to a distinct class of emergent phenomena. These ideas are more comprehensively reflected in the five orders of emergence suggested by Ellis (2006: 99–101). All these argue that the range and type of emergence possible in a system depend fundamentally on the

range and class of things agents are able to distinguish and the behaviour they are able to generate. If we are concerned to understand the emergence of norms in human societies we therefore need an adequate account of human cognition and the different aspects and facets and how they play a role in norms.

There has been considerable research directed at understanding the origins and developmental phases associated with distinctively human cognitive capabilities. Much of this has drawn on comparative neurology and sociological and psychological study of non-human animals, in particular apes. Insights are available also from the developmental psychology and neurology of the phases of development from infant to adult (see for example Reddy, 2008; Smith, 2005; Smith & Thelen, 2003).

Gardenfors (2006) identifies the following as among those needing explanation (presented in order of their apparent evolution): emotions, memory, thought and imagination, self-consciousness/theory of mind, free will and language. These are present to varying degrees in different organisms and develop at different stages in humans from infancy to adulthood. The degree of interrelatedness is not, however, straightforward. Apes for example demonstrate self-awareness and 'theory of mind' but do both without language, whereas in humans language appears to play a significant role in both.

Which of these cognitive capabilities are implicated in norms and in what way are considered briefly below and then developed throughout the rest of the chapter.

Cognitive Capabilities Implicated in Norms

Therborn has argued (2002: 868) that people follow norms for different reasons. He argues that at the more limited end of the range this involves habit or routine. Considering the cognitive capabilities implied in this, a simple capacity for remembering would be sufficient. He also argues that rational knowledge of consequences for the world may be involved. This implies agents capable of consciousness and free will or agency. Considering the implications of the previous discussion of orders of emergence, there may not, therefore, be a single emergentist account of norms, but rather a family of related ones. In other words, the overgenerality of the concept of norms may have led to a range of social behaviours being grouped together where very different generative mechanisms are implied. This is an important point from the perspective of this book as while those aspects of norms which are associated with memorised actions or unconscious patterning in decision making may lend themselves to being modelled with current approaches to social simulation, those which depend on conscious awareness do not. I say 'may' as recent research shows just how intertwined the evolutionarily older and more recent cognitive capabilities are in humans, where reflex, affect and rationality play out in complex ways in decision making (Lehrer, 2009); our physical experience of being 'in the world' informs our cognising and reason (Johnson, 1990; Lakoff & Johnson, 1999) and emotion permeates and is central to 'rational' action (Damasio, 2000, 2006).

In summary, understanding the mechanisms underpinning norms implies coming to terms with some of the most vexing aspects of social science: the problem of structure and agency or the micro–macro problem. Advances in our understanding of emergence have been driven by recent developments in complex systems as well as by the many and various examples of simulating both natural and social phenomenon. Through this work we have increasingly come to understand that different agent capabilities potentially give rise to different orders of emergence and that examples of emergence in the natural or the animal world may not help much with understanding how emergence works in human social systems—including norms.

A Narrative Account of Norms to Ground the Theoretical Discussion of the Role of Cognition

To explore the cognitive capabilities potentially involved with norms further, as well as to provide real-world grounding for the necessarily abstract discussion which follows, let us take a simple hypothetical narrative account of the operation of norms. In this simple narrative I will incorporate interactions which may play a role in the three questions with which I opened this section:

- Where do norms come from?
- How are they maintained?
- What leads them to change or to disappear?

I am a foreigner recently arrived in a new country. Walking the streets I follow the norm of my culture which is to acknowledge the presence of those (including strangers) I encounter. This pattern of engaging is for me habitual and unconscious. Let us assume that on first doing this my attempt to engage is ignored and gaze averted. My protagonist may also be acting out of habit. How do I know if this habit is based on a social norm and one that has salience to me? At this stage I do not and I may never do so, and yet I may still participate in maintaining or disrupting it.

On this first encounter, if I think about the reaction of my protagonist at all, I may conclude that he/she is simply acting out of an individual disposition—shyness perhaps. From his/her perspective I may be perceived as brash or threatening—again acting out an individual disposition rather than following a norm. The encounter may have registered unconsciously—we humans, like many animals, have evolved acuity to detecting patterns in behaviour which are contrary to our expectations (Lehrer, 2009). When I acknowledge my protagonist I may have triggered a physiological reaction. This may have included a tensing of the body, pulling away from me and the aversion of eyes as well as micro gestures of the face which suggest aversion—perhaps a flicker of shock or fear (Ekman, 1992). I may perceive these unconsciously at first through the somato-visceral system—I may experience negative affect, and I may become conscious of it as a feeling of surprise. This too may be unconsciously signalled through micro gestures although my protagonist may

not notice as he/she has already averted gaze. For both of us the reaction has 'meaning', and in the most general level this is one of threat.

If I had become conscious of the encounter I may describe it as having been rebuffed. My protagonist may report having been threatened. However, at any stage, neither the encounter nor the reaction enters conscious awareness.

Let us assume that over successive days the experience is repeated with different individuals. The negative affect experienced may lead me to unconsciously adjust my behaviour—I become less forthright or even mimic the response in order to re-establish a pattern that avoids the negative affect. If I mimic, and the response I encountered was indeed based on an individual disposition and did not reflect a social norm (there was a half-way home for paranoids nearby), then I may begin a norm as I now avert my eyes from even the non-paranoid and potentially change their behaviour. If avoidance was already a norm, then I now successfully contribute to its maintenance.

If at some point the interaction does enter my consciousness then a wider range of responses becomes possible. I may decide that the nationals of this country are antisocial and decide to 'play' with them, for example. I become even more intrusive—verbally greeting people to delight in their discomfort (rather like turning and facing people in a lift). Alternatively I might come to appreciate that this is a norm, but one particular to this place or to certain people within this place. This may help me be more tactical in the way I behave, choosing alternative ways to interact based on my appraisal of the situation and what I want to gain from my interactions with the others present. Over time this may become unconscious again—I hear a certain accent and I avert my gaze, a native of my own country, and I fully gesture acknowledgement.

In this account the degree of entanglement of cognitive abilities is illustrated. A norm may be effectively initiated or maintained without conscious awareness with signalling of conformance or non-compliance happening through subtle micro gestures out of awareness of one or more of the participating individuals but, equally, may be influenced by fully conscious processes. It may or may not involve deliberate action and consideration of own or others goals, interests or needs.

The encounter may only ever involve dyadic exchange—me and a particular protagonist. In that context neither of us can say anything about the presence or the absence of norms as we both lack the wider perspective to judge the behaviours as shared. The encounter cannot be understood without an appreciation that both of our reactions are the consequence of many past interactions which each of us has had within two different social contexts leading to the establishment of habits of action which maintain our social acceptance within that particular social context. Nevertheless, the social context determines what happens next.

If I am in a social context for which the habit is non-adaptive then the succession of disconfirming interactions and the affective impact this has on me will likely lead me to adjust my behaviour. Over time a new accommodation may be reached. If we were to go and seek out first-hand accounts of the experience of the encounter we would find very different attributions. I may describe being 'rebuffed by an antisocial person'. My protagonist may describe having been 'accosted by a foreigner'.

These accounts need not play any role in the process but they may. In making such an attribution I may decide to undertake a campaign to deliberately act so as to 'socialise the locals'. I use my agency to amplify my behaviour when I judge that it may be effective. In so doing I may generate even stronger reactions and deepen the norm among those I seek to influence—an unintended consequence.

Alternatively the rejection may lead me to give up and go and find more people whose self-narrative I share (hang out in expatriate pubs). Alternatively, within my social circles at least, I may succeed and over time this may propagate beyond my immediate interactions and change the established norm. All of which is to say that no individual needs to be aware of the 'norm' as norm, nor agree or consciously follow the deontic implied, indeed may even consciously and deliberately refuse to follow it and yet will participate in the maintenance (or potential change) of that norm. Whether or not the norm is maintained or changed will depend on the current state of the social system as a whole—including such factors as relative number of 'followers' compared to 'challengers' and how they have self-organised (distributed compared to ghettoised), the rate of introduction of individuals not accommodated to the patterns of the dominant social group. All of which is to argue, in the loosest possible way, that norms are indeed emergent. The challenge then is to more rigorously theorise what we can readily recognise.

Theorising the Mechanisms of Human Social Norms

Based on what has been presented above, the key point I wish to develop in this section is that both the prior social emergentist theory as well as a simple narrative account of norms in action imply that human social norms involve agent cognitive capabilities of various types operating at multiple levels. We need a theoretical account which can synthesise this into a framework which is compatible with an emergentist perspective and which can support practical experimentation and empirical investigation. I argue here that an enactive view is the best theory we currently have for this even though it is very much a work in progress and brings its own challenges.

In the remainder of this chapter I first provide an overview of key developments in an enactive theory of cognition and then examine the implications this has for the empirical study of norms as well as for their simulation.

Towards an enactivist account of norms

In the narrative encounter described above it is apparent that the history of past interactions in a particular social domain influences how each individual behaves instant to instant. This is consistent with the theoretical idea distilled from the many past attempts to come to terms with the interaction between micro and macro levels: structure and agency come together at the point of enaction. The fact that it is automatically reflected in all aspects of the agent (somato-visceral, affective and sensori-motor) indicates also that we are not talking just about deliberate action but

states of bodies as well as brains. For Bourdieu the habitus was the embodiment in each individual of the past as '*dispositions*, *schemas*, *forms of know-how and competence*'. For him also these were effective due to their being '*below the level of consciousness and language*, *beyond the reach of introspective scrutiny or control by the will*'. In discussing Bourdieu's account Crossley (2001: 83) states that, as a consequence, what was sought is '… *a conception of human action or practice that can account for its regularity*, *coherence*, *and order without ignoring its negotiated and strategic nature*'.

The construction of such an account has begun. It is being informed by developments in evolutionary biology, cognitive science, neurophysiology, robotics, artificial intelligence, artificial life as well as psychology, social theory and philosophy (Stewart, Gapenne, & Di Paolo, 2010). It represents an ambitious program to unite currently disparate perspectives on what it is to be an autonomous and intelligent agent. The wide scope of this enterprise presents a challenge in the context of this chapter: how best to summarise current development and link it to the theme of norms. Recent work by Barandiaran and Di Paolo et al. (Barandiaran, 2005; Barandiaran, Di Paolo & Rohde, 2009) as well as by Damasio reinforces a key theme—that the higher order abilities implicated in norms rest on the fundamentals of our living being and so we have to begin with biology, albeit emergentist biology.

The biological origin of what is meaningful and what is 'good' and 'bad' for an agent and therefore of what it 'ought' to do

In the account of norms provided by the philosophy of law discussed briefly in the opening section of this chapter, it was noted that norms imply a deontic—what 'ought' to be done. From this theoretical account the deontic is supplied by the wider society or by powerful social institutions within that society. This is in contrast with a dialectical account in that it provides no explanation of how such institutions come to take on significance or authority—to be meaningful—from the point of view of the individual, nor why individuals accede to them. The enactive account shows how this can come to be, and yet how the deontic has its origins in biological fundamentals. The account is a radical departure from how we habitually think about such things, and the following account may appear quite circuitous. It is necessary, however, to explain how some phenomenon comes to have 'meaning' for the agent.

The Biological Basis for Meaning

The transition between living and non-living has been argued, in emergentist terms, to result from self-organisation—more particularly a chain of autocatalysis resulting in the formation of self-producing autonomous (autopoietic) entities (Maturana & Varela, 1980). Recent extensions of this theory (see Barandiaran, 2005) have it that a minimal cognitive agent has a primary metabolic loop which serves to maintain its

biological viability and (at least) one other loop which links sensory surfaces with motor surfaces. This second loop adds significant plasticity within a behavioural rather than a metabolic domain (Moreno & Etxeberria, 1995: 168). Approached in this way '*minimal cognition is not so much a centralized property of the biological hardware of an organism, …*' as many theories of mind would have it, '*or a set of internally computed algorithms, …*' as assumed by first-generation artificial intelligence, including that which underpins much contemporary social simulation, '*but instead denotes an abstraction of organism environment reciprocity*' (van Duijn, Keijzer, & Franken, 2006: 165).

The most important implication of this is that the agent's classification of, and accommodation to, its environment is dynamic/homeostatic. Rocha uses the language of complex systems to elaborate on this, arguing that the order or the stability implied in the maintenance of agent viability—autopoiesis itself—is an attractor, as are the various metabolic and sensori-motor cycles involved in maintaining its relationship to a dynamic environment. States on these attractors constitute sources of input or reference to other attractors, and the current configuration of the nested attractors tells us something about the agent/environment accommodation at a particular point in time. As Rocha states it, these '*… perform environmental classifications … not all possible distinctions in some environment can be grasped by the self-organizing system: it can only classify those aspects of its environment/sensory motor/cognitive interaction which result in the maintenance of some internally stable state or attractor*' (Rocha, 1996). In other words, the range and type of environmental triggers that can be accommodated by an agent are necessarily constrained by the agent's biology, physiology *and* ontogeny and are reflected in its dynamical structure at any given point in time.

Importantly, those triggers which lead to a compensatory action can be said to be 'meaningful' from the perspective of the organism in that they have implications for its state and viability—what is 'good' for it or 'bad' for it—and may link directly to reflexes which serve to orientate it towards the 'good' (follow a nutrient gradient towards a source) and away from the 'bad' (move from an area of excessive or insufficient temperature). This is consistent with the position taken by Varela (Rudrauf, Lutz, Cosmelli, Lachaux, & Le Van Quyen, 2003; Thompson, 2004; Varela 1997) that what agents are sensitive to is determined by their own operation, not the environment. This establishes conditions of relative autonomy in that '*It is not the organism that matches the environment in a given specified way. On the contrary it is through the particular way in which the agent satisfies the homeostatic maintenance of essential variables that an adaptive environment* (*a world*) *is specified—cut out from a background of unspecific physical surroundings*' (Barandiaran, 2005).

However, this description of simple autonomy is still a long way from issues of higher cognition and norms. It is this connection I discuss next.

The idea that agent states define what is meaningful to them has direct parallel to the concept of affordances in social theory (Gibson, 1977). Particular organisms are capable of distinguishing particular stable structures in the environment, and these structures, when combined with the organism-specific capabilities, 'afford' those

organisms some opportunity. Looked at another way, material features of the world become tools, artefacts and technologies for that organism to the extent that they can extend that agent's cognitive range. In social systems also, existing social structures 'afford' opportunity and facilitate certain actions, extending the cognitive range and action potential of individuals. In part the argument here is that an agent's cognitive boundary may not be co-extensive with its physiological boundary.

With the account provided so far, we are building a layered model. The sensori-motor loop/s associated with a class of agent supports (support) distinct 'phenomenal domains'. These domains are loosely coupled to each other and to that generated by the metabolic processes associated with autopoiesis. In other words each domain has its '… *own internal coherency*' which constitutes a '*meaningful world in itself*' (Barandiaran & Moreno, 2006: 176). A social illustration of this partial autonomy or loose coupling of domains is the recent phenomena of suicide bombers. Here 'meaning' in one phenomenal domain (the belief in paradise) can trigger a behaviour which is inconsistent with the fundamental operation of the metabolic phenomena needed to maintain life. The organism is destroyed as a result of the operation of mechanisms which otherwise serve to extend and maintain its viability through inclusion within a particular social domain—in this case mutual acceptance around the norm of belief in fundamental precepts of a religion.

As we add layers of sensori-motor loops we need something to integrate then—a central nervous system. The advent in evolutionary terms of central nervous systems does not change the account of cognition provided so far in any significant way. Cognition does not now happen in brains: it is still in the agent/environment interaction. What is meaningful is not stored as a representation in memory; it is still in the dynamic maintenance of viability operating between the agent and its environment. All that has changed is that now this is facilitated by the nerve systems which link expanded points of interface with that environment. We can now say that it is the nervous system's structure—by which we mean the attractor states established within it rather than its physical architecture—that dictates which environmental perturbations can be a trigger (Mingers, 1991; Varela, Thompson, & Rosch, 1992) and therefore what will stand in a 'meaningful' relation to the agent. Just as with the amoeba, this has the implication that, as each organism traces a unique history, it specifies what is meaningful to it within its environment. Agents which trace similar or even share histories will generate similar domains of meaning (similar things in the environment as well as in the behaviour of each to the other will carry similar implications for their respective viability), while those which trace very different histories with little or no sharing may generate unique domains of meaning. We saw an example of this in a hypothetical human system with the two sets of cultural norms present in the narrative of a foreigner in a new country.

We may already talk about patterns in these resulting accommodations as 'norms' even if they are only coordinated by simple and largely innate reflex actions. Norms then are shared domains of 'meaningful' accommodations between agents. What is meaningful, and indeed the meaning conveyed, is referenced ultimately to that which is essential to maintaining the viability of the agent.

The Role of Affect and Emotion

A significant change in cognitive theory over recent times has been the growing acceptance that emotion is fundamental to cognition, including that of humans (Colombetti & Thompson, 2008; Damasio, 2006). This work suggests a complex relationship between aspects of the functioning of the body and is consistent with the intertwining of cognitive capabilities discussed earlier.

From the perspective of contemporary research, affect is argued to provide a rapid primary appraisal of presenting situations which operates in advance of conscious categorisation or assessment: affect directs the attention of the agent towards aspects of the environment or its own state which are relevant to its viability. This 'core affect' is argued on the basis of considerable empirical evidence to be a relatively un-differentiated state of arousal measured by the dimensions of valence (good/bad) and arousal (activated/deactivated) (Ryan & Deci, 2001). '*Core affect has been characterized as the constant stream of transient alterations in the organism's neurophysiological state that represents its immediate relation to the flow of changing events*'—*it is* '*a neurophysiological barometer of the individuals relation to an environment at a given point of time*' (Barrett, 2006: 31). Affective states then afford to an animal capable of supporting them what a simple sensori-motor reflex did for the amoeba, a means for classifying states in the agent/environment interaction as 'good' or 'bad'. Negative affect becomes associated neurologically with past experiences and conditions which were harmful and positive affect with ones that were beneficial to the agent.

What we call emotion is built on this core affect. Barrett (2006: 25) argues that '*The taxonomic structure of self-reported experiences of emotion does not support the view that anger*, *sadness*, *fear and so on*, *are qualitatively distinct and experientially primitive*'. This is to say that emotions are not biologically primitive like core affect but arise from a process of conceptual or perceptual categorization on top of or in relation to an affective response. What we commonly refer to as emotion (or *feelings* in Damasio's schema) are labels for a set of experiences represented in consciousness.

The position taken by many of these more recent emotion theorists is that these 'conceptualisations' are not abstracted from sensori-motor events and stored in propositional form, but exist as 'simulations' ('as if' states for Damasio) of the sensori-motor states that occurred with previous instances of a similar experience. When we see a picture of something frightening, we do not recover an abstract concept of fear to label the picture, rather we re-experience fear at a somato-visceral level, albeit in a low key way—the concept of fear is embodied.

Affect then represents a whole body state response to environmental triggers. Even when triggered by memories of events, they elicit a response that involves arousal and action—affecting the viscera, endocrine and motor systems in concert. This then presents no problems from the account of cognition being presented—affect and emotion merely form part of the continuum which may support qualitatively distinct domains of interaction and hence mechanisms for norm formation

and maintenance. A possible role for emotion in norm forming, maintenance and change was indicated in the narrative of a foreigner in a new country provided above. In this account, affect played a significant role in the immediate flow of events. Subsequent reflection on the emotional experience could have effected subsequent interactions to the extent that embarrassment was experienced in suffering a rebuttal to a social exchange, in endeavouring to avoid the experience in future. Each time the event is remembered, by the above account of the operation of emotion, the associated affective state will be re-experienced, serving to deepen the experience and aversion and perhaps the resolve to behave in some different way. For me to resolve to do something, however, I need to have conscious self-awareness and perhaps a sense of identity about who 'I' am as well as agency. These aspects too then can (but may not) play a role in the creation, maintenance or destruction of change of norms.

The 'Viability Set' Provides a Foundation for What Is 'Good' or 'Bad' at the Level of the Individual

Building on what has been argued so far we can say that 'cognitive agents' define what is meaningful to them in the environment—they place value on the stream of events they experience as they experience them. For living agents, at the most fundamental level, what stands as meaningful are those aspects of the environment essential to maintaining their viability as a living entity. For the most basic organisms (such as cells) their ability to adapt and remain viable in response to a change in their environment is quite narrow—specified by chemical and mechanical parameters fundamental to their metabolic pathways. However, once an organism has developed a sensori-motor loop in addition to the purely self-maintaining metabolic mechanisms it has the capacity to adapt behaviourally to its environment. Simple amoeba can, for example, propel themselves along a nutrient gradient using simple mechanisms such as flagella. This response capability is bounded: sensitive to only a limited range of changes with a limited set of response capabilities (flagella only work in fluids of limited range of viscosity). We can therefore conceive of a 'viability set': the range of events to which the organism can adapt and maintain its viability (Di Paolo, 2005). The basic sensori-motor mechanisms of reflex through to affective pre-appraisal (as just discussed) and then reaction through to conscious decision making and language (which I will consider in the next sections) all serve to expand the viability set.

As an agent begins to interact with others the response it engenders will be perceived as affirming or as a threat. With human agents this will most likely initially take place based on affective pre-appraisal (Damasio, 2000) as discussed above but may also involve more conscious deliberation as included in the account of the foreigner in a new country. The evaluation will lead to a behavioural response which is adaptive—based on the agent's history of interaction in particular social

domains in its history. The implicit goal will be to increase the chance of remaining viable in the current domain. It does not, however, do this in splendid isolation.

The discussion so far has focused on individual cognitive capacity—a very micro focused orientation. However the moment the effect of coupling between micro-agents is appreciated the pathway and mechanism by which social structures bootstrap from these interactions and back-propagate to constrain them become apparent. This next step is therefore key to an emergentist account of norms.

As agents interact with one another, their viability sets intersect. We could model this in the same way Kauffman (1993, 2000) has for fitness landscapes. The resulting 'viability landscapes' are coupled—the adaptations made by one agent change the landscape of the others with which it is interacting. In Froese and Di Paolo's terms '*… since the regulation of the interaction of one agent changes not only its own coupling but also that of the other agent, it follows that the agents can enable and constrain each others sense-making*' (2009: 9).

At the most general level norms can be conceptualised as relatively stable patterns on this coupled viability landscape—agents converge on viable accommodations of each other's accommodations. They form from the complex product of the response capability of the agents—affective, unconscious as well as rational conscious, but where each agent influences others through its behaviour (which may include subtle gestural aspects as well as the more overt). In this sense, norms are possible as agents make mutual accommodations to one another so as to maintain their viability within a particular social domain. However, if we are to make sense of behaviour such as the ignoring of norms then the mechanism described so far, that of viability maintenance, is insufficient. We need another idea: that of agency.

Agency

In considering a the role of agency in norm formation and maintenance we are concerned to distinguish between purely adaptive accommodations to environmental change, including that generated by the action of other agents in the coupled viability landscape, and agents which modulate their own behaviour so as to shape the trajectory of their interaction with the environment. Barandiaran et al. discuss it as follows:

> Environmental conditions are good or bad for the continuation of the system. This normative dimension is not arbitrarily imposed on the system by a designer or external agent that monitors the functioning of the system and judges according to her interests. It is the very organisation of the system which defines a set of constraints and boundary conditions under which it can survive. … This precariousness implies that whatever the organism is doing … there is something that it ought to do; not for an external observer but for itself, for the continuation of its very existence (2009: 375).

This quote illustrates why we had to go back to discuss fundaments of biology in order to understand norms. What is 'meaningful' to an organism and hence the base

for all subsequent accommodations and judgements about what is 'good' or 'bad' for it propagate from its biological viability. It will now be apparent to some readers that this presents a problem from the point of view of simulation. I will have more to say on that in the final section.

Barandiaran (2005) has argued that loose coupling between the metabolic domain and the sensori-motor domain allows an organism to exploit the rapid response times of the neural system in order to expand its viability set. Within the emerging field of neurodynamics (Cosmelli, Lachaux, & Thompson, 2007; Kelso, 1995; Rocha, 1996; Thompson & Varela, 2001; van Gelder, 1998) it is argued that this 'plasticity' is in large part due to the nervous system operating on a system of complex attractors, yielding quasi-stable emergent states. By these accounts it is the asymmetry between the combination of all possible configurations the agents biology and ontogeny afford it, and the (more limited) range of responses needed to maintain immediate regulation in a given environment, that gives rise to what we call 'agency': '*The higher the agent's capacity for adaptively guided self-restructuring* (*plasticity*) *the higher its behavioural adaptive autonomy and hence its agency*' (Barandiaran, 2005).

Peter Hejl (1993) also locates agency in 'cerebral overcapacity'. He notes that this conveys advantages and disadvantages. The advantage is in furnishing support for a wide range of possible responses and hence 'requisite variety' (Ashby, 1974). The disadvantage is that high plasticity contributes to the contingent nature of agent–agent and agent–environment interactions and thus instability. The advantages only hold sway over the disadvantages to the extent that the variability can be channelled or constrained in short time frames. As Hejl notes, '*The only* '*solution*' *to this problem … seems to be society*' (1993: 229). For Hejl then quasi-stable structures that arise through social interaction (such as norms) serve to reduce social complexity in the short term while keeping open a much wider range of possible adaptations and accommodations—through the change of existing norms or emergence of new ones appropriate to alternative contexts.

In short then, the 'surplus capacity' made available by an advanced neural system explains how a living system can come to have the potential to remain viable in changeable environments, but not how it exploits that potential. There is still a perspective missing. This is the perspective of how an agent can come to be conscious of its capacity for choice and use that choice in strategic and tactical ways.

As humans we can choose to ignore a norm—perhaps rationalising that it does not apply to us. As was illustrated in the narrative of a foreigner in a new country, we can also choose to maintain or to try and change a norm or begin a new one. All of this implies the use of agency in a strategic way—a purposeful striving. This only becomes possible if the agent can distinguish 'self' from 'other' and can act to advance its own or others' interests and intentions in a deliberate, selective and conscious way. Consciousness needs to be explained as a higher order cognitive function with significant potential implications for normative mechanisms, and, in the context of this chapter, it also needs to be placed within the wider enactive account being developed.

An Enactive Account of Consciousness, Self-Awareness and Identity

Thompson has argued that the sense of 'self' has its primary (pre-conscious) origin in an organism's capacity to use its own self-constitutive processes as a source of reference. Here the sense of self as a 'totality' or a stable whole is strongly associated with its biological autonomy (Thompson, 2005) and hence has its origin in fundamental biological processes such as those already discussed above. Similarly, Damasio (2000) distinguishes between proto-self, core consciousness and extended consciousness with each being developed on the former. The proto-self relies on the nervous system's capacity to use relatively stable internal states as a reference point. Damasio groups them under the heading of the 'internal milieu'. However, he also argues that this sense may be combined with proprioception and kinaesthetic mappings which identify the positioning of muscles and limbs in combination with the sense of 'fine touch' from the epidermis and thus use the body's interaction with the environment as a reference point for a sense of self as separate from environment. Either may provide a source that is relatively stable which can be used as a foundation for a distinct sense of 'self'. Importantly these sources are always available while the organism is alive and interacting in its environment. This is argued to provide a basis for consciousness to the extent that the organism can notice that actions have 'self' as an origin (are 'owned' by self) and that through such actions 'self' exerts agency on the environment.

This sense of self is further differentiated. Damasio uses the terms core consciousness and extended consciousness, while others refer to it as minimal self and narrative self. The former is associated with the agent's '*consciousness of oneself as an immediate subject of experience*, *unextended in time*' (Gallagher, 2000: 15) and the latter '*A more or less coherent self* (*or self-image*) *that is constituted with a past and a future in the various stories that we and others tell about ourselves*'. It is only this last form of 'self' or identity construction that requires language. As Menary argues '*First there are the experiences of a living body and then we turn those experiences into a narrative*' (Menary, 2008). Through narrative, however, a variety of alternative stories about self may be elaborated.

Narrative represents a means by which some socially located stability, such as 'norms', capture, propagate and give persistence to the unfolding dynamics of social interaction. They constrain individual action through their shaping of identity, without the individual having permanently to give up the full potential of the wider space of possibilities. They serve to smooth the otherwise turbulent 'push' and 'pull' of the accommodations individuals need to make to remain viable in different social domains. And, in so doing, they may stabilise the wider dynamics that results from structural coupling: forming another layer of constraint on the coupled viability landscapes already discussed.

At the level of the individual, the current state of their ontogeny is reflected in their narrative account of themselves at that time. That narration also reflects their location of themselves in a shared or a social history. Ochs and Capps state, '*The power*

to interface self and society renders narrative a medium of socialization par excellence' (Ochs & Capps, 1996: 31). Returning to the hypothetical story, when our foreigner jokes about the 'locals' with fellow expatriates he or she perhaps construct a narrative which locates 'us' as 'together against another'. The narrative reinforces the shared valuing of one set of norms (the ones shared with those present) and deprecates those of the 'other'. These exchanges, while undertaken in language, invoke emotive responses which become attached to the labels of 'us' and 'them' and will be regenerated in subsequent encounters, influencing behaviour.

While the proto-self is grounded in affect, the narrative self implies language and I have not yet accommodated language into the unfolding account of the relationship between cognitive capability and social norms.

Cultural Tools and Language

Ross (2007: 718) says of language, '*similar public linguistic representations cue similar behavioural responses in individuals with similar learning histories*, *as a result of conventional associations established by those histories*'. Thus, as Maturana has argued, a shared history of interaction leads to the establishment of a consensual domain (Maturana, 1978; Maturana & Varela, 1980). However, contrary to the conventional assumption this does not imply that language constructs a one-to-one denotative representation with objects or phenomena in the real world (Kravchenko, 2007). Rather language represents a particularly flexible form of behaviour by which one agent may attempt to influence another or others. If we concentrate on how people attempt to influence each other in language we will notice that it is not only, or even so much, the content of what is said that matters but more the manner of the saying and hearing. Linguistic interaction cannot be decoupled from the behaviour of talking and listening. Individuals are orientated to one another, and the reciprocal behaviours associated with a stream of 'communication' present each participant with many cues, some more subtle than others, about the others' orientation and intent with respect to the 'self' as well as their apparent purpose and what they intend for and from you. For Cowley and Macdorman (2006) talk is better approached as '*… a multimodal way of toying with persons*'.

If language is more indexical rather than symbolic, utterances and words, as well as the tone and style by which they are delivered, rely on some level of experiential grounding—a learned association gained through repeated exposure within a shared social domain. In this context a word is indexical of a gestalt of sensori-motor experience initially associated with particular contexts but which may become more generalised through increasingly diverse association. Lakoff and Johnson (1999) argue that this is so profound that many of our fundamental concepts 'borrow' from our experience in physical space. So when I say that to perform a task is 'below me' I use a physical metaphor (my experience in the world of things which are above and below one another) to tag an affect which cues me to my place within a social status norm within the society to which I belong. It is this fundamental

characteristic of language which supports the wide range of ways in which we use it—as metaphor, to invoke paradox, to hint at associations, ironically, to provoke, to stimulate and to frustrate, making it a powerful tool for influencing the behaviour of others and hence shaping the formation and transformation of norms.

Language too then plays a fundamental role in modulating shared viability sets.

Part 3: Implications for Simulation of Social Norms

I have now set out the key elements of an enactive account of aspects of human cognitive capability which may play a role in the initiation, maintenance and change of norms. The above account integrates existing psychological, sociological and cognitive theories of human action. It is also consistent with an emergentist approach applied to social systems. The account is far from complete however. It has drawn on recent developments in all of the contributing disciplines, including evolutionary robotics and artificial life and also some aspects of social simulation. It also has the potential to guide these more empirical sciences of sociality. It is to this that I wish to turn in concluding this chapter. In this final section I unpack the implications for how we might approach the simulation of norms mindful of what the enactive view has suggested as key mechanisms.

Our insights into and ability to theorise about the micro–macro interplay at the core of social phenomena have been greatly advanced by the possibility for computer simulation. Much of what we now understand about the behaviour of emergent systems has resulted from simulations. Theory and modelling have therefore moved hand in glove, and we might reasonably expect this to continue. The account set out above has a number of implications for how we choose to model and how we compare the model to the world.

Alternative Paradigms

There are a number of alternative ways in which simulation is being used to advance our insights, particularly into human social system behaviour, including that of norms. The three primary (paradigmatic) approaches are cognitivism, embodied cognition (Clark, 1998; Shapiro, 2011) and, more recently, enactivism (Stewart et al., 2010). Each represents a logical progression in that each is argued to address limitations and problems of those which have come before.

The message from the story recounted earlier is that the regularity which characterises norms is a product of contingent, situational specific striving of the participating agents, acting through a variety of motives, interpreting their situation differently and pursuing a mix of individual and collective goals with each influencing the other on a coupled viability landscape. If this is accepted then we can use this to examine which of the alternative paradigms may support simulation methods best equipped to deepen our understanding of different aspects of normative behaviour.

In this final section, therefore, I want to evaluate where we are in the development of alternative approaches to understanding and modelling norm-capable agents and how we might best advance theory and experimentation directed at better understanding how norms arise, are maintained, change and disappear.

Paradigms of Mind

As previously discussed, simulation including of simple 'dumb' agents or particles has given us a great deal of insight into mechanisms of emergence and will no doubt continue to do so. However, in order to extend this learning into the mechanisms of social behaviour we have needed to make assumptions about the nature of social agents. More particularly we have had to find ways to construct agents which reflect the cognitive capabilities associated with human social behaviour.

The science of artificial intelligence as well as of multi-agent systems has built upon cybernetics which itself drew on information theory and theories of universal computation to posit intelligence as a form of computation. The resulting paradigm has been labelled representationalism or cognitivism. In their book *The Embodied Mind*, Varela et al. (1992), argue that '*The central intuition behind cognitivism is that intelligence—human intelligence included—so resembles computation in its essential characteristics that cognition can actually be defined as computations of symbolic representations*' (Varela et al., 1992: 40).

Cognitivism therefore constructs a duality. The environment is experienced as a 'fact' external to the agent and is acted upon directly but is also conceived and symbolically represented in the 'mind'. This approach gave rise to two well-known and fundamental problems now referred to as the framing problem and the grounding problem. Both of these are relevant to understanding and simulating norm emergence.

As has been discussed, people unconsciously or consciously follow norms on some occasions and not on others. Within social theory this is usually explained by norms being context specific and by agents weighing the cost of adhering to norms against other alternative goals or drives. It is in relation to this aspect of norm following that the framing problem is an issue. Systems based on cognitivism cannot deal with dynamic and subtle variations in context. They require the designer to anticipate the range of environmental conditions the agent will encounter and design in a set of decision rules to support this.

The grounding problem is also invoked by the challenge of norms. As we have seen norms carry some implicit 'meaning' (or functional significance) for the agent. Cognitivism is based on the use of symbolic representation—some salient characteristic is represented in the mind as a symbol. In cognitivist systems the meaning of the symbol must be provided from outside or coded into the system.

In cognitivist approaches then, the frames the agent can use to judge the salience of a norm as well as any functional significance of that norm must be provided from outside and therefore are not under the control of the system. Such agents can generate

emergent behaviour but not in a manner analogous to the way humans appear to in relation to norms. To simulate the emergence of norms what is significant and meaningful must be allowed to change as a result of the interaction between agents as individual (micro) choices shape social (macro) consequences on coupled viability landscapes.

We must conclude, therefore, that it is difficult to do justice to the emergent nature of norms using cognitivist approaches.

The framing problem presented major problems for even simple robots attempting to navigate their way in relatively fixed environments. The solution was an approach to cognition which allowed agents to learn and evolve their parameters to deal with environments instead of attempting to program in the necessary contingency table. The resulting *connectionist* models (Brooks, 1991) invoke no symbols, thereby avoiding some aspects of the grounding problem. Rather than manipulating symbols which 'stand for something' in the agent's environment, meaning is embodied in fine-grained structure and pattern throughout the network. Connectionist approaches can derive pattern and meaning by mapping a referent situation in many different (and context dependent) ways. Meaning in connectionist models is embodied by the overall state of the system in its context. It is implicit in the overall '*performance in some domain*'. Connectionism led to a major leap forward in robotics. However Dreyfus has identified a residual challenge that confronts both cognitivism and connectionist approaches. This is how to '*directly pick up significance and improve our sensitivity to relevance*' … since this ability '*depends on our responding to what is significant for us*' given the current contextual background (Dreyfus, 2007: 30).

Linking this to thinking about norms, a connectionist model could converge on a pattern within its environment and develop an effective accommodation to it. If that pattern changes in a novel way—one not anticipated by the system designers—a connectionist system may still be able to accommodate that change within limits. What it still cannot do is make a judgement as to how the new pattern is in its interests and nor can it initiate strategies to attempt to influence that new pattern to turn it to its advantage, except to the extent that some representation or implicit design aspect framed from outside (i.e. through the hand and mind of the system designer) specifies where the boundaries of self interest are—it does not have and cannot develop the agency which, as has been discussed above, may play a role in normative action.

Connectionist approaches therefore support experimentation into aspects of norms where there is some scope for habits to form and adjust in relation to changing contexts, including the behaviour of other agents. However, as Froese et al. argue '*as long as there is no meaningful perspective from the point of view of the artificial agent, which would allow it to appropriately pick up relevance according to its situation in an autonomous manner, such a system cannot escape the notorious* "*frame problem*"' (Froese & Ziemke, 2007: 8).

This brings us to the argument for enactive approaches to artificial systems. Enactivism solves the framing and grounding problem in the manner already

described earlier in this chapter—the self-producing nature of the agents provides them with a fundamental goal—maintenance of self.

From what has now been considered throughout this chapter it is now possible to identify the minimum set of requirements for an approach to simulation capable of reproducing dynamics which are reasonable analogues of social norms in human social systems. To simulate norms we need agents:

- Who's state at any given time is a product of its interactions with other agents.
- Have a low-level goal (this presupposes a minimal condition which they seek to maintain such as their viability) and against which their actions and the actions of others can be evaluated.
- The range of emergent norms that will arise in such a system is influenced by the substantive constitutive nature of the agents—and hence the range of states they are capable of recognizing (perceiving), evaluating and responding to.
- This must involve more than a capacity to simply couple to the environment but a capacity to break symmetry (Barandiaran et al., 2009).

Towards an Emergentist Simulation of Norms

We are still a considerable way from being able to build systems with these capabilities. On the positive side we are getting closer to being able to specify what it will take.

1. We need to be able to model an agent as an operationally closed (autonomous) entity. This does not have to be at the level of biological process—the agent does not need to produce itself in a material sense; rather as Froese and Ziemke (2007) state the artificial system must be capable of generating its own systemic identity at some level of description. The level of description will be relative to our purpose for performing the simulation and which aspect of social (including normative) functioning we are attempting to explore.
2. An artificial system must have the capacity to actively regulate its ongoing sensori-motor interaction in relation to a viability constraint linked to the maintenance of its identity.
3. Agents need to be able to be assembled (or to self-assemble) onto coupled fitness landscapes where the fitness function is linked to the underlying viability set.

A recipe for working towards such an artificial system has been sketched by Morse et al. (Morse, Lowe, & Ziemke, 2008) and many simple practical experiments conducted in this direction (see for example Di Paolo n.d.; Di Paolo & Lizuka, 2007; Froese & Di Paolo, 2008; Montebelli, Lowe & Ziemke, 2009).

Patterns which emerge in the relationship between such agents would qualify as norms in that they would be genuinely emergent. They would represent quasi-stable patterns which satisfy the viability requirements of the participating agents. It will,

however, likely be very difficult for a human observer to understand in what way these patterns are 'meaningful' (i.e. functional with respect to the agents and/or the system they comprise) other than in the highly abstract context of the artificial world. It will be difficult to steer the emergence of such patterns towards particular experimental ends as well as to interpret what they suggest by way of outcome. De Loor et al. (n.d.) suggest that one approach to this problem may be to include a real human as a participating agent. No doubt we will discover more of how this might be possible as we progress towards the development of simulation platforms with these types of characteristics. If the slow rate of progress within AI is a guide, this will not be a rapid process.

Conclusion

Norm-following agents are characterised by being able to generate alternative response through their interaction with one another which serve to maintain each as a viable entity within particular social domains. Norms represent quasi-stable patterns or attractors generated by the process of mutual accommodation on coupled viability landscapes. These accommodations arise through multiple modes of interaction from reflex, through affectively modulated interactions through to tactical and strategic positioning made possible by different levels of cognitive capability extending in humans to agency and identity and the scope for language as a particularly flexible mode for mutual influence.

To date attempts to study norms within social science have failed due to the micro–macro divide—the inability by contemporary social science to provide an adequate account of the dialectic between macro social structures and individual dispositions and action. While systems thinking, particularly that associated with complex systems, has significantly advanced our understanding of mechanism of emergence and therefore served to illuminate mechanisms associated with this dialectical interpenetration of levels, it has done relatively little to date to contribute to our understanding of the particular way in which this may operate in human social systems. Nevertheless these advances, as well as rudimentary social simulations, made possible through cognitivist and more recent connectionist approaches to robotics, have proceeded hand in glove to help advance our understanding of human social system dynamics which extend well beyond what was achieved in the past several hundred years within social theory and philosophy alone.

Key and often fresh insights into what a next generation of social simulation platforms might look like can be drawn from recent advances in cognitive biology and evolutionary robotics. Unlike much social simulation, which has tended to stay with representational approaches, these other fields have taken seriously the questions posed by the entanglement of cognitive capabilities as well as the known problems with cognitivist approaches, in particular the framing and the grounding problems. This work has helped us to identify what the characteristics of a system need to be to support investigation of norms.

References

Archer, M. (1998). Realism in the social sciences. In M. Archer, R. Bhaskar, A. Collier, T. Lawson, & A. Norrie (Eds.), *Critical realism: Essential readings*. London: Routledge.

Archer, M., Bhaskar, R., Ciollier, A., Lawson, T., & Norrie, A. (1998). *Critical realism: Essential readings*. London: Routledge.

Ashby, W. R. (1974). Self-regulation and requisite variety. In F. E. Emery (Ed.), *Systems thinking*. Great Britain: Penguin.

Barandiaran, X. (2005). *Behavioral adaptive autonomy. A milestone on the ALife roue to AI?* San Sebastian, Spain: Department of Logic and Philosophy of Science University of the Basque Country.

Barandiaran, X., Di Paolo, E. A., & Rohde, M. (2009). Defining agency: Individuality, normativity, asymmetry and spacio-temporality in action. *Adaptive Behavior, 17*(5), 367–386.

Barandiaran, X., & Moreno, A. (2006). On what makes certain dynamical systems cognitive: A minimally cognitive organization program. *Adaptive Behavior, 14*(2), 171–185.

Barrett, L. F. (2006). Solving the emotion paradox: Categorization and the experience of emotion. *Personality and Social Psychology Review, 10*(1), 20–46.

Berger, P. L., & Luckman, T. (1972). *The social construction of reality*. Harmondsworth: Penguin.

Bhaskar, R. (1997). *A realist theory of science*. London: Verso.

Bhaskar, R. (1998). *The possibility of naturalism*. London: Routledge.

Brooks, R. A. (1991). Intelligence without representation. *Intelligence Without Reason, 47*, 569–595.

Castelfranchi, C. (1998). Simulating with cognitive agents: The importance of cognitive emergence. In J. S. Sichman, R. Conte, & N. Gilbert (Eds.), *Multi-agent systems and agent based simulation*. Berlin: Springer.

Clark, A. (1998). *Being there: Putting brain, body and world together again*. Cambridge, MA: The MIT Press.

Clark, A. (1999). An embodied cognitive science? *Trends in Cognitive Sciences, 3*(9), 345–351.

Coleman, J. S. (1994). *Foundations of social theory*. Cambridge, MA: Belknap.

Colombetti, G., & Thompson, E. (2008). The feeling body: An enactive approach to emotion. In W. F. Overton, U. Muller, & J. L. Newman (Eds.), *Developmental perspectives on embodiment and consciousness*. London: Lawrence Erlbaum Associates.

Cosmelli, D., Lachaux, J.-P., & Thompson, E. (2007). Neurodynamics of consciousness. In P. D. Zelazo, M. Moscovitch, & E. Thompson (Eds.), *The Cambridge handbook of consciousness*. Cambridge: Cambridge University Press.

Cowley, S. J., & Macdorman, K. F. (2006). What baboons, babies and tetris players tell us about interaction: A biosocial view of norm-based social learning. *Connection Science, 18*(4), 363–378.

Crossley, N. (2001). The phenomenological habitus and its construction. *Theory and Society, 30*(1), 81–120.

Damasio, A. (2000). *The feeling of what happens: Body, emotion and the making of consciousness*. London: Vintage Books.

Damasio, A. (2006). *Descartes' error*. London: Vintage Books.

De Loor, P., Manac'h, K., Fronville, A., & Tisseau, J. (2008). Requirements for an enactive machine: Ontogenesis, interaction and human in the loop. Plouzane, France: Universite, Europeenne de Bretagne.

Di Paolo, E. A. (2005). Autopoiesis, adaptivity, teleology, agency. *Phenomenology and the Cognitive Sciences, 4*(4), 429–452.

Di Paolo, E. A. (n.d.). Organismically-inspired robotics: Homeostatic adaptation and teleology beyond the closed sensorimotor loop. Brighton: School of Cognitive and Computing Sciences, University of Sussex.

Di Paolo, E. A., & Lizuka, H. (2007). How (not) to model autonomous behaviour. *Biosystems, 91*(2), 409–423.

Dreyfus, H. L. (2007). Why Heideggerian AI failed and how fixing it would require making it more Heideggerian. *Philosophical Psychology, 20*(2), 247–268.

Ekman, P. (1992). Facial expressions of emotion: New findings, new questions. *Psychological Science, 3*(1), 34–38.
Ellis, G. F. R. (2006). On the nature of emergent reality. In P. Clayton & P. Davies (Eds.), *The re-emergence of emergence: The emergentist hypothesis from science to religion*. Oxford: Oxford University Press.
Engels, F. (1934). *Dialectics of nature*. Moscow: Progress Publishers.
Franklin, S. (1998). *Artificial minds*. London: MIT Press.
Froese, T. & Di Paolo, E. A. (2008). Stability of coordination requires mutuality of interaction in a model of embodied agents. Paper presented at the Procedures of the 10th international conference on simulation of adaptive behavior, Berlin, Germany.
Froese, T., & Di Paolo, E. A. (2009). Sociality and the life-mind continuity thesis. *Phenomenology and the Cognitive Sciences, 8*(4), 439–463.
Froese, T., & Ziemke, T. (2007). *Enactive artificial intelligence*. Brighton: University of Sussex.
Gallagher, S. (2000). Philosophical conceptions of the self: Implications for cognitive science. *Trends in Cognitive Sciences, 4*(1), 14–21.
Gardenfors, P. (2006). *How homo became sapiens: On the evolution of thinking*. Oxford: Oxford University Press.
Gibbs, J. P. (1965). Norms: The problem of definition and classification. *American Journal of Sociology, 60*, 8.
Gibson, J. (1977). The theory of affordances. In R. Shaw & J. Bransford (Eds.), *Perceiving, acting, and knowing*. New York, NY: Lawrence Erlbaum.
Giddens, A. (1984). *The constitution of society: Outline of the theory of structuration*. Berkeley, CA: University of California Press.
Gilbert, N. (2002). Varieties of emergence. Paper presented at the social agents: Ecology, exchange, and evolution conference, Chicago.
Goldspink, C., & Kay, R. (2007). *Social emergence: Distinguishing reflexive and non-reflexive modes, AAAI fall symposium: Emergent agents and socialites: Social and organizational aspects of intelligence, Washington, DC*. Menlo Park, CA: AAAI.
Goldspink, C. & Kay, R. (2008). Agent cognitive capabilities and orders of emergence: Critical thresholds relevant to the simulation of social behaviours, AISB convention, communication, interaction and social intelligence, April 1–4, 2008, University of Aberdeen.
Hejl, P. (1993). Culture as a network of socially constructed realities. In A. Rigney & D. Fokkema (Eds.), *Cultural participation: Trends since the middle ages* (pp. 227–250). Amsterdam: John Benjamins Publishing Company.
Johnson, M. (1990). *The body in the mind: The bodily basis of meaning, imagination and reason*. Chicago, IL: The University of Chicago Press.
Kauffman, S. A. (1993). *The origins of order: Self organization and selection in evolution*. New York, NY: Oxford University Press.
Kauffman, S. (2000). *Investigations*. New York, NY: Oxford.
Kelso, J. A. S. (1995). *Dynamic patterns: The self-organization of brain and behavior*. Cambridge, MA: MIT Press.
Kravchenko, A. V. (2007). Essential properties of language, or, why language is not a code. *Language Sciences, 29*, 650–671.
Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to Western thought*. New York, NY: Basic Books.
Lehrer, J. (2009). *The decisive moment: How the brain makes up its mind*. Melbourne, VIC: Text Publishing.
Leont'ev, A. N. (1978). *Activity, consciousness and personality*. Englewood Cliffs, NJ: Prentice Hall.
Lewis, D. (1969). *Convention: A philosophical study*. Cambridge MA: Harvard University Press.
Luhmann, N. (1990). *Essays on self reference*. New York, NY: Columbia University Press.
Luhmann, N. (1995). *Social systems*. Stanford, CA: Stanford University Press.
Maturana, H. (1978). Biology of language: The epistemology of reality. In G. A. Miller & E. Lenneberg (Eds.), *Psychology and biology of language and thought: Essays in honor of Eric Lenneberg*. New York, NY: Academic.

Maturana, H., & Varela, F. (1980). *Autopoiesis and cognition: The realization of the living*. Boston, MA: Reidel.
Menary, R. (2008). Embodied narratives. *Journal of Consciousness Studies, 15*(6), 63–84.
Mingers, J. (1991). The cognitive theories of Maturana and Varela. *Systems Practice, 4*(4), 319–338.
Montebelli, A., Lowe, R., & Ziemke, T. (2009). *The cognitive body: From dynamic modulation to anticipation. Anticipatory behavior in adaptive learning systems* (Vol. 5499). Berlin: Springer.
Moreno, A., & Etxeberria, A. (1995). *Agency in natural and artificial systems*. San Sebastian, Spain: Department of Logic and Philosophy of Science University of the Basque Country.
Morse, A. F., Lowe, R., & Ziemke, T. (2008). Towards an enactive cognitive architecture. Paper presented at the international conference on cognitive systems, April 2–4, 2008, Karlsruhe, Germany.
Ochs, E., & Capps, L. (1996). Narrating the self. *Annual Review of Anthropology, 25*, 19–43.
Reddy, V. (2008). *How infants know minds*. London: Harvard University Press.
Rocha, L. M. (1996). Eigenbehavior and symbols. *Systems Research, 13*(3), 371–384.
Ross, D. (2007). H. sapiens as ecologically special: What does language contribute? *Language Sciences, 29*(5), 710–731.
Rudrauf, D., Lutz, A., Cosmelli, D., Lachaux, J.-P., & Le Van Quyen, M. (2003). From autopoiesis to neurophenomenology: Francisco Varela's exploration of the biophysics of being. *Biological Research, 36*, 27–65.
Ryan, R. M., & Deci, E. L. (2001). On happiness and human potentials: A review of research on hedonic and eudaimonic well-being. *Annual Review of Psychology, 52*, 141–166.
Sawyer, K. R. (2001). Emergence in sociology: Contemporary philosophy of mind and some implications for sociology theory. *American Journal of Sociology, 107*(3), 551–585.
Sawyer, K. R. (2005). *Social emergence: Societies as complex systems*. Cambridge: Cambridge University Press.
Shapiro, L. A. (2011). *Embodied cognition*. Bristol, PA: Taylor & Francis.
Smith, L. B. (2005). Cognition as a dynamic system: Principles from embodiment. *Developmental Review, 25*, 278–298.
Smith, L. B., & Thelen, E. (2003). Development as a dynamic system. *Trends in Cognitive Sciences, 7*(8), 333–348.
Sripada, C. S., & Stich, S. (2006). A framework for the psychology of norms. In P. Carruthers (Ed.), *The innate mind: Culture and cognition* (Vol. 2). New York, NY: Oxford Univer-sity Press.
Stewart, J., Gapenne, O., & Di Paolo, E. A. (2010). *Enaction: Towards a new paradigm for cognitive science*. Cambridge, MA: MIT Press.
Therborn, G. (2002). Back to norms! On the scope and dynamics of norms and normative action. *Current Sociology, 50*(6), 17.
Thompson, E. (2004). Life and mind: From autopoiesis to neurophenomenology, a tribute to Francisco Varela. *Phenomenology and the Cognitive Sciences, 3*, 381–398.
Thompson, E. (2005). Sensorimotor subjectivity and the enactive approach to experience. *Phenomenology and the Cognitive Sciences, 4*(4), 407–427.
Thompson, E., & Varela, F. J. (2001). Radical embodiment: Neural dynamics and consciousness. *Trends in Cognitive Sciences, 5*(10), 418–425.
van Duijn, M., Keijzer, F., & Franken, D. (2006). Principles of minimal cognition: Casting cognition as sensorimotor coordination. *Adaptive Behavior, 14*(2), 157–170.
van Gelder, T. (1998). The dynamical hypothesis in cognitive science. *Behavioral and Brain Sciences, 21*, 615–665.
Varela, F. (1997). Patterns of life: Intertwining identity and cognition. *Brain and Cognition, 34*, 72–87.
Varela, F., Thompson, E., & Rosch, E. (1992). *The embodied mind*. Cambridge, MA: MIT Press.
Vygotsky, L. S. (1962). *Thought and language*. Cambridge, MA: MIT Press.