Chapter 3
Robot Sociality: Genuine or Simulation?

Mark H. Bickhard

Abstract It is clear that people can interact with programs and robots in ways that appear to be, and can seem to participants to be, social. Asking the question of whether or not such interactions could be genuinely social requires examining the nature of sociality and further examining what requirements are involved for the participants in such interactions to co-constitutively engage in genuine social realities – to constitute genuine social agents. I will attempt to address both issues. A further question is "Why ask the question?" Isn't "sociality" like a program in that simulating the running of a program *is* the running of a program – so sufficiently simulated sociality is genuine sociality? What more could be relevant and why? As I will explain, there are at least two sorts of answers to the question of why the difference between genuine performance and simulation matters: (1) to better understand the ontology of sociality and thereby its potentialities and ways in which "merely" simulated sociality might fall short, especially of the developmental and historistic potentialities of sociality, and (2) to better understand the issues of ethics surrounding interactions among and between humans and robots.

Keywords Social ontology • Persons • Robot sociality • Interactivism

3.1 Introduction

This paper has three sequential issues: (1) The ontology of sociality, (2) Normative stakes in sociality, and (3) Robots and sociality. In the first discussion, I outline a model of social ontology, in preparation for the question of whether and in what sense robots might be capable of genuine participation in such an ontology. The second discussion focuses on an aspect of sociality that is, on the one hand, essential to genuine sociality, and, on the other hand, problematic for robots. The third

M.H. Bickhard (✉)
Department of Philosophy, Lehigh University, 15 University Drive, 18015, Bethlehem, PA, USA

Department of Psychology, Lehigh University, 17 Memorial Drive East, 18015, Bethlehem, PA, USA
e-mail: mhb0@lehigh.edu; www.bickhard.ws

R. Hakli, J. Seibt (eds.), *Sociality and Normativity for Robots*, Studies in the Philosophy of Sociality 9, DOI 10.1007/978-3-319-53133-5_3

section elaborates this problematic characteristic, and addresses whether it might be overcome, and some of its implications for the possibilities of robots as moral agents.

These three focal discussions are developed within a framework of a broader model that addresses the emergence of normative function, representation, cognition and language, social ontology, and the ontology of personhood. I will begin with an introduction to that framework.

3.2 Normative Emergence

3.2.1 Process and Emergence

A core point of this discussion is that human sociality involves forms of normativity that robots, in their current form, cannot capture, though they might be able to simulate that normativity up to a point. To explicate this point requires developing a model of normativity, especially normativity of representation and of motivation, and, in turn, of the emergence of such forms of normativity.

Emergence, however, is itself a problematic notion, so I begin with an argument that emergence is in fact not possible within a classical substance-metaphysical framework that we have inherited from the Greeks, but that it is possible, even ubiquitous, if we return "back to Heraclitus" and work within a process metaphysics. A process metaphysics not only, so I argue, makes genuine metaphysical emergence possible, but also overcomes some serious, if not fatal, problems with more standard substance or particle-based metaphysics.

I begin with Parmenides. Parmenides argued that change cannot occur (R. J. Campbell, 1992; Gill, 1989; Guthrie, 1965). This is at times interpreted as a response to Heraclitus' process model (Graham, 2006), though, whether or not that is correct, it certainly stands opposed to Heraclitus. The argument is that, when considering fundamental metaphysical reality, for something fundamental *A* to change into something also fundamental *B* would require that *A* disappear into nothingness and *B* appear – emerge – out of nothingness. "Nothingness" cannot exist: you cannot point to it (R. J. Campbell, 1992); "nothingness" here does not mean something like a vacuum (though that was a considered interpretation at times), but, rather, something itself fundamental, perhaps like a contemporary person trying to think about the "nothing" that "exists" outside of the universe. This argument has had an enormous influence on Western thought, in significant part because Plato and Aristotle gave it great credence.

Democritus and Empedocles both attempted to show how the appearance of change could occur without violating the Parmenidean argument. The key is to recognize that change does not occur at a fundamental level of metaphysics, but, rather, at a more "surface" level of appearances. For Empedocles, the fundamental metaphysics consisted of four substances, earth, air, fire, and water, and these did

not change. What could change, however, were mixtures of these. The combinations of earth, air, fire, and water could vary even though the four substances themselves did not change. Democritus, with respect to this issue, had a similar model: the configurations of atoms could change, but the atoms themselves did not.

Aristotle had a sophisticated elaboration of such a model. For Aristotle, earth, air, fire, and water could change into each other, but they were not fundamental. He maintained the constraint that change could only occur with respect to a (more) fundamental unchanging substrate for change, either in the form of Prime Matter that never changes, or in a still more sophisticated version that turned on unchanging substrates of hot or cold or wet or dry (Gill, 1989).

In any case, this is the basic constraint that has come down to us as a dominant background assumption: change occurs only with an unchanging substrate. In contemporary views, this level of substrate is usually assumed to consist of basic particles.

Adopting such a framework has several consequences. Here are three that are of importance for the following discussion:

1. Stasis, lack of change, is the default condition. Change requires explanation.
2. Emergence is impossible. You cannot, for example, get a new substance out of Empedoclean earth, air, fire and water. Emergence, in fact, is precisely what these were intended to preclude – they (claim to) account for change *without* having to posit emergence.
3. A "natural" world is limned consisting of substances or particles, "causal" relations among them, factual relations among them (e.g., configurations of particles). This realm is split from a realm of normativity and intentionality – a roughly "mental" realm that does not seem to fit into, or together with, the "natural" realm.

Given this split, there are only three general possibilities: (1) Assume dual metaphysical realms, such as Aristotle's matter and form or the Cartesian dual substances. (2) Try to account for the "natural" world in terms of the "mental" world, as in some versions of idealism. (3) Try to account for the "mental" world in terms of the "natural" world. This latter is the dominant position today, in which naturalism is assumed, and is assumed to consist of some sort of physicalism.

We would, perhaps, like to account for the normative, intentional world as *emergent* within the natural world, but emergence is precluded by precisely the substance/particle assumptions that create the split in the first place. Adopting a process metaphysics, however, overturns all three of these consequences:

1. Change becomes the default, and constancy or stability requires explanation.
2. Emergence becomes possible in organizations of processes.[1]

[1]For discussions of Hume's and Kim's arguments that preclude emergence, explicit for Kim and implicit for Hume (see Bickhard, 2009b, R. J. Campbell & Bickhard, 2011 and below). I argue that both presuppose a particle or entity metaphysics.

3. With barriers to emergence removed, it becomes possible to explore models of the emergence of normativity and intentionality – the possibility of (re-) integrating the two realms.

3.2.2 Why Process?

From the perspective of a substance or entity or particle metaphysics, organization is not a legitimate locus of causality. Organization is not a "thing", and only things can have basic causality. So any apparent causality that is manifest by some new organization of, say, particles, is just the particle causality working its way out starting from whatever configuration or organization they started in. Organization is just "initial conditions" and has no possible causality itself. But if emergence is supposed to account for new causal powers in new organization, then a particle framework makes this impossible by assumption.

Processes, however, are intrinsically organized, and have whatever consequences they have for the rest of the world – whatever causal influences they have – partly and necessarily in terms of their organization. Processes are intrinsically organized, and that organization is not just stage setting for their ongoing flow. Within a process metaphysics, (causal) influence on the world is necessarily dependent on (among other things) organization; organization cannot be excluded as a legitimate locus for causality without removing causality from the universe.

New or different process organizations, thus, can yield new or different consequences for the world. They can yield emergent forms of influences on the world.

So, process enables emergence. Are there other reasons to accept a process metaphysics? Yes.

First, in a pure particle metaphysics, the probability of one dimensionless particle hitting another are zero. In such a universe, nothing would ever happen.

A common view today is that the world is constituted as point particles that interact with each other not by contact, but via fields. This is strictly not correct (see below), but it already forces the acceptance of organization as potential locus of "causal" power: fields are organized, and have whatever influence on the world that they do have in part in terms of those organizations.

Second, our best contemporary physics tells us that there are no particles (Aitchison, 1985; Aitchison & Hey, 1989; Bickhard, 2003; Brown & Harre, 1988; Cao, 1999; Davies, 1984; Halvorson & Clifton, 2002; Huggett, 2000; Kuhlmann, Lyre & Wayne, 2002; Sciama, 1991; Weinberg, 1977, 1995). Everything is quantum fields, and what remains of the notion of particle is only that quantum field interactions are intrinsically quantized. This quantization is of either whole integer or half integer multiples, just as is the waveforms in a guitar string: if held at two points, the guitar string vibrates with whole integer numbers of waves, while, if held at only one point, it oscillates with half integer multiples (consider, for example, the half integer wavelengths in a rope that is waved from one end). This is quantization, and there are no guitar sound particles; nor are there physical particles.

It might be pointed out that our best theories in physics might nevertheless not be correct. Yes, and, in fact, we have good reason to think that they are false as they currently stand. Nevertheless, they do support a process metaphysics, not a particle metaphysics, and, perhaps more importantly, we have empirical evidence that makes it essentially impossible to ever return to a particle framework, no matter what changes are made in the theories. For example, in a vacuum, according to quantum field theory, there is constant oscillatory activity, with wave-like processes emerging and disappearing – a kind of froth or foam of activity. If two very smooth metal plates are held (very) close to each other, they act similarly to the holding of a guitar string at two places: only whole wavelength activity can occur between the plates. In this manner, the activity between the plates is reduced relative to the activity outside of the plates, and this induces a pressure pushing the plates toward each other. This effect has been detected and measured, completely in accord with theoretical predictions, and there are no particles in this story (Sciama, 1991). In effect, the localities and isolations of a particle based assumption have been shown to be false, and, in that sense, there is no possibility of return to a particle metaphysics.

So, a process metaphysics enables emergence, potentially enabling models of the emergence of normativity and intentionality (and related phenomena), and it is the only kind of metaphysics that is viable given contemporary physics, both theoretical and empirical.

3.2.3 Maybe Causal Emergence, but Normative Emergence?

Shifting away from substance models to process models is a basic theme in the history of science. We have given up models of fire as the substance phlogiston, heat as caloric, magnetism as magnetic fluid, life as vital fluid, all in favor of process models of combustion, kinetic energy, field models, and various kinds of self-organizing, self-maintaining, processes. Normativity and intentionality, however, seem to pose special problems. They are not only the core of what was split off by the ancient delimitation of a substance world, they are also the kind of phenomena that Hume supposedly showed could not be derived from fact: no norms from facts; no ought from is.

Hume did not actually develop an argument, and there have been multiple attempts to fill in what that argument might be (Schurz, 1997). Nevertheless, his conclusion has had powerful influence. Hume's "argument" consists of the phrase "what seems altogether inconceivable":

> I cannot forbear adding to these reasonings an observation, which may, perhaps, be found of some importance. In every system of morality, which I have hitherto met with, I have always remark'd, that the author proceeds for some time in the ordinary way of reasoning, and establishes the being of a God, or makes observations concerning human affairs; when of a sudden I am surpriz'd to find, that instead of the usual copulations of propositions, *is*, and *is not*, I meet with no proposition that is not connected with an *ought*, or an *ought not*.

> This change is imperceptible; but is, however, of the last consequence. For as this *ought*, or *ought not*, expresses some new relation or affirmation, 'tis necessary that it shou'd be observ'd and explain'd; and at the same time that a reason should be given, for **what seems altogether inconceivable**, how this new relation can be a deduction from others, which are entirely different from it. But as authors do not commonly use this precaution, I shall presume to recommend it to the readers; and am persuaded, that this small attention wou'd subvert all the vulgar systems of morality, and let us see, that the distinction of vice and virtue is not founded merely on the relations of objects, nor is perceiv'd by reason. (Hume, 1978, Book III. Part I. Section I. 469–470, emphasis supplied)

I offer here an interpretation that makes Hume's point valid, though, so I will argue, unsound. Consider a valid conclusion based on premises that, by assumption, contain only factual terms. If that conclusion contains new terms, not present in the premises, how could they have been introduced? By definitions. The only form of definition that Hume knew about, a form descending from Aristotle, was what might be called abbreviatory definition, in which a term or phrase is defined as being equivalent to a defining term or phrase. What is crucial for current purposes about such definitions is that they can be "back translated" through. That is, for all instances of a defined term or phrase, the defining term or phrase can be substituted. Perhaps there will still be new terms, but these too can be eliminated in favor of the defining terms or phrases, and so on until all remaining terms in the conclusion are those that were in the original premises. These, by assumption, are all factual, so no non-factual – e.g., normative – terms can be validly introduced in the conclusions.

Notice, that although directed at deriving "ought" from "is", the general form of this argument is that nothing beyond new grammatical (or logical) configurations of premise terms can be validly deduced in conclusions. This is reminiscent of a basic particle point from Democritus: only new configurations, not new particles. In this generality, Hume's "argument" is an argument against any kind of emergence.

But it is unsound: it involves a false premise. The false premise is the assumption that all legitimate forms of definition permit back translation. Hume knew of no other kinds, but, since Hilbert, we do know of other kinds of definitions today. In particular, implicit definition (as distinguished from explicit definition, which *does* permit back translation) does not permit back translation. Implicit definition was introduced by Hilbert (and other authors around the turn of the 20th century) with respect to formal axiom systems. Consider an axiom of geometry: two points determine a line. In classical views, the terms in this axiom would have to themselves be further defined in terms of, among other things, what a point and a line are. These ultimately would be taken to denote things in the real world, or in some kind of formal world. For Hilbert, their interpretations were left free. Any interpretations that satisfied the relations within and among the axioms were legitimate; any such interpretation (using Tarskian language) constitutes a *model* of the axioms. So, for example, two points determine a line in the sense of the line that runs between the points, but also two lines determine a point – reversing the interpretation – in the sense of determining the point of intersection of the lines (so long as points at infinity are accepted for parallel lines): the terms "point" and "line" can be interpreted either way.

A set of axioms *implicitly defines* its class of models. What is important for my current purposes is that such an implicit definition is, on the one hand, completely legitimate, and forms part of the basis for contemporary model theory (e.g., Chang & Keisler, 1990), but, on the other hand, cannot be back translated through. If the above construction of what Hume had in mind is correct,[2] then his conclusion is unsound because it makes a false assumption about the nature of definition.[3]

If we shift to a process metaphysics, then, emergence becomes possible, and if we recognize that Hume's argument is based on a false assumption about the nature of definition (as well as mirroring a basic particle metaphysics consequence), then perhaps normative emergence becomes possible.

3.3 The Emergence of Normative Function

Normativity and intentionality pose particular problems for attempts to model their emergence. Physics in general provides multifold distinctions among and within a large class of variables – location, velocity, charge, etc. – but there is no sense in which there is an asymmetry among those distinctions that the normative asymmetry between, for example, functional and dysfunctional or true and false or good and bad might be based. There is, however, one basic exception to this point: thermodynamics. And it is a thermodynamic distinction that forms the ground for the model of normative functional emergence.

Within a process metaphysics, change is the default, but there are many ways in which *organizations* of process can manifest persistences and stabilities. One major class of such stabilities is constituted by what can be called *energy well* stabilities. These are organizations of processes that remain persistent unless and until some above threshold energy impinges on them. An atom is a canonical example: it is a furious process among quarks, gluons, and electrons, but it remains stable in its basic organization for potentially cosmic time periods.

Another class of stabilities is at the opposite extreme of thermodynamics. An atom, or a rock, can be isolated from the world, and it will remain stable. It will go to thermodynamic equilibrium and stay there. A candle flame, on the other hand, cannot be so isolated: if isolated it will go to equilibrium and *cease to exist*. A candle flame is – ontologically – far from thermodynamic equilibrium, and it cannot be isolated without disrupting the maintenance of that far-from-equilibrium condition. Energy well stabilities persist at equilibrium, and do not require maintenance; far-from-equilibrium processes can persist, but only if they

[2]If it is argued that it is not correct, then the burden is on whoever offers such a claim to make good on what Hume's argument could possibly have been (Schurz, 1997). The abbreviatory definition interpretation fits what Hume knew about definition, and does support his "argument" as valid.

[3]There are also non-formal senses of implicit definition (e.g. Bickhard, 2009b; Hale & Wright, 2000).

are maintained at far-from-equilibrium conditions. Far-from-equilibrium processes require maintenance in order to persist, and energy well stabilities do not. This is the basic asymmetry upon which normativity is based: maintenance of the far from equilibrium conditions essential for a far-from-equilibrium process is contributory – *functional* – for the persistent existence of that process.

A candle flame also illustrates another important property: *self*-maintenance. A candle flame contributes to its own stability; it tends to maintain the conditions for its own persistence: It maintains above combustion threshold temperatures, induces convection which brings in oxygen and removes waste products, and so on. Far-from-thermodynamic-equilibrium conditions are essential for all self-*organizing* processes, but candle flames illustrate this further property of self-*maintenance*.

A candle flame is self-maintaining, but if, for example, it runs out of wax, it has no way of detecting that nor of doing anything about it if it were to detect it. A bacterium, however, illustrates a further crucial property: that of *recursive* self-maintenance (Bickhard, 2009b). If a bacterium is swimming up a sugar gradient, it will tend to keep swimming – toward more sugar. If it detects going down a sugar gradient – away from sugar – it will tend to tumble for a moment and then take up swimming again. In these circumstances, swimming is contributory, functional, for the bacterium so long as it is swimming toward more sugar. But it is dysfunctional if it is swimming toward less sugar. The bacterium is capable of detecting something about its circumstances in the world and adopting activities that are functional within those circumstances (D. T. Campbell, 1974). It is able to maintain its condition of being self maintenant across changing conditions: it maintains self-maintenance – recursive self-maintenance.

3.3.1 Etiological Models of Normative Function

The dominant models of normative function today are etiological models (Millikan, 1984, 1993). The central idea of these models is that some organ in the body *has a function* if it has the right kind of etiology, generally an evolutionary or a learning etiology. I will focus on Millikan's model, in which the required history for having a function is an evolutionary history: ancestral organs have been selected during evolution for having the consequences that constitute the function of the kind of organ involved – e.g., the kidney has the function of filtering blood because ancestral kidneys were selected for filtering blood.[4]

There are several questions and problems with such models. I will focus on a couple of them. One central problem is that having a function is *constituted* in

[4]Note that, insofar as such models were to succeed in modeling normative function, they would violate Hume's argument. I have argued that Hume's argument is unsound, but the point that either something is wrong with Hume's argument or else these models cannot be correct is generally overlooked or ignored.

having the right (evolutionary) history. But differing etiologies, evolutionary or otherwise, can yield the same organization of processes. That is, differing histories can yield dynamically, causally, identical systems. The problem that this produces is illustrated by a thought experiment that Millikan discusses: suppose that the molecules in the air come together for form a lion in the corner of the room that, by assumption, is molecule by molecule identical to a lion in the zoo. The lion in the zoo has organs with the right evolutionary histories to have functions, but the lion in the corner, even though identical in a dynamic sense to the lion in the zoo, has no relevant history, and, thus, its organs have no functions. In this case, we have two lions that are causally identical, but one has functions and the other does not: function, in this model, is causally epiphenomenal.

A related problem is that the evolutionary history has to have a beginning, and issues become problematic when such beginnings are considered. Consider the first time that some organ, or part or location, in an organism is selected for having some consequence. After some number of generations of this (how many is not clear) descendent organs will have that consequence as a function. It might be considered that such selection is strictly causal, with no normative properties at all, till some sufficient number of generations has passed, at which point normativity emerges. But being selected for something involves that something, that consequence, being positive for the organism, contributing to the organism's existence, being *functional for* the organism in the sense of the thermodynamics based model outline above. Whether or not this model of function is accepted, it is difficult to avoid the conclusion that the basic selections, no matter how many or few the number of generations involved, is already a normative kind of phenomenon. It is already normative in the sense of being useful to the organism, though, perhaps, not yet in the sense of being designed for serving that function (Christensen & Bickhard, 2002).

One basic difference between etiological models and the thermodynamic interactive model is that etiological models focus on what it is to have a function, which lends itself to a kind of evolutionary (or learning) design sense of what is involved, while the interactive model focuses on what it is to serve a function, to be useful, which is most primitively based on thermodynamic considerations. If having a function is central, then serving a function must be addressed, and that is generally addressed with the notion that something serves a function when it "succeeds" in having the consequence that is the function that it has. So, an organ serves a function when it does what its ancestral organs were selected to do. This makes it difficult, for example, to account for "accidental", not "designed", usefulnesses, such as the function of leg muscles in aiding blood flow on long airplane flights.

If *serving* a function is central, then the notion of *having* a function must be explicated. In general, the idea here is that various activities in an organism *presuppose* activities of other parts of the organism in the sense that the given activities can succeed in serving their functions so long as the presupposed activities and consequences are available. Such functional presuppositions that such and such a consequence will be realized in such and such a location – e.g., blood filtered in kidney locations – is what constitutes *having* a function in this model. In this model,

having a function is derivative from serving a function. There is no difficulty in this model in accounting for non-designed serving a function – serving a function that no organ *has*.[5]

3.4 Representation and Cognition

The notion of functional presupposition is crucial to a next major emergence: that of *representational truth value*. In general, when a living being engages in interactions with its environment, those interactions will be functional under some circumstances, but not functional under other circumstances. If an interaction is engaged in the "right" circumstances, it will succeed in serving whatever function(s) it might be associated with. Engaging in such an interaction, then, functionally presupposes that such supporting conditions hold, in the sense that engaging in such an interaction would not be functional if those supporting conditions did not hold, and the interaction in that case would not be true to its functional nature (R. J. Campbell, 2011): for the bacterium, swimming presupposes that its orientation is toward more sugar. This is an interaction level, environmental version of the functional presuppositions that hold among various organs within an organism that constitute those organs having functions.

Crucially, if an interaction is engaged in and supporting circumstances hold, then the presupposition of those circumstances is correct – is *true*. The interaction is true to its functional nature if its functional presuppositions are true (R. J. Campbell, 2011). If the bacterium is oriented up a sugar gradient, then the presupposition of the activity of swimming holds, and swimming will in fact be functional; it could be *dys*functional if the bacterium were swimming up a saccharin gradient, in which case the presupposition would be false. Here we have the emergence of representational truth value – the basic normativity of representation – within the framework of the functionality of various kinds of interactions between the system and its environment.

This is a primitive version of representation, but the central model has resources to account for more complex forms of representation. Consider a frog that might have a possibility of flicking its tongue in one direction to eat a fly, another direction to eat a different fly, and downward to eat a worm. The frog must have some functional way in which to indicate these possibilities, and some way of selecting among them. For my current purposes, the central point is that indications of interactive possibilities can branch into multiple possibilities.

They can also conditionally iterate: the frog, to continue that example, might also have the possibility of shifting to its left, at which point a different worm might become accessible – some interactions can create the possibilities for other interactions. Such branching and iterating indications of interactive potentialities

[5]For further discussion and comparison, see (Christensen & Bickhard, 2002; Bickhard, 2009b).

can, in more complex organisms, form vast webs of indications, and it is within such webs that more complex forms of representation can emerge.

Consider a child's toy wooden block. There are many interactions possible with such a block, including multiple visual scans and manipulations, as well as throwing it, leaving it on the floor, and so on. There is a sub-web of such possibilities that has certain special properties: every scan or manipulation is reachable from every other scan or manipulation, with appropriate intermediary manipulations, and this internally completely reachable sub-web is itself invariant under various other things that can be done or could happen, such as throwing, leaving on the floor, etc. Some things, however will destroy the opportunities of the web, such as crushing or burning the block. In this manner, representations of, for example, small objects can be constructed. This is essentially a borrowing of Piaget's model of the representation of small objects into the interactive model.[6] Such borrowing is possible because both Piaget's and this model are action based.[7]

A complex agent, thus, must characterize its environment in terms of what interactions with that environment are possible (Bickhard, 2009b).[8] An agent must have such a functional characterization in order to be able to select among those possibilities in accordance with criteria such as goals – e.g., a goal of eating if the internal state is one of hunger (e.g., low blood sugar).[9] Such possibilities can be complex, and are constantly changing due to environmental processes as well as activities of the agent itself, so organizations – webs – of such *anticipations* of possibilities must be constantly maintained and updated. I call such complex webs the organism's *situation knowledge* – knowledge of what potential interactions are available in the current environment – and the maintenance and updating of situation knowledge is called *apperception*.[10]

[6]As in Piaget (1954). There is a substantial literature that claims that such knowledge is innate, and, thus, not really constructed. But there is also a large family of reasons to consider such claims to be based on incorrect theory and systemically faulty methodology (Allen & Bickhard, 2013).

[7]For discussion of these resources in more detail, including resources for representations of abstractions such as the number three, see, e.g., Bickhard (2009b).

[8]It should be noted that this model of representation is a model of functional emergence, not of consciousness or conscious representation. Neither the bacterium nor the frog, for example, have reflective consciousness. The overall model, however, does have some strong implications regarding consciousness, and entails: (1) that consciousness is not the unitary phenomenon that is commonly assumed, (2) that *some* properties of consciousness are in fact emergent in fairly simple organisms, and (3) that much of the mystery of how to naturalize consciousness evaporates when it is realized that mystery is a product of underlying false assumptions being made in the literature (Bickhard, 2005).

[9]Processes of selecting among trajectories of indicated interaction possibilities constitute the domain of *motivation*.

[10]These anticipatory webs have some similarities with Gibson's notion of affordance, though Gibson's affordances cannot have the kind of structure and organization mentioned above. For discussion of Gibson, see Bickhard & Richie (1983).

3.4.1 Encoding Models of Representation

Extant models of representation are versions, sometimes ingenious versions, of what can be called *encodingism*: the assumption that all representation is some form of encoding. There are multiple versions of encoding models, but they have in common an assumption that there is some special correspondence between representation and represented that constitutes a representational relationship. This special relationship is variously modeled as being causal, informational, structural, etiological, and so on. One further common characteristic is that such models assume that the special representational relationship is some version of looking backwards down the input stream to try to "represent" where that input stream is coming from–spectator models, in the language of Dewey (1960; Tiles, 1990).

Encodings certainly exist, witness Morse code, in which, for example, "· · ·" encodes "s". Such codes can be useful: dots (and dashes) can be sent over telegraph wires while characters, such as "s", cannot. For another example, most of us would starve if computer codes were suddenly abolished: they are very useful in contemporary society. Furthermore, codes do not have to be conventional like Morse code or computer codes: Consider a neutrino count in a gold mine in North Dakota that encodes properties of fusion processes in the sun.

What all genuine codes have in common, however, is that in order for the code to exist, both ends of the encoding relationship must be known, and the relationship between the ends must be known: both "· · ·" and "s", both the neutrino count and the fusion properties, and the relationships between them. Codes change the form of representation, and give us access to doing things with those encoding representations, such as inferring fusion properties or transmitting a message over a wire, that would otherwise be not possible or much more difficult. Crucially, codes require that the encoded end of the relationship *be already known*. Encountering "· · ·" per se cannot give you any sense of "s", nor neutrino counts per se anything about fusion properties. Encodings cannot constitute representation of anything not already known about.

3.4.2 Problems with Encodingism

There is, in fact, a large family of fatal problems with encoding models, some known for millennia, some of more recent provenance. But encoding assumptions still dominate because there has not been an alternative on offer. The interactive model outlined above belongs in the tradition of pragmatism, which is historically relatively recent, and avoids the problems of spectator models in virtue of being future oriented, not past oriented.

One problem with encoding models is that they cannot account for the emergence of representation. You have to already have a representation of "s" in order to define "· · ·" in terms of it. But what constitutes the representation of "s", and

where did it come from? Encoding definitions can iterate, some things defined in terms of others, which are, in turn, defined in terms of others, but this must halt at some level of basic representations out of which all others must be defined. Fodor argues that this base must be innate (Fodor, 1981). But if representation cannot emerge, then evolution cannot generate representation any more than can learning or development. Conversely, if evolution *can* generate emergent representation, then Fodor offers no argument about why learning and development could not do the same. But representation did not exist 13 billions years ago and it does now: it has to have emerged. In fact, Fodor at times acknowledges something like this:

> I am inclined to think that the argument has to be wrong, that *a nativism pushed to that point becomes unsupportable, that something important must have been left aside.* What I think it shows is really not so much an a priori argument for nativism as that *there must be some notion of learning that is so incredibly different from the one we have imagined* that we don't even know what it would be like as things now stand.[11] (Piattelli-Palmarini, 1980, p. 269)

Another problem with encoding models is that they cannot account for organism-detectable representational error. In fact, they tend to have difficulties accounting for the very possibility of error at all, and *organism-detectable* error is not addressed. Attempts at modeling the "mere" possibility of error are from the perspective of some external observer of the organism, who is supposed to determine whether the organism's representations are correct or incorrect about something being represented in the environment – e.g., I think I see a cow, but my "external observer" can tell that it's really a horse on a dark night: in what sense is my representation of "cow" wrong and how can the external observer determine that?

But, if organism detectable error cannot occur, then error guided behavior and learning are not possible. We know that error guided behavior and learning occur, so any model that cannot account for organism-detectable error is thereby refuted. In general, the problem is that the only way I have to check my representation to see if it is correct is to use that same representation again. This is circular, and does not provide any knowledge of error.[12] Every standard model in the literature fails this criterion: Fodor, Millikan, Dretske, Cummins, symbol systems, connectionist models, and so on (Bickhard, 2009b, 2014).

[11] This is one of multiple places in which Fodor acknowledges that there are serious problems with his model. See, for another example, Fodor (1990).

[12] That this problem is equivalent to the classic radical skeptical argument provides some sense of its difficulty. The radical skeptical argument points out that, in order to check my representation, I would have to somehow step outside of myself and gain independent epistemic access to what I am trying to represent – become my own external observer – to be able to compare my representation with what is being represented. I cannot step outside of myself, so I cannot check my own representations. One common intuition is that I can check *consequences* of my representations: walk up closer to the "cow" and discover that it is in fact a horse. I think there is a germ of a correct approach here (anticipation), but, as stated, it simply checks one representation with another, and gives no reason to accept that either one is true or false. Such considerations can lead to pure coherence models. For further discussion, see Bickhard (in preparation).

The interactive model solves or dissolves both of these problems. Representation is *emergent* in particular kinds of functional processes (indications of interaction possibilities), and so does not require some base of already available representational atoms. And the possibility of *organism-detectable* error emerges in virtue of the future orientation of interactive anticipation: if the indicated interaction is engaged in and it does not proceed as anticipated, then the anticipation is falsified. Error detection is constituted not as peering back in time down the input stream, but in checking whether or not the future turns out to be as anticipated.

3.5 The Ontology of Sociality

Something interesting emerges when two or more complex agents are in each other's interactive presence: Complex agents pose special problems for each other, and solutions to those problems constitute social realities.[13] Sociality, then, – social ontology – arises intrinsically among complex agents (assuming that they arrive at solutions to the intrinsic problems).

3.5.1 Two Agents

Consider now two complex agents in each other's presence. Each seeks to interactively characterize their environments, which includes interactively characterizing the other agent. But the other agent's interactive potentialities are constituted in part by their own characterization of their environment, which includes the first agent. Each agent, then, needs to interactively characterize the other, including in terms of that agent's characterization of itself – which is constituted (in significant part) by the characterization of the other. There is regress here.

Each agent has an interest in resolving this characterization problem, but they cannot do so without arriving at compatible mutual characterizations of the situation together with each other's presence in that situation. This constitutes a coordination problem in the sense of Schelling (1963), and, thus, a resolution of the problem is (roughly) a convention in the sense of Lewis (1969).[14] I argue that social realities – *all* (forms of) social realities – are constituted by such conventions (Bickhard, 2008, 2009b). These are conventions concerning the interactive nature of the situation – *situation conventions*.

[13] Social realities range from momentary common understandings among participants to a conversation, to two oncoming pedestrians passing each other on the right, to a check out clerk relationship to a customer, to institutions of government, to language, to friendships, and so on. There are many kinds of such realities, and, generally, many instances of such kinds.

[14] For some discussion of important differences in the underlying model of convention from that of Lewis, see Bickhard (2008, 2009b).

3.5.2 Conventions

Lewis developed his model of convention most centrally in order to account for language as conventional, and he also generalized the model to other kinds of what might be called *institutionalized conventions* – those that repeat across people and occasions, such as driving on the right hand side of the road. I do not disagree with the characterization of such phenomena as conventional, but the model offered here arises not from considerations of language, though it has much to say about language (see below), but rather as a kind of fixed point problem that arises intrinsically whenever complex agents encounter each other. The problem is to find a fixed point, one that is relatively coherent and thus successful in the mutual situation knowledge anticipations, in the mutual characterizations of the agent-containing situations. This involves some changes in Lewis's underlying model, changes which have some potentially important consequences.

For a first point, not all situation conventions have to be institutionalized. In every conversation, for example, at any given point (assuming that the conversation is succeeding in being mutually intelligible) there is a mutual characterization of the situation among the participants to the conversation concerning the current status of the conversation – a kind of ongoing common knowledge. This ongoing, continuously updated, common understanding is what permits, for example, common resolution of pronouns, and, thus, maintenance of the ongoing common knowledge. The apperceptions of the participants to the conversation must maintain the coherence that constitutes the situation convention, if the conversation is to not fail in some way. Such momentary situation conventions may have never occurred before, and may never again. Such highly *non*-institutionalized situation conventions cannot be modeled with Lewis's model based on *regularities of behavior* – there are no repetitions across which to be regular, and the crucial coherences are among interactive characterizations, not among behaviors.

Another difference from Lewis's model is a more central involvement of normativity. One criticism of Lewis's model is that it does not account for the normativities of conventions and of upholding them among participants (e.g., Gilbert, 1989). There is a kind of normativity in that participants have a common interest in arriving at a solution to the underlying coordination problem, and that interest constitutes a form of normativity, but the convention itself, defined in terms of regularities of behavior, either exists or it doesn't. There seems to be simply a matter of fact there, without any strong normativity. Driving on the right hand side of the road, however, seems to involve a strong normativity, not just a conventional convenience. Similarly, presenting oneself in the role of teacher or judge is normative, not just a convenient solution to a coordination problem about how to act.

In the model outlined above, situation conventions arise not just as regularities of behavior, but as solutions to the basic problem of how to understand one's situation, and one's place – and others' places – in that situation. Such characterizations, in general, are central to one's functioning in the world, and one's place in that world. This is a much stronger sense of normativity in social situations.

For example, it is essential that one's *presentation of oneself* as a competent, reliable, social interactor with sufficient integrity to support the flow of situation conventions *be accepted by others* in order to function socially at all. Again, this is a much more central involvement of normativity than can be found in Lewis's model.[15]

3.5.3 Institutionalized Conventions

Social realities in the moment are constituted by situation conventions, but social realities across society are built on *institutionalized* conventions. Driving on the right hand side of the road is one example, but there are multiple other kinds of institutionalization that are important.

For example, the anticipative interactive coherences that constitute a situation convention may involve different anticipations for differing participants in the situation. They may be understood by all involved to be going to be engaged in differing courses of further interaction: one might drive animals toward another, while that other hides and waits for the opportunity to kill an animal and share it with the group. Such interrelated kinds of anticipations can constitute roles, and typifications of such interrelations can constitute types of roles (Berger & Luckmann, 1966). Uncles, for example, might be expected to engage in certain kinds of activities and stances with respect to nieces and nephews, but the role of "uncle" might also become typified such that a person who is not a biological uncle could take on the role of Uncle.

It was and is a major development when a society develops the notion of an institutional role in terms of its relations to other roles, and distinct from the person who might at a given time occupy that role. This is the foundation for bureaucracy, which, for all of its inefficiencies and frustrations, is much more efficient and powerful than individual (e.g., feudal) or biologically (e.g., clan) based social organization.

The model of situation conventions, thus, can address social realities from momentary common understandings and anticipations to broad social institutions. That includes, so I argue, language.

[15] Another criticism of Lewis's model is that it depends on a rather rigid model of rationality, and of unrealistic assumptions about how rational thought works (Gilbert, 1989). I will not present the interactive model of rationality here, but suffice it to say that, whether or not those criticisms are valid against Lewis, this model does not involve such models of rationality (see Bickhard, 2002, 2008).

3.5.4 How Are Situation Conventions Created and Evoked?

How do conventions emerge; how are they created or modified? One important answer is: By precedence and habituation. Suppose we happen to meet at a restaurant next Tuesday and have lunch together. It is a pleasant lunch, and the following Tuesday, perhaps we both go back to that restaurant and have lunch together again. Some Tuesday hence, after meeting at that restaurant several times, we will have a convention between us – a mutually coherent organization of our anticipations – that we meet at that restaurant on Tuesdays for lunch. Crucially, we need never have talked about or negotiated this arrangement in order for it to emerge and constitute a convention.[16]

Lunch on Tuesday conventions are evoked by the arrival of Tuesdays. Many conventions, however, are evoked by activities on the part of one or more participants, which may or may not have calendar or other kinds of contextual constraints. I may, for example, invoke the convention of walking to the right of oncoming pedestrians by tending slightly to the right. That is, I do something that *presupposes*, and thereby evokes, the convention to be invoked. If this is understood and accepted by the oncoming people, we act in accordance with that convention. The *presupposition-in-action* yields the convention presupposed.

Of course, I may be in a country in which it is customary to walk to the left, or the oncoming person may misunderstand my tending to the right "gesture" as indicating, for example, that I am heading into a store. The invocation of conventions is by presupposition of those conventions, but the epistemology about such conventions is inherently tentative and fallible.

Note that the sense in which the mutual anticipations are coherent is an implicit sense. It need not involve any explicit representations of the fact that any such convention exists, or even that conventions as a "kind" exist. What is required is a coordinative framework of anticipations that mutually works as those anticipations are engaged. What prevents or annihilates convention is a failure of the anticipations to be interactively coherent, or, in some more sophisticated cases, an explicit negation on the part of one (or more) of the parties involved about the anticipations to be played out – such as a con man who wishes to deliberately invoke certain social realities that he knows are not correct (do not actually exist) because there is an explicit intention on his part, thus anticipation, to violate that anticipatory framework at some opportune time.

Convention, in these senses, does not require explicit representation of convention or of the situation as being conventional. It is because of this sufficiency of implicitness that even infants (and adults) can participate in conventional interactions, such as peek-a-boo, without having to think through any layers of "I know

[16]This kind of possibility for the establishment of convention is of central importance for Lewis (1969), who wants to characterize language as conventional: if the only way to establish a convention is via negotiation, then what is the language in which the conventions that constitute language are negotiated?

that you know that I know …". What is required is interactive characterization coordinative coherence *in fact*, and a lack of doubt or negation (at any reflective level) of such coherence.

Convention emerges, when it does, in processes of apperception concerning social situations. Apperception is the process of updating and maintaining "knowledge" of the interactive nature of a situation: apperception can and does change prior or ongoing situation knowledge.

So, convention emerges in apperceptive processes, processes that are evoked by, for example, actions that presuppose convention. Actions may also evoke processes that *change already existing convention*: action can evoke apperception, which updates situation knowledge, and, if social participants' situation-knowledges are updated in a coherence maintaining fashion, some new convention will now obtain. Still further, such convention altering effects can be themselves conventionalized: all convention invocation is apperception invocation, and apperception is inherently context dependent, so when that context dependency is itself conventionalized, we have conventional means for invoking or changing convention.

Insignia of rank, for example, evoke a conventional framework of expectations and obligations in contexts in which those insignia are relevant and operative. They change the respective situation knowledges of the participants. An even more powerful tool for such alterations of and operations on situation conventions is language (see below).

3.5.4.1 Emergence and Implicitness

It is worth further emphasizing here two properties of convention in this model: emergence and implicitness. It is because of the possibility of *implicitness*, for example, that hierarchies of reflective thoughts are not required for the existence of conventions and conventionalized processes – for infants, toddlers, and adults.

It is because of the *emergent* nature of convention that no basic primitive social ground has to be posited, such as we-intentionality (Searle, 1995) or the we-mode (Tuomela, 2013). In both cases, these are not derivable from the "I"-version. These are explicitly *not* emergence models, and one important reason why they cannot be is because of the background assumption that representation and reason are necessarily explicit, based generally on explicit propositional encodings (see also Bratman, 2007; Gilbert, 1989), and there is no way to account for the emergence of foundational encodings (Bickhard, 1993, 2009b). Implicitness makes possible the emergence of *in fact* interactive anticipatory coordinative coherence without anyone having to know or explicitly represent that is what is going on, let alone to think through some hierarchy of reflections on such "knowledge".

3.5.5 A Word About Language

The model outlined here does not and cannot support standard encodingism frameworks – usually of an information semantic form, though not necessarily – that posit information "coming in" via perception, being processed in cognition, and then yielding outputs in the forms of action and utterances. That is, it cannot support standard approaches to perception, cognition, or language (Bickhard, 2009b).[17] Here I would like to say a bit more about how this model nevertheless addresses language – arguably with much more success than standard approaches, though elaborating that claim must be reserved for elsewhere.[18]

In this model, utterances modify social realities, via apperceptions of those utterances – where social realities, as outlined above, are constituted as coordinative relations among various social participants' interactive characterizations of their situation. These characterizations, in turn, are constituted as representational anticipations of interactive possibilities: thus, utterances *modify* representations, but do not *transmit* them via encodings. They "operate on" conventions (via apperception) about the current situation.

Social realities are solutions to coordination problems concerning what future possibilities might be anticipated. In human social realities, those future potentialities are massively, though not entirely, constituted in potentialities for further language. Thus, language is ontologically a major part of those social realities.

Language, then, functions in accordance with conventions concerning the apperceptive effects of utterances and parts of utterances (Bickhard, 1980). In a broad sense, it is an institutionally conventionalized toolbox for the construction of utterances which have conventionalized effects on situation conventions.

3.5.5.1 Sub-utterances and Grammar

A "full" utterance in this model elicits a full transformation of a situation convention. Partial or sub-utterances can create (*linguistic* situation convention) circumstances in which other kinds of sub-utterances can function. For example, one partial

[17]Information semantics is the currently dominant framework within psychology, cognitive science, and philosophy. It constitutes a fundamental equivocation between technical information – a condition of being correlated with – and semantic, or representational, information. If X is correlated with Y, then knowing X can permit inferring something about Y – but only if X and Y and the correlation are already known and represented. If all of these are known, then X can be used as an encoding of (properties of) Y, but, as usual, encodings require that all of the relevant representations be already available. Correlational information cannot, in itself, constitute representation. This point is relatively well known, though often ignored, but what else is required has no resolution. The problem of organism-detectable error, discussed above, is one of several reasons why it *cannot* be resolved. See Bickhard (1980, 1993, 2009b, 2014) for further discussion.

[18]For more developed discussions of language, and also of how perception can be modelled within this framework, see, for example, Bickhard (2009b; Bickhard & Richie, 1983).

utterance may differentiate some part or aspect of the overall commonly understood (or at least commonly and coordinatively anticipated) situation convention, making it a focus for the second part, while the second partial utterance might modify that part or aspect. In "The ball is red", for example, "the ball" will, if successful in creating and maintaining situation convention, differentiate a ball that is represented in common among the participants, while "is red" will, perhaps, modify those representations to include the property of being red, or at least will bring into the common knowledge of the situation convention the fact or assumption that the ball is red. By differentiating *types* of incomplete utterances in this manner, a grammatical framework emerges that has important similarities to categorical grammars (though also some important differences) (Bickhard & Campbell, 1992).[19]

3.5.6 Persons

There is a special emergence, according to this model, of social persons. Persons are agents, and they are social agents, but their sociality is fundamentally different from that of, for example, social insects. Ants are social, and they are emergently social, but the sociality is at the level of the ant nest as a whole. There is no social emergence at the level of individual ants.

In contrast, persons are intrinsically social agents at an individual level. Human infants are intrinsically open to developing as social persons, but the manner in which they do so and the particular agents that they become is deeply dependent on familial, social, and cultural contexts (Berger & Luckmann, 1966). Human infants become social persons within that context in the sense that they become agents that can participate in the social process and realities of that family, society, and culture – and thereby participate in the constitution of those social realities.

Persons are developmentally emergent social agents: agents that have developed as participants in socio-cultural social realities, and that thereby co-constitute those social realities. Their ontology is at the same level as human sociality per se. In particular, they constitute an emergent kind of process within the general framework of complex agents (Bickhard, 2013).

This is important for the current discussion in that it entails that persons are intrinsically involved, ontologically involved, in the normativities of social processes and realities.

[19]Note also some similarities here to file-change models, though without any assumptions of or restrictions to explicit propositional encodings (Heim, 1983; Lewis, 1979).

3.6 Stakes in Sociality

Persons, then, are ontologically constituted as socio-cultural interactive agents. As such, persons have normative stakes in their sociality – for example, successful functioning in social realities requires sufficient establishment and maintenance of those social realities, which, in turn, requires presenting oneself as – and being accepted as – reliable, competent, and trustworthy social agents. Successful functioning in social realities constitutes the processual existence of the social person, similar to the sense in which "successful" burning constitutes the existence of a candle flame – a strong stake indeed.

3.6.1 Sociality and Historicity

Social realities require constant maintenance, repair, and development. To be a person is, among other things, to be able to participate in such maintenance, repair, and development. This requires the ability to respond to failures of social realities, ranging from misunderstandings to encounters that go beyond current social realities, such as maintaining and developing a relationship in encountering a crisis for one or more parties to that relationship – or, of importance for current purposes, in unfolding further emergent aspects and characteristics of social realities, such as friendships. That is, social realities necessarily have histories of development – and deep social realities, e.g., long standing friendships or marriages, have deep histories of development – "deep" in the sense that more and more aspects of the emergent person are dependent for their interactive realization on those friendships or partnerships or other kinds of relationships.

Such relationships are themselves conventions, generally between two or a few people, that have developed and are developing over longer time frames and with respect to more central concerns and values of the persons involved. The historistic development of such relationships requires values and normative stakes with respect to which the learning and development of the relationships can occur – to constitute criteria of appropriateness and inappropriateness, success and failure, in the further development of a relationship, and, thus, of the persons involved. The relationships are in important respects the conventions of such *development*, not just of particular interactive common understandings in the moment.

In sum, such learning requires normative stakes relative to which learning can occur, and, thus, such developmental historicities can be realized.

3.7 Robots and Sociality

Robots can simulate complex forms of social interaction, and they could be programmed to detect various pre-designated kinds of "failures" and to use various heuristics to recover and learn from them – a simulation of normativity, and of the learning and development that is based on such normativity. But they have no intrinsic stake in the world, nor in their existence in the world, nor in their existence as social agents.[20] Thus, they can only simulate interaction forms and criteria for failures that have been anticipated in their design. They are limited to what their design can anticipate. In that sense, they cannot have genuine developmental historicity, thus not genuine social being.

3.7.1 Is This Limitation Necessary?

Could, however, artificial agents be developed that *do* have intrinsic sociality and its normativities? I would argue "Yes", but they would have to have inherent emergent stake in their existence as social agents, which, in turn, requires having a stake in their existence in the world.

To accomplish this would require, among other things, a model of emergent normativity as it occurs in biology, and human beings in particular. An artificial agent, then, could presumably realize or instantiate such normativity.

As discussed above (and, e.g., Bickhard, 2009b), this requires certain kinds of far-from-thermodynamic-equilibrium processes. That is, normativity emerges most primitively as normative function, and normative function emerges in certain kinds of processes that are far from thermodynamic equilibrium. Far-from-equilibrium processes have a stake in the world in the sense that their existence requires maintenance of the far-from-equilibrium conditions that existence depends on. Their existence is at stake in whether or not those far from equilibrium conditions are in fact maintained. Robots, as we currently understand them, do not realize such processes – they are (primarily) not far from equilibrium – but artificial systems could, in principle, realize such far-from-equilibrium processes, and, thus, realize emergent normativities of various sorts.[21]

[20] With no stake in their existence in the world, they cannot have a stake in their existence in the world as social agents.

[21] For further discussion of this issue regarding artificial agents, see Bickhard (2009a).

3.7.2 Robots as Moral Agents

The issue of having a stake in their own existence is also of relevance for moral issues. In particular, this would seem to be at the center of being an object of moral regard – a moral patient. Moral regard is, in this sense, a kind of respect for the stake that other beings have in their own existences. Having such a stake, however, is not, in general, sufficient for being a moral *agent*. For example, an infant has a stake in the world, and in its own existence in the world, and, thus is an object of moral regard, but it is not (yet) a moral agent. A robot, thus, that has a stake in its own existence would be a an object of moral regard; there would be, perhaps, moral responsibilities toward such beings. But they would not just in virtue of that be moral agents, agents with moral responsibility. What is further required for moral agenthood must be addressed elsewhere.

3.8 Conclusion

There are multiple reasons for working with a process metaphysics, and multiple important consequences to doing so. Among these consequences are the possibilities of emergence, and, in particular, normative emergence. With an initial emergence of normative function in far-from-thermodynamic-equilibrium processes, a directly related model of the emergence of representation, with its normative dimension of truth and falsity, is easily constructed. This is in terms of the normative characteristics of an agent's indications or anticipations of what it could do next, and the possibility that those indications or anticipations might be false – indicating that some interactive possibility holds presupposes that sufficient supporting conditions hold, and they might not.

With such an inter-action based model of representation and cognition, a special problem arises when two (or more) complex agents are in each other's presence: characterizing the situation requires characterizing the other agent, including that agent's characterization of the first agent, which includes the first agent's characterization of the second, and so on. Insofar as both agents have an interest in arriving at a successful characterization, this constitutes a coordination problem in the sense of Schelling, and, thus, any solution to this problem constitutes a convention in a (broadened) sense of Lewis. I argue that conventions, modeled in this broader manner, capture all social realities, from the momentary to the institutional to language to the deep, such as a friendship. One important characteristic is that the persons who participate in and constitute such social realities have normative stakes in their participations, and in their existence and acceptance as agents who are capable of the creation, maintenance, and development of such conventions. This is especially important for social processes, such as friendships, that develop over longer time periods and engage central values of the persons involved.

It is such stakes that current robotics does not capture, and, thus, contemporary robots cannot be genuine social agents. Robots as currently designed are limited to various kinds and degrees of simulation of sociality. For some purposes, that may suffice. They cannot be genuine social and moral agents, however, without having an intrinsic stake in their existence as social agents. Among other consequences: they cannot, in contemporary form, be genuine friends. That will require different kinds of design and development principles than are currently available.

Acknowledgements This is an expanded version of Bickhard, M. H. (2014). Robot Sociality: Genuine or Simulation? In J. Seibt, R. Hakli, M. Nørskov (Eds.) *Sociable Robots and the Future of Social Relations*. Series: Frontiers in Artificial Intelligence and Applications. (81–86). Amsterdam: IOS Press.

References

Aitchison, I. J. R. (1985). Nothing's plenty: The vacuum in modern quantum field theory. *Contemporary Physics, 26*(4), 333–391.

Aitchison, I. J. R., & Hey, A. J. G. (1989). *Gauge theories in particle physics*. Bristol/England: Adam Hilger.

Allen, J. W. P., & Bickhard, M. H. (2013). Stepping off the pendulum: Why only an action-based approach can transcend the nativist-empiricist debate. *Cognitive Development, 28*, 96–133.

Berger, P. L., & Luckmann, T. (1966). *The social construction of reality*. New York: Doubleday.

Bickhard, M. H. (1980). *Cognition, convention, and communication*. New York: Praeger Publishers.

Bickhard, M. H. (1993). Representational content in humans and machines. *Journal of Experimental and Theoretical Artificial Intelligence, 5*, 285–333.

Bickhard, M. H. (2002). Critical principles: On the negative side of rationality. *New Ideas in Psychology, 20*, 1–34.

Bickhard, M. H. (2003). Variations in variation and selection: The ubiquity of the variation-and-selective retention ratchet in emergent organizational complexity, part II: Quantum field theory. *Foundations of Science, 8*(3), 283–293.

Bickhard, M. H. (2005). Consciousness and reflective consciousness. *Philosophical Psychology, 18*(2), 205–218.

Bickhard, M. H. (2008). Social ontology as convention. *Topoi, 27*, 139–149.

Bickhard,M. H. (2009a). The biological foundations of cognitive science. *New Ideas in Psychology, 27*, 75–84.

Bickhard, M. H. (2009b). The interactivist model. *Synthese, 166*, 547–591.

Bickhard, M.H. (2013). The emergent ontology of persons. In J. Martin & M. H. Bickhard (Eds.), *The psychology of personhood: Philosophical, historical, social-developmental, and narrative perspectives* (pp. 165–180). Cambridge: Cambridge University Press.

Bickhard, M. H. (2014). What could cognition be, if not computation ... or connectionism, or dynamic systems? *Journal of Theoretical and Philosophical Psychology, 35*(1), 53–66.

Bickhard, M. H. (in preparation). *The whole person: Toward a naturalism of persons—Contributions to an ontological psychology*.

Bickhard, M. H., & Campbell, R. L. (1992). Some foundational questions concerning language studies: With a focus on categorial grammars and model theoretic possible worlds semantics (target article for commentary in a special double issue). *Journal of Pragmatics, 17*(5/6), 401–433.

Bickhard, M. H., & Richie, D. M. (1983). *On the nature of representation: A case study of James Gibson's theory of perception*. New York: Praeger Publishers.
Bratman, M. E. (2007). *Structures of agency*. Oxford: Oxford University Press.
Brown, H. R., & Harre', R. (1988). *Philosophical foundations of quantum field theory*. Oxford: Oxford University Press.
Campbell, D. T (1974). Evolutionary epistemology. In P. A. Schilpp (Ed.), *The philosophy of Karl Popper* (pp. 413–463). La Salle, IL: Open Court.
Campbell, R. J. (1992). *Truth and historicity*. Oxford: Oxford University Press.
Campbell, R. J.(2011). *The concept of truth*. New York: Palgrave Macmillan.
Campbell, R. J., & Bickhard, M. H. (2011). Physicalism, emergence and downward causation. *Axiomathes, 21*, 33–56.
Cao, T. Y (1999). *Conceptual foundations of quantum field theory*. Cambridge: University of Cambridge Press.
Chang, C. C., & Keisler, H. J. (1990). *Model theory*. Amsterdam/New York: North Holland.
Christensen, W. D., & Bickhard, M. H. (2002). The process dynamics of normative function. *Monist, 85*(1), 3–28.
Davies, P. C. W.(1984). Particles do not exist. In S. M. Christensen (Ed.), *Quantum theory of gravity* (pp. 66–77). Bristol, England: Adam Hilger.
Dewey, J. (1960). *The quest for certainty*. New York: Capricorn Books. (Originally published in 1929).
Fodor, J. A. (1981). The present status of the innateness controversy. In *Representations: Philosophical essays on the foundations of cognitive science* (pp. 257–316). Cambridge, MA: MIT Press.
Fodor, J. A. (1990). Information and representation. In P. P. Hanson (Ed.), *Information, language and cognition* (pp. 175–190). Vancouver: University of British Columbia Press.
Gilbert, M. (1989). *On social facts*. Princeton, NJ: Princeton University Press.
Gill, M. L. (1989). *Aristotle on substance*. Princeton, NJ: Princeton University Press.
Graham, D. W. (2006). *Explaining the cosmos*. Princeton, NJ: Princeton University Press.
Guthrie, W. K. C. (1965). *A history of Greek philosophy* (The presocratic tradition from Parmenides to Democritus, Vol. II). Cambridge: Cambridge University Press.
Hale, B., & Wright, C. (2000). Implicit definition and the a priori. In P. Boghossian & C. Peacocke (Eds.), *New essays on the a priori* (pp. 286–319). Oxford: Oxford University Press.
Halvorson, H., & Clifton, R. (2002). No place for particles in relativistic quantum theories? *Philosophy of Science, 69*(1), 1–28.
Heim, I. (1983). File change semantics and the familiarity theory of definiteness. In C. S. R. Bauerle & A. Stechow (Eds.), *Meaning, use, and interpretation of language* (pp. 164–189). Berlin: de Gruyter.
Huggett, N. (2000). Philosophical foundations of quantum field theory. *The British Journal for the Philosophy of Science, 51*(supplement), 617–637.
Hume, D. (1978). *A treatise of human nature*. Oxford: Oxford University Press. (Index by L. A. Selby-Bigge; Notes by P. H. Nidditch).
Kuhlmann, M., Lyre, H., & Wayne, A. (2002). *Ontological aspects of quantum field theory*. River Edge, NJ: World Scientific.
Lewis, D. (1969). *Convention*. Cambridge, MA: Harvard University Press.
Lewis, D. (1979). Scorekeeping in a language game. *Journal of Philosophical Logic, 8*, 339–359.
Millikan, R. G. (1984). *Language, thought, and other biological categories*. Cambridge, MA: MIT Press.
Millikan, R. G. (1993). *White queen psychology and other essays for Alice*. Cambridge, MA: MIT Press.
Piaget, J. (1954). *The construction of reality in the child*. New York: Basic.
Piattelli-Palmarini, M. (1980). *Language and learning: The debate between Jean Piaget and Noam Chomsky*. Cambridge, MA: Harvard University Press.
Schelling, T. C. (1963). *The strategy of conflict*. New York: Oxford University Press.

Schurz, G. (1997). *The is-ought problem: An investigation in philosophical logic* (Trends in logic, Vol. 1). Dordrecht/Boston: Kluwer Academic.
Sciama, D. W. (1991). The physical significance of the vacuum state of a quantum field. In H. R. B. S. Saunders (Ed.), *The philosophy of vacuum* (pp. 137–158). Oxford: Clarendon.
Searle, J. R. (1995). *The construction of social reality*. New York: The Free Press.
Tiles, J. E. (1990). *Dewey*. London/New York: Routledge.
Tuomela, R. (2013). *Social ontology: Collective intentionality and group agents*. Oxford: Oxford University Press.
Weinberg, S. (1977). The search for unity, notes for a history of quantum field theory. *Daedalus, 106*(4), 17–35.
Weinberg, S. (1995). *The quantum theory of fields* (Foundations, Vol. 1). Cambridge: Cambridge University Press.