ARTICLE IN PRESS

PERGAMON

New Ideas in Psychology ■ (■■■■) ■■■–■■■

NEW IDEAS IN PSYCHOLOGY

www.elsevier.com/locate/newideapsych

Critical principles: on the negative side of rationality

Mark H. Bickhard*,[1]

Department of Philosophy, Lehigh University, 15 University Drive, Bethlehem, PA 18015, USA

Abstract

A naturalized model of rationality is developed, with a focus on an important but largely neglected aspect: knowledge of error, or "negative" knowledge. The development of knowledge of what counts as error occurs via a kind of internal variation and selection, or quasi-evolutionary, process. Processes of reflection generate a hierarchy of principles of error, a hierarchy that frames and constrains positive rationality. The dynamics of rationality is an internalization of processes of reflected-upon variation and selection. The nature and origin of logic are addressed from within this framework, and the overall rationality model is applied to three central issues in the philosophy of science: the rational function of truth and realism in science, the nature of progress in science, and the rationality of certain induction-like considerations. © 2001 Published by Elsevier Science Ltd.

Keywords: ■; ■; ■

Rational activity requires knowledge of how to do things in ways that tend to avoid or overcome error. This knowledge involves regulation of those activities, and there are at least two fundamental sorts of such regulation: (1) regulation of the processes of *interaction* between the rational system and its environment (including autoregulation of the system's regulatory processes) (Hooker, 1995), and (2) regulation of the processes of the *construction* of new interactive (sub)systems—a kind of meta-regulation. In both cases, error feedback is involved. In the case of interaction between a system and its environment, error feedback can be an aspect of the interactions per se: knowledge of error via encounters with that error. Hopefully, these encounters are via lower cost vicariants of—surrogates for—actual error

*Tel.: +1-610-758-3633.

E-mail address: mhb0@lehigh.edu, mark.bickhard@lehigh.edu (M.H. Bickhard).

[1] http://www.lehigh.edu/~mhb0/mhb0.html

PII: S0732-118X(01)00010-1

selection pressures (Campbell, 1974). In the case of system constructions, error feedback and error knowledge—knowledge of possible errors, or *negative knowledge*—take on a somewhat different form. These considerations are the primary focus of this essay. I argue that such negative knowledge, and the intrinsic tendency towards its construction, are central to rationality.

Regulation via error feedback—whether interactive or constructive—constitutes an internalization of the basic variation and selection processes of evolution. It is useful to internalize this process, as much as is possible, because of the potentially high cost of encountering full unmitigated selections. In some species, in fact—such as human beings—the ability to developmentally internalize variation and selection processes at the level of the individual organism is among the most important attributes. This ability is the result of a massive macro-evolutionary trend in that direction: increasing abilities to learn and to learn to learn (Bickhard, 1980b; Campbell & Bickhard, 1986).

I propose that rationality in a broad sense is precisely this internalization of variation and selection processes. Even relatively simple organisms can have error-corrected and error-guided *actions*. In human beings, there is a powerful inherent tendency to further internalize variation and selection processes in individual *development*. This internalization will yield not only processes for accomplishing various tasks, but also methodological processes for constructing and evaluating such processes, metamethodologies, and so on. I will be emphasizing in particular the internalization of knowledge of selection pressures—errors—with regard to the constructive processes of further development. This focus is complementary to that of Hooker (1995).

1. Preliminaries

1.1. Toward a naturalism of rationality

The deeper project, of which this offers a small part, is that of a thorough and strict naturalism (Bickhard, 1993, 1998a, in preparation-a; Brown, 1988; Hooker, 1987, 1995). It is now accepted that fire and heat and life, for example, are all natural processes in the world. It is widely assumed that the mind, and persons more broadly (Bickhard, in preparation-a), are also natural parts of the world, but understanding how that is so, or even possible, is still a severe problem—so much so that there is serious scepticism that it will ever be accomplished (Chalmers, 1996; Geach, 1977; McGinn, 1993; Nagel, 1986). The aspect of thought that we call "rationality" is one of the particularly difficult challenges to naturalism. It involves not only multiple properties of mind and thought per se, such as intentionality and motivation, but also a normativity about thought. Normativities of all kinds pose interesting problems for naturalism: they all pose some version of the question of how an "ought" could possibly derive from or emerge from an "is" (Bickhard, 1998b).

1.2. Requisites for naturalism

It might seem that naturalism is obviously correct, and its importance dismissed precisely because it seems so obviously true. Such a stance, however, would seriously underestimate the difficulties of carrying through a project of naturalism: It is distressingly easy to espouse naturalism, but nevertheless to fail in a project of naturalism—such failures can be subtle and far from obvious. Discovering such errors, then, can be important and non-trivial, and pointing out such errors can be an important and non-trivial form of criticism. Naturalism is a powerful position from which to evaluate models of mind and persons. Many models fail to be consistent with naturalism in spite of the best intentions of their authors.

The criteria that must be met in order for a naturalistic model to be successful in that naturalism are manifold, and the details of such criteria can vary with the history of science—with our current overall understanding of the nature of the natural world. One criterion that I have found to be of particular importance and power, and that will be a tool in the following discussion, is the problem of origins. If a model of a particular phenomenon makes it impossible for that phenomenon to have come into existence, then the model cannot be correct (Bickhard, 1979, 1991c; Bickhard & Campbell, 2000). Naturalistic models must be consistent with our best current science (unless there are good reasons to challenge parts of that science), and one aspect of current science is that most of everything we might be interested in understanding—stars, life, mind, rationality—once did not exist. They did not exist at the moment of the Big Bang, for example. But if they once did not exist, and they now do, then they have to have come into existence somehow—they have to have emerged (Bickhard, 1993, 1998b; Bickhard & Campbell, 2000; Campbell & Bickhard, in preparation; Horgan, 1993; O'Conner, 1994). An essential characteristic of any naturalistic model of any phenomena, therefore, is that it be consistent with the natural emergence of those phenomena. Many contemporary models in the study of the mind and persons fail this criterion of emergence (Bickhard, 1991a, 1993, 1998a,b; Bickhard & Christopher, 1994; Bickhard & Terveen, 1995).

1.3. A framework of assumptions

I will not undertake a naturalism here starting from basic physics, but will assume a framework of various phenomena as the tools with which to model an aspect of rationality, with the underlying assumption that these phenomena either already have been, or at least in principle can be, naturalized themselves. I will be assuming the acceptability of functional forms of analysis and modeling, and therefore that *normative function* itself can be naturalized—a non-trivial problem. I will also be assuming a particular naturalistic model of *representation*, called interactivism—also highly non-trivial—and several properties and possibilities that (naturalistically) follow from that favored model of representation: specifically, an *implicitness* and intrinsic *modality* of emergent representation, an *evolutionary epistemology*, and the possibility of *reflection*, of representations of representations. Here are the core intuitions of these models. More elaborate presentations of these models can be found elsewhere (Bickhard, 1993, 1998a,b; Bickhard & Terveen, 1995; Campbell & Bickhard, 1986).

1.3.1. Function

A function of a system or process is one or more of its causal consequences (Wimsatt, 1972). But all systems and processes have many consequences that are not involved in their functions, assuming they have any functions at all. The heart, for example, not only pumps blood, but it also creates heart beats and consumes oxygen. The latter two are not normally considered to be functions of the heart. But, then, on what basis can we determine which consequences are functions? What of hearts that are not pumping properly: do they have a function or not? The notion of "*not* pumping properly" relies on a notion of "functioning properly": how can we distinguish function from dysfunction? The problems of a naturalization of function are not simple.

I will be relying on a model of natural function—not designed or intended function—as being emergent only in certain kinds of open systems. The intuition of the model is that open systems (a flame, for example) require the maintenance of various conditions (such as combustion temperature) in order for the systems themselves to continue to exist—and, therefore, to continue to have their natural causal influences in the world (whatever they may be). Functional consequences are those that contribute to the maintenance of such conditions, and thus that contribute to the maintenance of the existence of the open system itself (and conversely for dysfunctional consequences). In turn, they contribute to the maintenance of whatever natural consequences that system might have (Bickhard, 1993, 1998b, forthcoming; Christensen & Bickhard, in press).[2]

1.3.2. Representation

Just as one of the serious problems for a naturalization of function is to make sense of dysfunction, one of the serious problems for a naturalization of representation is to make sense of (the very possibility of) representational error (Fodor, 1987, 1990a). There is a large literature on this task, but virtually none of it would satisfy a strict naturalism even if it succeeded on its own terms—and even "its own terms" are generally conceded to have not been met: "Deep down, I think I don't believe any of this. But the question what to put in its place is too hard for me" (Fodor, 1990b, p. 190).

One reason that success in the attempted models of representation would not yield a naturalistic model is that they involve an intrinsic and ultimately circular dependency on an observer. The primary criterion that is set for modeling representational error is that an observer of the supposed representational system could distinguish correct representations from false representations (as, for example,

[2] It should be noted that the dominant etiological approach to modeling function (Godfrey-Smith, 1994; Millikan, 1984, 1993) suffers from a serious failure of naturalism: etiological function cannot be constituted in the current state of a system, but only current state can be causally efficacious. The etiological model of function, then, is a model of causally epiphenomenal function (Bickhard, 1993, 1998b, forthcoming; Christensen & Bickhard, in press).

if Fodor's asymmetric dependency condition succeeded "on its own terms"-—Loewer & Rey, 1991). Success in this endeavor would at best yield a notion of representational error, and, thus, representation, that was dependent on such an observer. But observers are precisely what we are attempting to model in the first place—or at least their representations—so this dependency introduces a fatal circularity.

Furthermore, if representational error can be distinguished *only* from the perspective of an observer, then the offered model of representation cannot be accepted even if the circularity per se is ignored. There are many processes and activities that are guided by representational error—error feedback in interaction and error selection in learning, for example—and these require representational error that is detectable *by the system itself*, not just by an observer. A system cannot guide interaction or learning with respect to error that it cannot detect. None of the standard approaches to representation can satisfy this fundamental criterion (Bickhard, 1999, in press-c; Bickhard & Terveen, 1995).[3]

Note that a central argument of radical skepticism is that any check on representational error is circular, and, therefore, knowledge is not possible. If epistemic access to the world is in terms of epistemic correspondences, then correspondences cannot be checked except via those correspondences: circularity. This argument has had a rather long and successful career (Annas & Barnes, 1985; Barnes, 1990; Burnyeat, 1983; Groarke, 1990; Hookway, 1992; Popkin, 1979; Rescher, 1980). But the dependency of action and learning on precisely such error suggests that skepticism poses serious problems for the scientist as well as for the philosopher: representational feedback interactions and representational learning cannot be modeled in any framework that does not avoid the skeptical argument concerning error. Conversely, that career suggests that this is a problem that is non-trivial to solve.

I have argued that the error problem cannot be solved within the standard approaches to representation—approaches that have dominated thinking about the mind and representation for a very long time. Such approaches, which I collectively call *encodingism*, assume that representational relationships are some special form of correspondence relationships—correspondences between representations and the represented (e.g., Clark, 1997; Dretske, 1981, 1988; Fodor, 1987, 1990a,b, 1998; Hanson, 1990; Smith, 1987).[4] Correspondences, however, are ubiquitous throughout the universe, and very few of them are representational. Attempting to figure out which ones are, and how they are, has consumed a great deal of effort.

[3] Fodor wishes to postpone such issues of epistemology in favor of a "metaphysics first" strategy (Fodor, 1998). This is potentially an acceptable strategic move, but not if the proffered metaphysics makes an epistemology *impossible* (Levine & Bickhard, 1999).

[4] Millikan's model of representation (Millikan, 1984, 1993) is not straightforwardly a correspondence model, but the epiphenomenality of the etiological model of function visits itself on the derivative model of representation, creating an epiphenomenal—non-naturalistic—model of representation. Cummins (1996) has also introduced a model of representation that does not immediately fail the error criterion, but it suffers from other fatal problems (Bickhard, in press-c).

I argue that the *only* genuine correspondence representations are those that make use of the very representational properties that we want to account for in a model of emergent and original representation. Morse code is an example of genuine correspondence representations: "dot dot dot" is in correspondence with "S", and it is so because it is known to be so by everyone that uses Morse code. This works just fine for genuine encodings (or ciphers), such as Morse code, but is fatally circular if we are attempting to account for representation per se: Morse code correspondences are representational only because we, who already possess the capacity to represent, represent them that way.

Morse code is conventional, and it might be thought that renders the example nugatory. Consider, then, a natural example: neutrino fluxes "encode" properties of fusion processes in the sun, but do so only insofar as the physicists involved already represent the fluxes and their detection, the solar fusion processes, and the relationships between them. Again, codes require that both ends of the encoding relationship and the relationship itself be already represented. This is not a problem for genuine codes, but it is a fatal problem if encodings are taken to be the fundamental nature of representation.

The alternative that I offer, and will be assuming in the following discussion, is called *interactivism*. Interactive representation is emergent in interactive system organization. In this view, representation is a phenomenon of pragmatics, not just the processing of inputs. The framework for the model is consonant with the action focus of Peirce (Hookway, 1985; Joas, 1993; Mounce, 1997; Rosenthal, 1983), rather than the focus within a passive consciousness of Plato or Aristotle.

The intuition of the interactive model of representation begins with the recognition that the course of an interaction between a system and its environment can be functionally anticipated in and by the system. This could be accomplished by, for example, pointers to (sets of functionally anticipated) "next states", or functional readinesses to process restricted classes of potential inputs, or functional readiness to proceed in restricted forms of further interaction, and so on. Such functional anticipations can be *falsified* by the actual course of the interaction—the environment might not satisfy or fulfill the anticipations. Such potential falsification constitutes the emergence of truth value, the fundamental, criterial, property of representation (Bickhard, 1993, 1999, in press-c). In particular, error in such anticipations is not only definable, but is also detectable in and by and for the system itself: failure of functional anticipations is a *functionally* detectable condition. In this view, the skeptical argument that discovering epistemic error is impossible is valid, but unsound: it is based on a false assumption about the basic category within which representation is to be understood—correspondence.

Such a minimal emergence of naturalistic truth value is a long way from accounting for more familiar representational phenomena. The minimal model might account for the representations of very simple organisms, but what about representations of objects, space and time, numbers and other abstractions? What is interacted with for such abstract representations? And what about other phenomena that are commonly thought to be representational in nature, such as language? I will not address these issues here, but will assume that all are more complex versions of interactive representation, and interactive phenomena in general (see, in addition to references mentioned above, Bickhard, 1980a, 1987, 1992c, 1995, 1996, 1998a; Bickhard & Campbell, 1992; Campbell & Bickhard, 1992b).

1.3.3. Implicitness

One important property that is inherent in interactive representation is that of implicitness. Interactive representations differentiate those environments, and environmental properties, that would support the interactive anticipations from those that would not—but those differentiations are strictly implicit. There is no explicit representation of those environments or properties per se. There *can* be explicit representations of *relationships* among such implicitly defined categories: for example, any encounter with anything in "this" implicitly defined class of environments will also be an encounter with "that" implicitly defined class of environments (Bickhard, 1998a).

One important aspect of representational implicitness is that environmental equivalence classes can be differentiated without having any further knowledge or representation about any specifics of the environments in those classes; this stands in contrast to encoding models, in which all representation is explicit (witness the encodingist arguments about whether frogs, which will tongue-flick at flies, but also at BBs, "really" represent flies or BBs or little black dots, etc., Fodor, 1987, 1990a; Loewer & Rey, 1991). A related aspect is that implicit representations are unbounded: there is no a priori bound on what will fall in a differentiated class.

1.3.4. Modality

The anticipations that constitute interactive representation are anticipations of possibilities—of possible further interactive process. An indication that a class of interactions is possible, or would be possible if such-and-such an intermediate interaction were engaged in, does not *force* any of those interactions to occur. It is not a simple causal relationship. In general, in fact, there may be *many different* further interactive possibilities indicated, not all of which can be pursued simultaneously. A frog, for example, might have an occurrent indication of the possibility of tongue flicking and eating, evoked by a fly, and simultaneously an indication of the possibility of pain or worse, evoked by a shadow of a hawk. The frog will most likely jump into the water instead of go for the fly.

Furthermore, a particular class of interactions being indicated may be unbounded (this is related, but not identical, to the environmental unboundedness mentioned above). A single loop in a routine that is being indicated as possible may yield that that routine is competent to an unbounded class of possible interactions. Indications of further interactive processing, then, can be multiple, unbounded, and complex.

In particular, such indications are indications of *possibilities*, of *potentialities*. Interactive indication, and, therefore, interactive representation, is inherently *modal*. There is, in fact, a rich set of resources of modal representation and implicit definition in the interactive model (Bickhard, 1998b; Bickhard & Campbell, 1992;

Campbell & Bickhard, 1992b; Bickhard & Terveen, 1995). Such modality of representation will be of importance in later discussions of logic and necessity.

1.3.5. Evolutionary epistemology

Correspondences with the world might be impressed into a passive substrate—e.g., forms into a waxed slate. If representations are thought to be some version of such correspondences, then it is tempting to think that representations could be created or invoked by such impressions into a mind: point-in-time *trans*ductions into encodings, or scratching-the-wax over time *in*ductions into encodings. Note that the mind is epistemically passive in such models. Any activity or action is epistemically superfluous, even if biologically understandable. Not so in the interactive model.

Interactive representation is constituted in interactively competent system organization, not in correspondences. There is in this view no temptation to model the origins of representations in terms of impressions into an epistemically passive system.

All representation—indeed, all system organization—must be constructed. Such constructions, however, cannot be assured of being correct, so they must be tried out and selected out if inadequate. The interactive model of representation, thus, forces a variation and selection constructivism; it forces an evolutionary epistemology (Campbell, 1974; Hahlweg & Hooker, 1989; Radnitzky & Bartley, 1987; Wuketits, 1990).

1.3.6. Reflection

Issues about reflection merge inherently with those about consciousness and reflective consciousness. On the one hand, any issue about the *existence* of such reflection is settled by any reflection on the question. On the other hand, a naturalistic model is much more problematic.

I will not need and will not assume a full naturalistic model of reflective consciousness in the following discussion (even though such a model will be offered elsewhere: Bickhard, in preparation-a; see also Bickhard, 2000, in press-a, in press-b, forthcoming). What I will be making use of is a notion of epistemic reflection: representations of (properties of) representations.

This possibility derives readily from the interactive model of representation. Just as a system can interact with its environment and can represent various things about that environment, so also can a second level system interact with and represent things about a first level system. This intuition is developed into a notion of levels of knowing: at level one, the system interactively represents its environment; at level two, it interactively represents (properties of) level one; level three represents (properties of) level two; and so on. Such a hierarchy of levels of knowing is a hierarchy of *possibility*—(levels of) possibilities of things to be known—not a hierarchy of forced or inherent actual knowing systems. Most epistemic systems, in fact, are restricted to level one, and even human beings seem mostly to be limited to just a few levels (Campbell & Bickhard, 1986). But it is those (few) levels of reflection in human beings that I will be making use of in the discussion of rationality to follow.

1.3.7. Dynamics and rationality

The model of rationality that I am proposing centers around what I call *critical principles*—knowledge of error, negative knowledge, knowledge of principles that (if articulable) yield grounds for criticism. Internal knowledge of error—internal ability to (learn to) avoid error—constitutes an internalization of selection pressures, a beginning internalization of variation and selection processes more broadly.

For complex systems, such as human beings, that possess the properties discussed above, such as levels of knowing, this internalization of variation and selection processes can proceed complexly and unboundedly. I argue that there is an inherent tendency toward that internalization—and that it constitutes an inherent tendency toward the development of rationality.

2. Errors and error vicariants

The simplest form of error internalization occurs in systems that can learn to avoid error in their interactions with their environments. The central dynamic for such error internalization is a *destabilization* of interaction *readiness* if error is encountered. If some particular form of readiness for interaction becomes destabilized, then that form of readiness will tend to change in some way that may no longer produce those kinds of interactions, and, therefore, potentially no longer produce interaction error. That is, forms of interaction that tend to yield error will come to be avoided, via a variation and selection process occurring on the processes of setting up, constructing, forms of interactive readiness. This outline requires considerable filling out, but I will not focus on this level of primitive error dynamics in this paper (see Bickhard, in preparation-b). Instead, I will assume such simple internalizations of selection pressures, and proceed to consider more sophisticated error knowledge and its implications.

2.1. Epistemic error vicariants

I turn now to *representations* of error, epistemic surrogates or vicariants for error (Campbell, 1974), that permit us not only to avoid error but to think and reason about it. This involves both the dynamic interactive model of the nature of representation per se, and the model of levels of knowing, or levels of potential reflexive representation. In particular, potential errors at one level of knowing can be represented at the next higher level. That is, it becomes possible to *represent* possible constructive errors, not just to detect or avoid them.

I argue elsewhere that the possibility of reflection is the result of a very long macroevolutionary trajectory of increasing adaptability—increasing ability to anticipate and take into account temporal complexity in the environment. Human beings seem to be the current most advanced versions of this particular macroevolutionary trend; we are, above all else, adapted to niches of (that require) adaptability (Bickhard, 1980a,b, in preparation-a; Campbell & Bickhard, 1986; Bickhard & Campbell, forthcoming).

In particular, we are the inheritors of a nervous system with two innate levels of epistemic processes, one interacting with the external environment, and the second interacting with the first. These constitute the first two levels of the knowing levels hierarchy. They constitute an advance in the adaptation to adaptability in that they permit, for example, internal planning and internal anticipation of tentative interaction plans: the second knowing level can examine the first and make use of the interactive information resident there—information about the interactive potentialities of the environment and of the organism.

A species that has well developed second level knowing capabilities, in turn, is cognitively potentiated to develop full learned language—cultural language that can accumulate knowledge socially, not just genetically as in the language of, for example, social insects. The crucial emergent property is the ability to consider language meaning in itself, and not just respond to the immediate environmental implications of "language" as with social insects. Only with such a partial decoupling of language from immediate situation can complex language evolve (Bickhard, 1980b; Campbell & Bickhard, 1986).

One critical further potentiation is that language, in turn, makes possible the ascent through higher knowing levels—beyond the second—in a strictly functional manner without having to evolve still further physically distinct layers of central nervous system. Ascent from one knowing level to the next is essentially an instance of Piagetian reflective abstraction (Piaget, 1974, 1976, 1978, 1975, 1985, 1987, 1977, 2001; Campbell & Bickhard, 1986; Vuyk, 1981). The emergence of epistemic reflection and the emergence of language, then, potentiate and scaffold each other's further development.

2.1.1. *Reflective abstraction*

With the neural maturation of the second knowing level in human beings (at about age four; Bickhard, 1992a), developmental construction of interactive and representational organizations at that second level begins. Such second level processes can occur concomitantly with first level processes. Ascent beyond the second level involves externalization of indices of internal processes—usually in language—followed by abstraction from those indices of representations of properties of the processes that manifested those indices. A historical example is Aristotle's abstraction of the general forms of syllogisms: as abbreviations of terms became variables, the abstracted form of sentences and arguments emerged (Bochenski, 1970; Campbell & Bickhard, 1986; Kneale & Kneale, 1986). Such a process is reflective in that it involves consideration of and representation of lower level processes; it is abstractive in that there is no mirroring—no encoding—of lower levels into higher levels, but, rather, an abstraction of lower level properties in higher level representations. The epistemic relationship from higher to lower level is not that of an epistemically passive, higher level encoding perceiver of the lower level, but instead a higher level version of the first level's *interactive* epistemic relationship to the external environment.

2.1.2. Values

In a first level interactive system, some representations can function as goals in the sense that failure of the environment to satisfy them guides further interaction as interactive error: e.g., the system tries again, tries a different way, or in some other way tends to persist until the environmental goal conditions (differentiations) are satisfied (Bickhard, 1993; Bickhard & Terveen, 1995; Miller, Galanter, & Pribram, 1960). Other interactive subsystems may be evoked and regulated as being instrumental, or potentially instrumental, toward the satisfaction of such goals. That is, other subsystems can be evoked for the sake of changes they might make in the environment that are at least heuristically instrumental toward goal satisfaction.

Such functional goals can also occur at higher knowing levels. But at higher knowing levels the satisfaction conditions will be conditions of—properties of—lower level process and organization. Satisfying such higher level goals, then, will be constituted by the system itself (at lower levels) being or functioning in certain ways. Instrumental processes at higher knowing levels with respect to such goals will tend to change lower order system so that the system is or functions in those particular ways.

In a broad cognitive sense, such higher level goals constitute *values* about the lower level system. (I will not address emotional or motivational aspects of values here, nor will I address self-referential values that can be about the system as a whole; see Bickhard, 1997, 2000, in press-b; Campbell & Bickhard, 1986.) In their general ontology, such values are not differentiated by domain, though they may come to be so insofar as the system itself differentiates domains of being and functioning and develops values that are specific to those domains. For example, values might be about interpersonal interaction (a classic Kantian or utilitarian domain of ethics)—or about notions of good character and good life (a eudaimonistic domain of ethics)—or about conditions under which systems and representations are accepted as bases for further functioning (an epistemic domain)—or about how to best engage in attempting to accomplish other goals and values (a methodological domain). Both developmentally and historically, differentiation of such domains is itself a discovery of the basic evolutionary epistemological processes (Campbell & Bickhard, 1986; Shapere, 1984). One form that such domain differentiation could take would be the construction of a value (on system organization, functioning, and further construction) that *imposed* such a differentiation: that excluded organization, functioning and further construction that did not honor such a differentiation. We learn as children, for example, that considerations having to do with numbers of units—e.g., marbles—are independent of considerations have to do with the spatial distribution and arrangement of those units. Rearranging a bunch of marbles does not change their number.

Developmentally, domains often arise from the discovery of new kinds of error possibilities: exploration of related errors and of ways to avoid those errors can fill out a domain (Bickhard, 1992b, Campbell & Richie, 1983; Campbell & Bickhard, 1986). Developing a domain for number requires the development of criteria for what counts as error with respect to number.

3. Rationality and critical principles

3.1. The domain(s) of rationality

The ontology and development of domains is itself a complex subject (Bickhard, 1992b,c; Campbell, 1993; Campbell & Bickhard, 1986, 1992a; Campbell & Richie, 1983; Campbell & Christopher, 1996). Domains can differentiate out of prior domains; domains can combine (parts of) prior domains; domains can abstract aspects of other domains. I suggest that, as a domain, rationality is such an abstraction of an aspect of multiple other domains.

In particular, in some domains, there are inherent forms of error: attempting to walk across a chasm to get to the other side does not in general succeed. In some domains, there may be instrumental error in reaching or failing to reach goals, but the selection of goals per se is not subject to inherent error: you may or may not succeed in getting the flavor of ice cream you want, but the choice of ice cream flavor to seek is not generally subject to inherent error. In some domains, there are at best very general criteria of error, but, nevertheless, specific criteria of error are developed and invoked and changed, often for the sake of the pleasure involved in masterfully avoiding such errors. Many domains of esthetics are of this form. Historical and cultural and even individual forms of music, painting, and so on are not in themselves better or worse (except in the richness or lack of it that they permit), but they are necessary for esthetic appreciation and exploration.

In domains for which error is relevant, whether the error is inherent or created, there will in general be (at least heuristic) methodologies for avoiding those errors, for succeeding in whatever the domain tasks are—methodologies created by variation and selection processes with respect to those error criteria. There may also be meta-methodologies for *improving* those primary methodologies. Note that this makes methodological improvement with respect to a particular domain a kind of abstracted domain.

It happens to be the case (in our best wisdom) that issues of methodological improvement, although usually involving significant detail that is *specific* to particular domains, also involves issues and procedures that *overlap* those of other domains, and some issues and procedures that seem to be *universal*, or nearly so, to methodological improvement in all domains. Such issues and procedures will include criteria of evaluation—for success and failure—and heuristic procedures for satisfying those evaluative criteria. Criteria of symmetry are powerful sources of evaluation in contemporary foundational physics, for example, while criteria of logical validity are powerful sources of evaluation of reasoning methodologies in almost all domains. That is, the meta-methodologies of methodological evaluation and improvement themselves form an abstracted domain.

3.2. Rationality

The concept of rationality is applied ambiguously to several aspects of this (meta-) methodological domain. Rationality involves knowledge of how to do things right, and that necessarily involves knowledge of what would be in error. The concept of rationality, then, is applied to the processes, both interactive and constructive, that are guided by various error avoiding methodologies. And it is applied to the knowledge involved concerning what counts as success and error, and the heuristics of how to satisfy those criteria. Most deeply, I propose, it applies to the inherent tendency to develop and improve such processes, methodologies, and meta-methodologies.

The engine of development is variation and selection. In the context of knowing levels and cultural language, that will inherently tend to yield an internal version of variation and selection processes, an internalization of processes of evolutionary epistemology.[5] Internalized evolutionary epistemology, in turn, will tend to involve criteria and heuristics for interaction and construction—methodologies—and criteria and heuristics for the evaluation and further construction of such methodologies—meta-methodologies.[6] In this view, rationality is an intrinsic tendency, an intrinsic theme, of development. Derivatively, it is also the products of that developmental tendency. Rationality emerges as an explicit domain, both culturally and individually, insofar as values emerge that are general to meta-methodological concerns—values concerning knowledge, and the further development of knowledge, of what constitutes error and of how to avoid error.

3.2.1. *Unfolding*

Values at higher knowing levels can serve to constrain the functioning and the construction of the lower knowing level system. Interestingly, there is a converse sense in which lower level systems constrain the construction of higher level values. This constraint emerges intrinsically in the process of the construction of higher level interactive systems.

Construction at the first knowing level will be constrained by what will succeed in its interactions with the environment. System constructions will tend to be stable only insofar as they successfully anticipate and control their courses of interaction. In appropriate configurations, that constitutes successfully representing those environments.

For construction at higher knowing levels, it is the lower knowing levels that constitute the interactive environment. Congruent to the case of the first level and the external environment, higher knowing level constructions will tend to be those that

---

[5] This is an internal development of processes of evolutionary epistemology—a kind of process that continues to occur externally. Internal variation and selection processes emerge because of the advantages of adaptability that they offer the organism. They are not in any sense a bringing into the organism or impressing into the organism of something from outside. The internal emergence is parallel to, not instructed from, external processes—except, of course, via the selection effects of external processes. This is in contrast, for example, to the notions of internalization of Piaget and Vygotsky, which involve specific external structures and organizations being brought into the organism, and at least suggest an encoding of those structures and organizations (Bickhard, 1995).

[6] The ability to develop heuristic processes of variational construction imposes its own non-trivial requirements on underlying dynamics and architecture (Bickhard & Campbell, 1996).

successfully interact with lower level systems. In appropriate configurations, these will constitute representations of lower level properties.

This constraint will hold for the construction of higher level values as much as for any other kind of higher level system. With regard to values, the constraint manifests itself as a tendency for new higher level values to be values that are *already satisfied* in the lower level systems with respect to which they have been constructed. That is, higher level constructions will tend to be of properties, including values, that are already instantiated at lower levels, just as representations at level one will tend to be of properties that are already instantiated in the external environment. Higher level values, then, will tend to *unfold* values that are implicit in existing lower level system organizations and processes (Campbell & Bickhard, 1986).

Such unfolding is not the only theme in the construction of values. If it were, then values would serve little function beyond "reflecting" what was already present. In particular, conflict can be encountered, forcing further constructive accommodation. This can occur in at least two ways: (1) Two values unfolded at a particular knowing level may be inconsistent with each other. Each one might accurately unfold a value that is implicit in some part of the lower level, but each would then be inconsistent with the lower level base for the other: conflict, which was implicit, now becomes explicit. And (2) even if there is no second value, an unfolded value with respect to one lower level organization might conflict with, fail to be satisfied by, other lower level organizations. Encounter with such a conflict again introduces instability and forces some sort of accommodation (which, of course, is not necessarily a successful accommodation—or a rational one): (i) Such lower level conflicting systems might be taken as *counterexamples to* the higher level values, forcing change in the higher level values (which might, in turn, induce change in whatever they unfolded from in the first place). (ii) A conflicting lower level system organization might be taken as a *violation of* the higher level values, to be changed itself until it no longer constitutes such a violation. (iii) One or more higher level values might be changed such that its *boundaries of application* no longer include the prior conflict; and so on. Resolution of such conflicts is itself a proper subject for the development of rationality.

There is an intrinsic tendency for unfolding to encounter such conflicts and to eliminate them (again, perhaps badly from some broader perspective, but eliminate them nevertheless; see Bickhard & Christopher, 1994). Implicit conflicts may encounter resultant failures and destabilizations, but, as implicit conflicts unfold into *explicit* conflicts, the likelihood of destabilization and consequent change to reduce or eliminate the conflict increases. That is, there is an intrinsic tendency for development to increase the *consistency* of interactive, regulatory, constructive, representational, and evaluative processes. The intrinsic developmental tendency toward rationality, then, also involves an intrinsic developmental tendency toward consistency.

3.3. Critical principles

Values, insofar as they are associated with heuristics for the accomplishment of their satisfaction, can constitute *positive knowledge*—knowledge of how things should be, how they should function, and how they should be regulated, including how constructions should proceed. But there is an inherent asymmetry between such positive knowledge and the *negative knowledge* of what constitutes error. In particular, negative, or error, knowledge need not necessarily be associated with any knowledge of how to avoid that error. Negative knowledge can differentiate and represent error without necessarily providing *any* guidance to further process beyond that error feedback per se. This asymmetry holds for interactions, for constructions, and, in particular, for values.

Values can represent negative knowledge, knowledge of error, as well as positive knowledge. A child can know how to check for errors in addition (multiplication, etc.) before knowing how to add (e.g., count the union of the relevant sets of units—marbles). In fact, negative knowledge is a form of knowledge that requires less information than positive knowledge; it will tend to be an earlier aspect of knowledge construction. Knowledge of error precedes knowledge of how to avoid that error. The lack of positive heuristics for value satisfaction does not prevent negative knowledge from being functional in the system: it constitutes error vicariants or surrogates, and such surrogates can guide construction via internalized selections against constructions that lead into error (that violate the error criteria). Negative knowledge can guide (otherwise) blind variation and selection—yielding a minimalist evolutionary epistemology.

Negative knowledge values, when they can be articulated, constitute the grounds for *criticism*. They are the grounds for rejection or refutation. Correspondingly, I call them principles of criticism, or *critical principles*.

3.3.1. Hierarchies of critical principles

The crucial property that becomes possible with higher knowing levels is that higher level representations can represent lower level system properties. This holds for values, including negative knowledge values—critical principles—as well as for other representations. In particular, values can represent (properties of) lower level values, which might, in turn, represent still lower level values; the representational relationship can iterate up the knowing levels. Values, and critical principle values, then, can form hierarchies. In physics, for example, principles of symmetry or invariance represent and generalize particular forms of conservation, such as of momentum or charge (Kaku, 1993; Ryder, 1985; Sudbery, 1986).

Higher order values can *affirm* lower order ones, in the sense, for example, of providing a deeper or broader criterion by which the lower order value is correct. They can *infirm* lower order values, in the sense, for example, of representing the lower order value to be in error in some way. They can even do both simultaneously: for example, a higher order value might deepen the rationale for a lower order one, while simultaneously restricting the boundaries of its application. Symmetry principles affirm lower level physical conservations—up to the limits of various kinds of symmetry breaking.

3.4. Movement away from error

Constructions that ascend the hierarchy of possible values yield increasing knowledge, and knowledge about knowledge. Ascent through the hierarchy of critical principles constitutes increasing knowledge about error, including errors in what might be taken to be in error. Insofar as the system becomes capable of satisfying those ascending critical principles, then, it becomes increasingly able to avoid error. Ascent of the hierarchy of critical principles constitutes *movement away from error*. Such ascent constitutes progressive internalization of evolutionary epistemology, including of the values and critical principles—the internal vicariants—that make it possible.

This tendency toward movement away from error is intrinsic to the intrinsic tendency toward rationality. It constitutes a basic coherence in rationality: a reflexive sense in which rationality is rational by its own standards. Evolutionary epistemology is driven by error guidance. The internalization of evolutionary epistemology, including in particular the tendency toward rationality and away from error, enhances the ability of the system to avoid error, including most deeply the increasing ability to avoid errors of the evolutionary epistemological processes themselves.

3.4.1. The asymmetry between positive and negative knowledge—again

Positive knowledge is successful only insofar as it avoids relevant error. As new kinds of errors are discovered, old positive knowledge may no longer be acceptable—it might succeed in avoiding the errors represented by old critical principles, but fail to avoid newly discovered errors. Positive knowledge that *does* succeed in avoiding those new errors may or may not have any particular constructive relationships with the old positive knowledge: the new knowledge, when eventually constructed, might simply modify the old knowledge, but it might replace it with something entirely different. The ontology of the space-time of special relativity is ontologically fundamentally different from the space and time of Newtonian mechanics (Friedman, 1983; Longair, 1984; Misner, Thorne, & Wheeler, 1973; Torretti, 1983; Wald, 1984), and the ontology of caloric is nothing like that of the kinetic theory of heat (Harman, 1982). Positive knowledge can be highly unstable relative to the ascent up the rationality hierarchy of critical principles.

Critical principles, on the other hand, tend to build on earlier critical principles. This is not necessarily, however, a simple cumulation of critical principles, or even a simple cumulation of a hierarchy of critical principles. Critical principles can infirm, and even reject, earlier critical principles, but the *aboutness* relationship that moves up the hierarchy holds even in such cases of rejection: there is something specific that is being rejected. Often such hierarchical infirmations become sedimented in a culture and in the manner in which the relevant domain is taught. For example, Russell's paradox is learned after some understanding of naive set theory—that is, only after gaining some understanding of something that violates the criterion that Russell's paradox provides. There are occasions in which critical principles simply disappear, rather than being historically or developmentally sedimented. One example might be the criterion by which the discovery of a moon of Jupiter was rejected because there could not be other than exactly seven heavenly bodies since there were exactly seven orifices in the head (Hempel, 1966): the metaphysics concerning God's design of the universe, and the supposed constraints of coherence that that imposed on how the universe could possibly be, has now been rejected, and, along with it, the critical principles that presupposed it.

Nevertheless, the asymmetry strongly tends to hold. Ascent up the hierarchy of critical principles tends to be "progressive", even when it involves infirmations, in the sense of the tendency to move away from error. In contrast, positive knowledge can much more readily be strongly replaced and abandoned as new negative knowledge is discovered. Positive knowledge can be "progressive" in the sense of avoiding larger ranges of error and deeper errors, but it is often not *cumulative* in the sense of incorporating and building on old positive knowledge.

This asymmetry suggests a metaphor in which negative knowledge, critical principles, form the skeleton of rationality. The hierarchy is what positive knowledge is built on and around, and what positive knowledge must comply with. The overall system can do very little *without* such positive knowledge—like a skeleton with no musculature—but it can do nothing at all without negative knowledge, not even attempt to construct positive knowledge. Positive knowledge is positive knowledge only relative to negative knowledge; knowledge is fundamentally knowledge of how to avoid error.

4. Naturalism, formalism, and logic

4.1. Rationality and logic

The model of rationality just outlined does not look much like popular notions in which rationality is equated to logic. In particular, to be rational, in these views, is to in some form honor the formal rules of valid logical reasoning—a formalism of rationality. There is an at least prima facie appeal to formalism: logic does seem to be intimately involved in rationality, and simply equating the two has made sense to many throughout Western history. Logic, after all, has been considered to be the "laws of thought" (Boole, 1958).

It is incumbent on me, therefore, to compare the two approaches and to show why the critical principles model presented above should be preferred to formalism. I will address this comparison with respect to two issues: (1) formalism fails the problem of origins, while the critical principles model does not, and (2) the critical principles model is itself capable of accounting for logic.

4.2. Formalism and origins

Any formalistic approach begins with two classes of assumptions: (1) the basic representations, usually propositions, about which reasoning is to occur, and (2) the rules by which valid reasoning can occur (Brown, 1988). The rules permit the inference of new propositions on the basis of some set of initial propositions, preserving or cumulating (in the case of inductive logic) the warrant or truth of those initial propositions. If the process is initiated with propositions with high warrant—preferably with certainty—then the results are assured of being of as high a warrant as possible.

New propositions can be derived from initial propositions, and new rules can even be derived from initial rules, but, in both cases, there must be some foundation of rules, and, for any given instance, some foundation of propositions. Formalism, in other words, yields foundationalism. Formalism requires foundational propositions and foundational rules.

There have been millennia of attempts to make good on providing or accounting for those foundations. None work. I will not rehearse details of the many arguments involved, but formalism falls because it requires foundations—foundations of warranted representations and foundations of warranted rules—and there is no account of the origins of those warrants (Brown, 1988; Hooker, 1995).

There is, in fact, a second sense in which formalism requires an impossible foundationalism: a foundationalism of content. Both the presumed foundational propositions and the rules are representations, and must have representational content. Just as for warrant, the only way in which such content can be provided is in terms of some set of foundational representations with foundational content out of which all further content can be generated. But, if the only way to get new representational content is to begin with content already available, then it is not possible in principle to account for the foundational level of content. Similarly, if the only way to get new warrant is to begin with warrant already available, then it is not possible in principle to account for the foundational warrant.

Formalism, then, fails to account for the origin of warrant and for the origin of representational content. Foundationalisms, of all kinds, are intrinsically not naturalistic. They cannot naturally account for the origins of their own foundations.

4.3. *Critical principles and origins*

The problem of the naturalistic origins of critical principles does not encounter a foundationalist aporia (Bickhard, 1991b). The dynamics of constructive destabilization upon encountering error, in fact, is precisely an account of how and why critical principles could be expected to naturally emerge in systems with certain properties (Bickhard, in preparation-b). In particular, the internalization of processes of *evolutionary epistemology* can be expected for systems that are capable of such internalizing *constructions*, and, if they are also capable of ascent up the knowing levels, critical principles are part of an inherent tendency of development. Critical principles do not have to be constructed out of already available critical principles. The dynamic model accounts for their emergence out of prior forms of process via a kind of internalization of variation and selection processes. Critical principles, then, do not yield a foundationalism.

Similarly, the warrant for critical principles, and thus also for positive knowledge that satisfies them, is emergent in the avoidance of the (fallible) negative error knowledge that is constituted by those critical principles. Justification, or warrant, for critical principles is also not vitiated by a foundationalism.[7]

Finally, the representational content of critical principles, and for all interactive representation, is emergent in interactive system organization of certain kinds. There is no foundationalism, no aporia, of representational content.

The critical principles model, in other words, is consistent with a strict naturalism: it does not founder on the problem of origins. Formalism, in contrast, commits to foundationalism in multiple ways, and cannot escape from the impossibilities of origin of any of them. Formalism is not naturalistically possible; it fails as a natural model of rationality.

4.4. Critical principles and logic

If the critical principles model were inherently incapable of accounting for logic, it would at best be an incomplete model of rationality—and perhaps worse. Demonstrating how logic could be modeled within the framework of critical principles, therefore, is a crucial task.

The general manner in which that can be accomplished is to model the nature and emergence of various logical criteria—logical critical principles. The rules of logic constitute positive knowledge of how to avoid logical error. The essential aspect of logic to account for, then, is the negative knowledge of what constitutes logical error.

I will illustrate how that can be done, with a focus on the notion of validity. I will explicate validity in terms of some related concepts—logical necessity, in particular—and then model necessity as a critical principle. The model turns out to accommodate multiple kinds of necessity, and to suggest an intimate relationship between truth and necessity.

4.4.1. Validity

Valid reasoning preserves Truth value. In particular, a valid argument cannot begin with true propositions and yield false propositions. A valid argument form has no exceptions, no counterexamples—no *possible* exceptions or counterexamples—to its maintaining Truth value. Any particular argument that begins with true

[7] Broader questions can be raised at this point concerning the origins of any kind of normativity. In general, kinds of normativity emerge, often from lower level forms of normativity, but the possibility of such emergence, of the naturalization of normativity in this way, would seem to be precluded by the general impossibility of deriving norms from facts. There are two levels of response to such challenges, neither of which will be developed here: one is to show that the arguments for the impossibility of such naturalization, of deriving norms from facts, are themselves unsound (Bickhard, 1998b, in press-c); the other is to provide a model of such emergence(s). In the general interactive model, the normativity of biological function emerges in certain kinds of far from equilibrium systems; the normativity of representation emerges as a kind of biological function (of anticipation); and the normativity of rationality emerges in the domain of biological function and representation and the evolutionary epistemological reflections upon them (Bickhard 1993, 1998a,b, 1999, 2000, in press-c; Christensen & Bickhard, in press).

propositions and ends with false propositions is a counterexample to any claim, any hypothesis, that the form instantiated by that argument is a valid form. Such counterexamples are essentially selections against forms of argument. Valid forms of argument have no such counterexamples; they survive the selection pressures.

Valid forms of argument, then, *necessarily* preserve Truth value. There are no exceptions to that in the entire modal space of possible exceptions. If we could model logical necessity, then, we could model validity. More generally, validity is a modal notion—a type of necessity. I will outline how the critical principles model can account for logical validity by more generally outlining how it, and, in particular, the interactive model of representation upon which it is based, can account for modality.

4.4.2. Necessity

Necessity is a condition of having no counterexamples within a space of possible counterexamples. Even one such counterexample annihilates necessity. Necessity, then, is a critical principle; it is a way in which a relationship can fail—it can fail by having counterexamples. Necessity is the critical principle of having no such counterexamples.

The modeling aspect of this that might appear to be most problematic is the representation of the space of possible counterexamples. Given such a representation, say "A", it is not difficult to model a monitor for any counterexamples in that represented space—something with the power of a little bit of predicate logic will suffice: "There are no A's that have the property 'counterexample' ".

But the space of possible counterexamples is a space of possibility; its representation is a modal representation. So, validity leads to necessity, which, in turn, leads to modality more generally. How can spaces of possibilities be represented?

4.4.3. Modal representation

As outlined earlier, modal representation does not pose in-principle problems for interactive representation. Interactive representations are intrinsically representations of spaces of possible interactions; interactive representations are intrinsically modal. Modality is not something added on top of a more fundamental non-modal form of representation. Insofar as interactive representation is capable of *any* form of representation, it will represent modally.

This, of course, reverses the apparent problematic. Now the question is how *non*-modality could be represented—how *actuality* could be represented. There are two parts to the answer to this question within the interactive model. The first is to simply note that the outcome of an actual interactive representational interaction indicates an encounter with an *actual* instance of whatever the representation represents. The second is to note that the *extension* of a representation is a property of that representation, and, thus, in principle capable of being itself represented from a higher knowing level. Given representations of extensions, properties of those extensions can be represented, such as that they are unit sets or that they are nonempty. Actuality, then, is recovered by differentiation out of an initially undifferentiated modal form of representation.[8]

Given the possibility of representations of possibility, representation of necessity follows. Possibility, necessity, and actuality, in fact, must be progressively differentiated from each other in development through the knowing levels. Representation of necessity, in turn, permits representation of validity, which yields a critical principle whose satisfiers will be logically valid forms of reasoning.

4.4.4. Kinds of necessity

The representation of necessity here is as an absence of counterexamples in a specified class of possible counterexamples. This rather naturally yields notions of different kinds of necessities, with kinds of necessities varying with the kinds of counterexamples being considered. Thus we find not only logical necessity, but also physical necessity, legal necessity, existential necessity, and so on, each characterized as exceptionlessness in the relevant class of possible counterexamples, of possible exceptions.

4.4.5. Truth

Note that if the specified class of relevant possible counterexamples is the class of actualities,[9] then the condition of having no actual exceptions is an interesting candidate explication for Truth. In this view, truth and necessity are indeed intimately related.

---

[8] This differentiation, in fact, is the developmental progression in children. "Consider the following example from a child of 4 years and 11 months who was asked to indicate all the ways that a toy car could get from a point A to a point B in a room:

> Pie (4;11) "Show me all the ways one can go from A to B." *Straight ahead.* "Can you make another?" *No.* "Try it." *You could put the car in the garage* (he repeats the straight path). "But do another one." He describes a slight curved line. "And another." *No.* "There are only two to do?" *Yes.* "Why?" *Because there's only one car*. We set up the post [a post set on the floor in between A and B]. "Now, do it." *It's impossible, because there's a post, so we can't go to B, it would make an accident*. "Try." He makes a curved path. *I got around it*. "And another." He repeats the same curved path, but turns back at the post, having bypassed it, instead of going to B. "Another." A curve from A to B, bypassing the post at the right instead of left. *That's not the same*. "Are there others?" *No.* "When you go to school, you always take the same way?" *No.* "And from A to B? Always the same?" *Yes* (Piaget, 1987, p. 19) (from Bickhard, 1988, pp. 502–503).

Trying to sort out just what is going on in this example is non-trivial (Piaget, 1987). What is clear, however, is that, although notions of possibility, impossibility, and so on are not unknown to Pie, nevertheless he has the field of modality extremely confused and mixed up—undifferentiated. The course of development involves, among many other things, a progressive differentiation out of such initial beginnings.

[9] The notion of "actuality" is itself subject to variation. It acts, in fact, as a kind of dual to the various kinds of necessity: an actuality in this room; an actuality in this physical world; an actual physical possibility; an actuality in a model, or in some space of possible models, and so on. The notion of Truth climbs up these spaces dually to the notion of necessity descending them. Truth and necessity are dual notions, with Truth focusing on exceptionlessness with respect to one space of consideration and necessity focusing on spaces of possible variations in that space of consideration.

This notion of Truth yields a limit notion of "true representation" in the sense that, if a representation were to be continually revised so as to exclude exceptions, true representation would emerge as the limit—the asymptotic limit—of that process of revision. It is of interest to note that any assumption of the existence of such a limit, and certainly of its uniqueness, involves *additional* assumptions concerning the topology of the space in which the limiting process of exception-exclusion is taking place (cf. Groeneveld, 1994; Gupta & Belnap, 1993). This limit notion of true representation, therefore, makes connection with the classical notions of truth as unique correspondence, but only asymptotically, and only with the addition of further topological assumptions (assumptions that may not hold: Sher (1998/1999) argues that Truth is a multi-faceted concept, and does not have a single substantive character—with logical form [see below] being one of its facets). The classical correspondence notion of truth is an asymptotic ideal of interactive representation (Bickhard, 1980a; Bickhard & Richie, 1983).

(Note, however, that even if it is assumed that unique differentiation—correspondence—is ever attained, or could be attained, that does not yield a model of *representation*. The incoherence of assuming any representational content of what is on the other end of the correspondence remains.)

4.4.6. Logical consequence

I began the discussion of logic with the notion of validity. I will end the discussion by returning to it, but now in a broader and more careful perspective, making use of the conceptual tools introduced. In particular, I will return to validity in the form of valid, or logical, consequence.

Logical consequence is a kind of necessary consequence. As mentioned, it is consequence that is necessarily valid by virtue of the *form* of the reasoning, but "form", and, therefore, "formally necessary", have not yet been carefully defined. The conceptual apparatus for explicating the notion of formal necessity has already been introduced. In particular, "formal" is taken to refer to properties that hold across, are invariant with respect to, variations in possible extensions of representations—across "semantic" models (Sher, 1991, 1996a, b).[10] Formally necessary, then, is necessity independent of representational contents—invariant with respect to particular kinds of variations in those contents. It is only the logical form of the inference that remains invariant under such transformations: This is the manner in which the intuition that

[10] In particular, with respect to automorphisms of models. Note that automorphisms are not the only kind of possible variation of extensions: restriction of consideration to automorphisms yields a particular, very powerful, notion of the nature of logic (Sher, 1991). That restriction, in turn, follows from the necessary indeterminism of the particulars of (interactive) representational extensions: representation is fundamentally implicit, not explicit, so set theoretic structural properties, which are preserved by (structurally invariant under; structurally isomorphic with respect to) automorphisms, are the limit of what can be considered about extensions in general. (More will be represented, of course, about [elements of] the extensions of particular representations.)

logical consequence is necessary in terms of the *form* of inference can be modeled.[11]

Note that, because extensions of representations are higher level properties of those representations (Bickhard & Campbell, 1986), formal properties are second knowing level properties (Sher, 1991, 1996a, b): "some" = "nonempty extension"; "every" = "empty complement of extension"; "and" = "intersection of extensions"; and so on. Furthermore, extensions can only be represented implicitly, not explicitly (except in special cases, such as the explicit listing of members of a (finite) set as definitive of that set). To assume otherwise is to presuppose explicit knowledge of what is differentiated, therefore implicitly represented, by an interactive representation: such prescience is not possible, and to assume it is to assume precisely the representational knowledge which the model of representation is to account for—it is circular. (This is an instance of the assumption that an element in a representational encoding correspondence somehow announces that it *is* in such a correspondence *and* that it announces what it is in correspondence with. In this case, the assumption is that the announcement is somehow of everything that the correspondence *could* be with—the extension. Elsewhere, I call this assumption "the incoherence of encodingism"—Bickhard, 1993, 1996; Bickhard & Terveen, 1995.) This necessary implicitness of the extensions of representations, in turn, requires that logical formality cannot depend on either a full metaphysics of what might be in such extensions, nor on any particular system of representations ("terms") for elements in those extensions (Sher, 1996a, 1998/99, 1999a).[12]

Still further, the second order character of formal properties explains why children are relatively inept with such logical considerations (especially those involving logical negation, which involves some notion of a universal extension) until the advent of second knowing level. The manner in which formal properties involve invariances of *possible* extensions—a modal notion—illustrates one of the crucial ways in which modalities are differentiated by children in the course of development out of an initial lack of such differentiation among actuality, possibility, and necessity (Bickhard, 1988; Piaget, 1987).

The historical discovery, explication, and refinement of the details of characterizations of logic and logical consequence, and the fact that such creation of more and more careful characterizations are still occurring (Sher, 1991, 1996b, 1998/1999, 1999b), is further demonstration that logic is created within rationality, rather than rationality being subsumed under logic. Furthermore, this creative construction is itself, as mentioned above, not a- or ir-rational. Instead, it explores and shows how

[11] Note that this notion of "formal" is related to, but is not the same as, the formal approach to rationality. In particular, this notion of formal "simply" means *in terms of particular identifiable* (*formal*) *properties*, and has no assumptions or implications of foundationalism.

[12] Invariance under automorphisms of models prescinds both from a full metaphysics of possible elements of extensions, and from particularities of systems of representation. Logic abstracts properties that are invariant with respect to structures of extensions, ignoring the particularities of extensional elements and of representational systems (Sher, 1991, 1996a,b, 1998/1999, 1999a,b). Such structural invariance, however, does depend on, among other things, the notion of and criteria for identity of elements of those models—criteria of "entityness" and of identity.

to honor ever more finely differentiating critical principles concerning form, identity, language, consequence, model, sets, and so on. The history of logic is itself a complex and rich demonstration of the history of the growth of a hierarchy of critical principles and of demonstrations of how to satisfy them, or that they cannot be satisfied.

Logical consequence, then, involves a critical principle of exceptionlessness with respect to a space of possible formal exceptions, where formal properties are properties invariant across (structural variations in) possible extensions. Logical consequence is necessary validity in virtue of logical form.

4.4.7. Rationality and logic

Just as rationality emerges as an abstracted domain from inherent tendencies of development, so also will critical principles of various kinds of exceptionlessness emerge. Exceptionlessness is "just" the principle that further processes of exposure to possible selections will not in fact select against the candidate representation or procedure; exceptionlessness, thus necessity and validity, emerge as natural notions from an evolutionary epistemological internalization of evolutionary epistemological processes. The critical principle model, then, not only can account for the individual and historical emergence and development of logic, it shows how logic too is an inherent tendency of such development, not just a contingent accident.[13]

5. Philosophy of science

The critical principles model has thus far been presented with two primary aims: (1) to demonstrate how critical principles can be expected to evolve naturally given a species with a particular set of characteristics—characteristics characteristic of human beings—and (2) to demonstrate how a critical principles approach to rationality can incorporate the existence and function of logic. I will turn now to illustrating how critical principles can help understanding of certain aspects of

[13] The discussion has focused on recovering, modeling, logic from within the critical principles framework. One consequence of the general form of that model that I would like to point out (but will not develop here) is a natural solution to a fundamental epistemological problem. The subject matter of logic is proposed as the invariances that hold under automorphic shufflings within extensions of representations, that hold in the properties of and relationships among those implicitly defined extensions. I would suggest that the subject matter of mathematics can be understood as the domain of possibility, the domain of interactively implicitly defined possibilities, their properties, and their relationships when those extensions are associated with principles of unitization and individuation (making it intimately related to logic). This requires considerable development, but all that I want to point out here is that such implicitly defined properties and relations are *internally* related to the physical/functional processes that implicitly define them, and the processes of reflection do have causal access to those physical/functional processes, and thus to the internally related properties that they carry. This is unlike the standard Platonic model, for example, in which there is no causal access to the presumed subject matter of, say, numbers. In other words, via the *internally* related properties and relations of real physical/functional processes, the model offers a route toward a solution to the tension, or conflict, between semantics and epistemology that Benacerraf has pointed out (Benacerraf, 1996).

rationality that otherwise remain difficult or inexplicable. In particular, I will address three issues of epistemology from the philosophy of science:

- the rational function of truth and realism in science,
- the nature of progress in science, and
- the rationality of certain induction-like considerations.

5.1. Truth and realism in science

The truth of scientific theories, and the realism of the entities to which they refer, can pose some perplexing problems (Psillos, 2000). In particular, Laudan has pointed out that, given the history of apparently well established theories—and their presumed ontologies—being overturned and rejected in favor of later developments in science, we would seem to have a rather strong "negative induction" concerning the acceptance of the truth or realism of *any* theories in science. But, if it is not rational to believe the truth or realism of scientific theories, what rational role could truth and realism possibly play in science? One conclusion might be that they play *no* rational role, and that scientific rationality has to be understood in different terms (Laudan, 1977). This is strongly counter-intuitive, yet—unless the implications of the historical negative induction can somehow be blocked (Hooker, 1987)—it would seem to be strongly supported.

The critical principles model offers a direct transcendence of this apparent dilemma. The key recognition is that critical principles, as negative knowledge, knowledge of kinds of possible error, can be rationally applied—to a theory, for example—to fallibly check if that theory does in fact make that error, even if there could never be appropriate and sufficient warrant to believe that the theory does *not* make that error. That is, it can be rational to check if and how a theory might fail criteria of truth or realism, even if it would never be rational to believe that the theory is true or that its ontologies are real; it can rational and useful to attempt to find out how a theory *fails* to be true or real, even if never rational to believe that it *is* true or real. It does not have to be accepted that *anything* satisfies critical principles in order for it to be rational to apply those critical principles.

Particle physics of the 1960s and 1970s offers a brief historical illustration of this point. In the mid-1960s there were two rival theories that addressed the zoo of particles and their interactions and decays that had been observed: Reggeon theory and quark theory. From an instrumentalist problem solving perspective, there was nothing available on which to base a choice between the two. Reggeon theory did not yield an ontology, and few accepted that there might be any reality for quarks, so there was no difference in their accepted status from a realist perspective—they were both taken to be likely nothing more than instrumental, and they made the same instrumental predictions about particle interactions. But when realist questions *were* posed, it was found that quark theory yielded predictions about certain kinds of high energy interactions with nucleons that Reggeon theory did not. When relevant experiments were performed, those realist predictions were supported, and we now seldom hear of Reggeon theory (at least not in anything like that earlier form) (Dodd, 1984; Riordan, 1992).

Here is a case in which it was clearly rational and useful to pursue a critical principle of realism (truth of reference), even though it is not even now accepted that, say, quantum chromodynamics is a true and realistic theory. It was rational and useful to apply a critical principle of realism even though it did not yield a rationally certain belief in that realism.

This point about critical principles and truth and realism in science is "just" an application of the general asymmetry between "confirmation" and "falsification". But, until it is recognized that knowledge of *kinds* of falsifications or refutations—critical principles—is itself a crucial form of knowledge, it is difficult to make this extension of the asymmetry.

5.2. Progress

Scientific progress seems superficially to consist of the accumulation of more and more knowledge. Attempts to abstract principles of rationality in science have often presupposed this view: they tend to yield notions of rationality that are self-consistent—in which it is rational to be rational—because rational thought and action is supposed to move closer to the truth, or to yield better approximations to the truth, or to yield higher probabilities of the truth, and so on. Laudan's rejection of the rational role of truth or realism in science mentioned above was in part a reaction against the repeated failures of models of rationality that attempt to make such claims to support them convincingly. In addition to the multifarious logical and conceptual errors in such models, there is the straightforward question: if scientific rationality provides such guarantees of cumulative progress of science, then why do we find the negative historical induction? Why has science overthrown such wildly successful and supported theories?

A simple reaction against such major repudiations in the history of science, of course, is to posit an irrationality of some sort in the nature of science (Feyerabend, 1975; Kuhn, 1962; Lakatos & Musgrave, 1970). An unsustainable faith in the cumulative progress of science or a rejection of science as being fundamentally rational at all is not a palatable choice. Critical principles provide a transcendence of this dilemma too.

The key to this transcendence has already been introduced. Critical principle rationality is self-consistently progressive in the sense that it tends to move away from error, not in the sense that it provides some guarantee of (even a tendency toward) movement toward truth. In this view, science is not necessarily *cumulative* in its positive knowledge, but is *progressive* in that the positive knowledge, even when utterly rejecting previous positive knowledge, succeeds in avoiding more and deeper errors than previous positive knowledge. Without the recognition of the special nature and role of negative knowledge—critical principles—the only possible form of progressivity seems to be cumulativity, and positive cumulativity is contradicted by history.

A historical progression that illustrates this point can be found in a cumulative sequence of critical principles that have been involved in the history of physics. In Aristotle's physics, the laws of physics varied from place to place; the laws for the heavens, for example, were not the same as those for earth. The revolution of Copernicus, Galileo, and Newton introduced a principle of criticism of such notions: the heavenly laws and those on earth should be the same (Brown, 1988; Kuhn, 1957). The laws of physics should be place invariant. The Special Theory of Relativity maintained this criterion of place invariance, and added a criterion of velocity (first time derivative of place) invariance. The General Theory of Relativity, in turn, maintained both of these criteria, and added a criterion of acceleration (second time derivative) invariance (Friedman, 1983; Longair, 1984; Lucas & Hodgson, 1990; Misner, Thorne, & Wheeler, 1973; Torretti, 1983; Wald, 1984; Weinberg, 1972).[14]

The positive theories, and their underlying ontologies, were radically rejected in each of these moves, but the negative knowledge, the structure of critical principles, was strictly cumulative. It is the cumulativity of these critical principles, I suggest, that accounts for the progressivity of the respective theories. An example mentioned earlier—the rejection of metaphysical critical principles based on notions of God's design of the universe—illustrates critical principle progressivity even when there is an infirmation, not a confirmation or cumulation, between critical principles.

5.3. Induction

Induction has provided a scandalous situation for epistemology. On classical grounds, there seems to be no rationality to "inductive" inferences (Hume, 1978; Schacht, 1984; Stroud, 1977). In fact, induction fails on two fundamental grounds, both instances of failures to account for origins (that is, of an implicit anti-naturalism). The standard, and correct, view of the failure of induction is that it fails

[14] Invariance is one of the most important kinds of critical principle (Bickhard, 1980a; Hooker, 1992, 1995). Differing forms of invariance can be found with respect to the cognitions of physical objects; conservations, such as of number, mass, or volume; the differing kinds of invariance that generate the various kinds of geometry; and global and gauge invariances in theoretical physics. The invariances discussed with respect to the Theories of Relativity are examples of global invariances. Earlier, the domain of logic was construed as being constituted as a kind of invariance: invariance under isomorphic structural transformations of abstracted representational extensions (Lindenbaum & Tarski, 1934–1935, 1983; Mautner, 1946; Lindström, 1966a,b; Mostowski, 1957; Tarski, 1966/1986; see especially Sher, 1991, 1996a, b, 1998/99, 1999a). Invariance is a—perhaps the—primary form of differentiation from, independence of, the processes of an epistemic agent. The invariances of objects yield a stability in time with respect to most ensuing events, which, in turn, makes possible the representation of a relatively stable world transcending the immediate perceptual environment of the organism—your home remains relatively invariant, and can be represented as such, even when you are away. The invariances of physics prescind from particularities of the situation of observers and the origins, orientations, and time derivatives of measuring frames. Logic is the domain of properties that are invariant under isomorphic structural transformations of representational extensions—it is the domain generated by prescinding from the particularities of representational agents and their situations. Invariance is the general form of understanding and representing the world as (relatively) independent of the observer; invariance is agent-decentering.

to provide *warrant* in any of the ways that have been proposed for it (Lakatos, 1968; Popper, 1959, 1985).

I will not rehearse these very familiar points, but will simply point out one additional failure of induction to account for origins: its failure to account for the origins of the representational *content* of inductive inferences. Not only do transductions and inductions fail to provide warrant for the impressions and scratchings into the waxed slate (sensory receptors, memory banks), they fail to provide any representational content for those impressions and scratchings about which there might even be any issues of warrant (Bickhard, 1993). Popper pointed out that what is usually called induction—a cumulation of warrant from cumulations of positive instances of some relationship—in fact requires that the representation of the relationship (the hypothesis) be already present in order to notice even the first such positive instance. That is, the cumulation of positive instances does not provide the representation of what they are instances of. But, if that is the case, then what is involved is hypothesis testing, not induction (Popper, 1959, 1965, 1972, 1985).

But a puzzle does remain. We do generally grant increasing warrant in most cases of the cumulation of positive instances. Is this simply irrational? And, if not, in what sense is it rational?

Again, I wish to suggest that the critical principles model offers an escape. The cumulation of evidence for a theory or hypothesis offers increasing rational warrant *insofar as* that evidence excludes more and more errors which that theory or hypothesis—or those tests of the theory or hypothesis—might be committing. There will generally be many kinds of error—critical principles—that might be involved, and, as such kinds of errors are themselves rationally tested and excluded, the rationality of accepting that the theory or hypothesis avoids such errors increases. The increase of warrant from the cumulation of evidence, then, is an increase of warrant for accepting the relevant forms of error avoidance.

Cumulation of evidence is fallible but rational evidence that the considered theory avoids the tested forms of error, and, therefore, is to be rationally preferred to any otherwise equivalent alternative that commits any of those errors. That is, the warrant is not warrant for belief, but instead it is warrant for comparative theory evaluation and selection (cf. corroboration or verisimilitude, Niiniluoto, 1985; Popper, 1959, 1965, 1972, 1985). The sense in which all rational warrant is comparative rather than absolute follows readily from the construal of rationality in terms of movement away from error. Positive knowledge is always rationally accepted relative to the current state of negative knowledge, and relative to alternative candidates for positive knowledge. No form of absolute acceptance is supported by the critical principles model, and seems contradicted by the historical "negative induction" of the eventual overthrow of positive knowledge (Campbell, 1974, 1988).

This view not only explicates the rationality of weight-of-evidence warrant per se, but it also makes sense of characteristics that can otherwise seem intractable. First of all, not all evidence is equal. Evidence that rules out rationally more important alternatives is more important, gives more warrant, than that which rules out less important or trivial alternatives. For example, repeating the same experiment once may serve to rule out the moderately important alternative that the first time was a fluke or simple mistake in some way, but repeating it over and over again gives less and less marginal warrant because the alternatives that are thereby ruled out are more and more bizarre or trivial. For another example, evidence that rules out a possible methodological error in an earlier study gives less warrant than evidence that rules out a theoretical alternative. Movement away from error is structured, structured by the partial ordering of the critical principles hierarchy, and is not subject to any unitization or measure. Without such a structure of alternatives and critical principles that apply to them, there is no non-ad hoc way to weight the epistemic warrant of evidence.

A second characteristic that emerges naturally from this view is that the rational warrant for a theory can change *without any change in direct evidence at all*. If a new theoretical alternative is discovered that has not been considered before, and if it itself is not infirmed or eliminated by current principles and evidence, and, even more, if it is satisfying of critical principles that do not apply to the earlier theory, then the sense of warrant for that earlier theory, will, so long as this situation holds, be rationally diminished from before. The new alternative rationally indicates new ways in which the old alternative might be incomplete or wrong. Conversely, the elimination of an old alternative, even if by evidence or considerations that do not apply to a given alternative, will increase the warrant of alternatives remaining (Campbell, 1974, 1988; Laudan, 1981, 1993, 1996). Such characteristics are intrinsically mysterious from the perspective of any model of rationality as the increase of truth content or the accumulation of support.

6. Conclusion

Naturalism provides a powerful set of critical principles. The problem of origins, to mention but one, eliminates most contemporary models of representation, and of rationality. Naturalistic models of function arguably require open (interactive) systems dynamics: function is a natural emergent only in far from equilibrium open systems. Issues of the anticipatory regulation of interactive processes in such systems, in turn, yield the interactive model of representation—and interactivism forces a constructivism of variation and selective retention, an evolutionary epistemology.

Jumping over a rather long macro-evolutionary trajectory to a species that is capable of reflection and language, we find an inherent developmental tendency toward the internalization of variation and selection processes—vicariant evolutionary epistemology. This essay has focused on the necessary negative knowledge aspect of that development—the development of constructive error surrogates, or critical principles—and has argued that the developmental tendency in general, and its products in particular, constitute the framework of rationality. Rationality is the domain of getting better at the avoidance of error. The claim to naturalism of this model is prima facie strong by virtue of its derivation from more general natural dynamic considerations.

No viable model of rationality can be incompatible with the role of logic in rational thought, in spite of the fundamental naturalistic inadequacy of formalist approaches, which *equate* rationality with logic. The critical principles model suggests an inherent tendency to develop critical principles about the formal properties of reasoning, such as validity. Positive knowledge of how to satisfy those critical principles, in turn, is what we identify as the formal rules of logic. The path to this derivation of the domain of logic is via the intrinsic modality of interactive representation. Logic, then, emerges quite naturalistically.

Finally, the fundamental asymmetry between positive and negative knowledge is exploited to explain several otherwise inexplicable phenomena in the philosophy of science. In particular, truth and realism are argued to have rational roles to play in science as critical principles, in spite of there being little rational reason to actually believe truth or realism of any particular theories. Progress in science is modeled in terms of the movement-away-from-error tendency of the critical principles hierarchy, rather than in terms of the cumulation of positive knowledge—which tends to be periodically overthrown, not cumulated. And the rationality of seemingly inductive practices is explained in terms of the structured exclusion of possible errors by cumulating evidence.

Rationality does not look like a good candidate for a counterexample to naturalism. But to accomplish a naturalism with respect to rationality does require abandoning many contemporary frameworks in the philosophy of mind. Among others, it requires rejecting encodingist models of representation, and formalist or foundationalist models of epistemology and rationality. Process models of an evolutionary epistemology of open systems dynamics offer an alternative metaphysics to the atomistic foundationalisms of encodingism and formalism.

Acknowledgements

Thanks are due to the Henry R. Luce Foundation for support during the preparation of this paper and to Cliff Hooker and Alex Levine for very useful comments on earlier drafts.

References

Annas, J., & Barnes, J. (1985). *The modes of scepticism*. Cambridge: Cambridge University Press.
Barnes, J. (1990). *The toils of skepticism*. Cambridge: Cambridge University Press.
Benacerraf, P. (1996). Mathematical truth. In W. D. Hart (Ed.), *The philosophy of mathematics* (pp. 14–30). Oxford: Oxford University Press.
Bickhard, M. H. (1979). On necessary and specific capabilities in evolution and development. *Human Development*, *22*, 217–224.

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

Bickhard, M. H. (1980b). A model of developmental and psychological processes. *Genetic Psychology Monographs*, *102*, 61–116.

Bickhard, M. H. (1987). The social nature of the functional nature of language. In M. Hickmann (Ed.), *Social and functional approaches to language and thought* (pp. 39–65). New York: Academic.

Bickhard, M. H. (1988). The necessity of possibility and necessity: Review of Piaget's possibility and necessity. *Harvard Educational Review*, *58*(4), 502–507.

Bickhard, M. H. (1991a). The import of Fodor's anticonstructivist arguments. In L. Steffe (Ed.), *Epistemological foundations of mathematical experience* (pp. 14–25). New York: Springer.

Bickhard, M. H. (1991b). A pre-logical model of rationality. In L. Steffe (Ed.), *Epistemological foundations of mathematical experience* (pp. 68–77). New York: Springer.

Bickhard, M. H. (1991c). Homuncular innatism is incoherent: A reply to Jackendoff. *The Genetic Epistemologist*, *19*(3), 5.

Bickhard, M. H. (1992a). Commentary on the age 4 transition. *Human Development*, *35*(3), 182–192.

Bickhard, M. H. (1992b). Scaffolding and self scaffolding: Central aspects of development. In L. T. Winegar, & J. Valsiner (Eds.), *Children's development within social context: Research and methodology* (pp. 33–52). Mahwah, NJ: Erlbaum.

Bickhard, M. H. (1992c). How does the environment affect the person? In L. T. Winegar, & J. Valsiner (Eds.), *Children's development within social context: Metatheory and theory* (pp. 63–92). Mahwah, NJ: Erlbaum.

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

Bickhard, M. H. (1995). Intrinsic constraints on language: Grammar and hermeneutics. *Journal of Pragmatics*, *23*, 541–554.

Bickhard, M. H. (1996). Troubles with computationalism. In W. O'Donohue, & R. F. Kitchener (Eds.), *The philosophy of psychology* (pp. 173–183). London: Sage.

Bickhard, M. H. (1997). Is cognition an autonomous subsystem? In S. O'Nuallain, P. McKevitt, & E. MacAogain (Eds.), *Two sciences of mind* (pp. 115–131). Amsterdam: John Benjamins.

Bickhard, M. H. (1998a). Levels of representationality. *Journal of Experimental and Theoretical Artificial Intelligence*, *10*(2), 179–215.

Bickhard, M. H. (1998b). A process model of the Emergence of Representation. In G. L. Farre, T., & Oksala (Eds.), *Emergence, complexity, hierarchy, organization. Selected and Edited Papers from the ECHO III Conference. Acta Polytechnica Scandinavica, Mathematics, Computing and Management in Engineering Series* No. 91, Espoo, Finland, August 3–7, pp. 263–270.

Bickhard, M. H. (1999). Interaction and representation. *Theory & Psychology*, *9*(4), 435–458.

Bickhard, M. H. (2000). Motivation and emotion: An interactive process model. In R. D. Ellis, & N. Newton (Eds.), *The Caldron of consciousness* (pp. 161–178). Reading, MA: J. Benjamins.

Bickhard, M. H. (in preparation). Autonomy, function, and representation. *Communication and Cognition*.

Bickhard, M. H. (in preparation-a). *The whole person: Toward a naturalism of persons*.

Bickhard, M. H. (in preparation-b). Error dynamics: The dynamic emergence of error avoidance and error vicariants.

Bickhard, M. H. (in press-a). The emergence of contentful experience. In T. Kitamura (Ed.), *What should be computed to model brain functioning?* Singapore: World Scientific.

Bickhard, M. H. (in press-b). Function, anticipation, representation. In D. Dubois (Ed.), *Fourth international conference on computing anticipatory systems*. American Institute of Physics.

Bickhard, M. H. (in press-c). The dynamic emergence of representation. In H. Clapin, P. Staines, & P. Slezak (Eds.), *Representation in mind: New approaches to mental representation*. New york: Praeger.

Bickhard, M. H., & Campbell, D. T. (2000). Emergence. In P. B. Andersen, C. Emmeche, N. O. Finnemann, & P. V. Christiansen (Eds.), *Downward causation* (pp. 322–348). Aarhus, Denmark: University of Aarhus Press.

Bickhard, M. H., & Campbell, D. T. (forthcoming). Variations in variation and selection: The ubiquity of the variation-and-selective-retention ratchet in emergent organizational complexity.

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. *Journal of Pragmatics*, *17*(5/6), 401–433.

Bickhard, M. H., & Campbell, R. L. (1996). Topologies of learning and development. *New Ideas in Psychology*, *14*(2), 111–156.

Bickhard, M. H., & Christopher, J. C. (1994). The influence of early experience on personality development. *New Ideas in Psychology*, *12*(3), 229–252.

Bickhard, M. H., & Richie, D. M. (1983). *On the nature of representation: A case study of James J. Gibson's theory of perception*. New York: Praeger.

Bickhard, M. H., & Terveen, L. (1995). *Foundational issues in artificial intelligence and cognitive science—impasse and solution*. Amsterdam: Elsevier Scientific.

Bochenski, I. M. (1970). *A history of formal logic*. New York: Chelsea.

Boole, G. (1958/1854). *An investigation of the laws of thought*. New York: Dover.

Brown, H. I. (1988). *Rationality*. London: Routledge.

Burnyeat, M. (1983). *The skeptical tradition*. Berkeley: University of California Press.

Campbell, D. T. (1974). Evolutionary epistemology. In P. A. Schilpp (Ed.), *The philosophy of Karl Popper*. LaSalle, IL: Open Court.

Campbell, D. T. (1988). Science's social system of validity-enhancing collective belief change and the problems of the social sciences. In E. S. Overman (Ed.), *Methodology and epistemology for social science* (pp. 504–523). Chicago: University of Chicago press.

Campbell, R. J., & Bickhard, M. H. (in preparation). Physicalism, emergence, and downward causation.

Campbell, R. L. (1993). Epistemological problems for Neo-Piagetians. In the architecture and dynamics of developing mind. *Monographs of the Society for Research in Child Development*, *58*(5–6), 168–191.

Campbell, R. L., & Bickhard, M. H. (1986). *Knowing levels and developmental stages*. Basel: Karger.

Campbell, R. L., & Bickhard, M. H. (1992a). Types of constraints on development: An interactivist approach. *Developmental Review*, *12*(3), 311–338.

Campbell, R. L., & Bickhard, M. H. (1992b). Clearing the ground: Foundational questions once again. *Journal of Pragmatics*, *17*(5/6), 557–602.

Campbell, R. L., & Christopher, J. C. (1996). Moral development theory: A critique of its Kantian presuppositions. *Developmental Review*, *16*(1), 1–47.

Campbell, R. L., & Richie, D. M. (1983). Problems in the theory of developmental sequences: Prerequisites and precursors. *Human Development*, *26*, 156–172.

Chalmers, D. J. (1996). *The conscious mind*. Oxford: Oxford University Press.

Christensen, W.D., & Bickhard, M. H. (in press). The process dynamics of normative function. *Monist*.

Clark, A. (1997). *Being there*. Cambridge, MA: MIT/Bradford.

Cummins, R. (1996). *Representations, targets, and attitudes*. Cambridge, MA: MIT.

Dodd, J. E. (1984). *The ideas of particle physics*. Cambridge: Cambridge University Press.

Dretske, F. I. (1981). *Knowledge and the flow of information*. Cambridge, MA: MIT.

Dretske, F. I. (1988). *Explaining behavior*. Cambridge, MA: MIT.

Feyerabend, P. (1975). *Against method*. London: NLB.

Fodor, J. A. (1987). *Psychosemantics*. Cambridge, MA: MIT Press.

Fodor, J. A. (1990a). *A theory of content*. Cambridge, MA: MIT Press.

Fodor, J. A. (1990b). Information and representation. In P. P. Hanson (Ed.), *Information, language, and cognition* (pp. 175–190). Oxford: Oxford University Press.

Fodor, J. A. (1998). *Concepts: Where cognitive science went wrong*. Oxford: Oxford University Press.

Friedman, M. (1983). *Foundations of space-time theories*. Princeton: Princeton University Press.

Geach, P. (1977). *The virtues*. Cambridge: Cambridge University Press.

Godfrey-Smith, P. (1994). A modern history theory of functions. *Nous*, *28*(3), 344–362.

Groarke, L. (1990). *Greek scepticism*. Montreal: McGill-Queens.

Groeneveld, W. (1994). Dynamic semantics and circular propositions. *Journal of Philosophical Logic*, *23*(3), 267–306.

Gupta, A., & Belnap, N. (1993). *The revision theory of truth*. Cambridge, MA: MIT.

Hahlweg, K., & Hooker, C. A. (1989). *Issues in evolutionary epistemology*. Albany, NY: SUNY Press.

Hanson, P. P. (1990). *Information, language, and cognition*. Oxford: Oxford University Press.
Harman, P. M. (1982). *Energy, force, and matter*. Cambridge: Cambridge University Press.
Hempel, C. G. (1966). *Philosophy of natural science*. Englewood Cliffs, NJ: Prentice-Hall.
Hooker, C. A. (1987). *A realistic theory of science*. Albany, NY: SUNY Press.
Hooker, C. A. (1992). Physical intelligibility, projection, objectivity and completeness: The divergent ideals of Bohr and Einstein. *British Journal for the Philosophy of Science*, *42*, 491–511.
Hooker, C. A. (1995). *Reason, regulation, and realism: Towards a regulatory systems theory of reason and evolutionary epistemology*. Albany, NY: SUNY.
Hookway, C. (1985). *Peirce*. London: Routledge.
Hookway, C. (1992). *Scepticism*. London: Routledge.
Horgan, T. (1993). From supervenience to superdupervenience: Meeting the demands of a material world. *Mind*, *102*(408), 555–586.
Hume, D. (1978). *A treatise of human nature*. Oxford: Oxford University press.
Joas, H. (1993). American pragmatism and German thought: A history of misunderstandings. In H. Joas (Ed.), *Pragmatism and social theory* (pp. 94–121). Chicago: University of Chicago Press.
Kaku, M. (1993). *Quantum field theory*. Oxford: University of Chicago Press.
Kneale, W., & Kneale, M. (1986). *The development of logic*. Oxford: Clarendon.
Kuhn, T. S. (1957). *The copernican revolution*. Cambridge: Harvard University Press.
Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago: University of Chicago Press.
Lakatos, I. (1968). *The problem of inductive logic*. Amsterdam: North Holland.
Lakatos, I., & Musgrave, A. (Eds.). (1970). *Criticism and the growth of knowledge*. Cambridge: Cambridge University Press.
Laudan, L. (1977). *Progress and its problems*. Berkeley: University of California Press.
Laudan, L. (1981). A problem-solving approach to scientific progress. In I. Hacking (Ed.), *Scientific revolutions* (pp. 144–155). Oxford: Oxford University Press.
Laudan, L. (1993). Comparative theory appraisal. Manuscript.
Laudan, L. (1996). *Beyond positivism and relativism*. Boulder, CO: Westview.
Levine, A., & Bickhard, M. H. (1999). Concepts: Where fodor went wrong. *Philosophical Psychology*, *12*(1), 5–23.
Lindenbaum, A., & Tarski, A. (1983). On the limitations of the means of expression of deductive theories. In A. Tarski (Ed.), *Logic, semantics, metamathematics* (pp. 384–392). London: Oxford.
Lindström, P. (1966a). First order predicate logic with generalized quantifiers. *Theoria*, *32*, 186–195.
Lindström, P. (1966b). On relations between structures. *Theoria*, *32*, 172–185.
Loewer, B., & Rey, G. (1991). *Meaning in mind: Fodor and his critics*. Oxford: Blackwell.
Longair, M. S. (1984). *Theoretical concepts in physics*. Cambridge: Cambridge University Press.
Mautner, F. I. (1946). An extension of Klein's Erlanger Program: Logic as invariant-theory. *American Journal of Mathematics*, *68*, 345–384.
McGinn, C. (1993). *Problems in philosophy: The limits of inquiry*. Oxford: Blackwell.
Miller, G. A., Galanter, E., & &Pribram, K. H. (1960). *Plans and the structure of behavior*. New York: Holt, Reinhart, and Winston.
Millikan, R. G. (1984). *Language, thought, and other biological categories*. Cambridge, MA: MIT.
Millikan, R. G. (1993). *White queen psychology and other essays for alice*. Cambridge, MA: MIT.
Misner, C. W., Thorne, K. S., & Wheeler, J. A. (1973). *Gravitation*. San Francisco: Freeman.
Mostowski, A. (1957). On a generalization of quantifiers. *Fundamenta Mathematicae*, *44*, 12–36.
Mounce, H. O. (1997). *The two pragmatisms*. London: Routledge.
Nagel, T. (1986). *The view from nowhere*. Oxford: Oxford University Press.
Niiniluoto, I. (1985). The significance of verisimilitude. In P. Asquith, & P. Kitcher, *PSA 1984*, Vol. 2. East Lansing, Michigan: Philosophy of Science Association.
O'Conner, T. (1994). Emergent properties. *American Philosophical Quarterly*, *31*(2), 91–104.
Piaget, J. (1976). *The grasp of consciousness* (S. Wedgwood, Trans.). Cambridge, MA: Harvard University Press (original work published 1974).
Piaget, J. (1978). *Success and understanding* (A. Pomerans, Trans.). Cambridge, MA: Harvard University Press (original work published 1974).

Piaget, J. (1985). *The equilibration of cognitive structures: The central problem of intellectual development* (T. Brown, & K. Thampy, Trans.). Chicago: University of Chicago Press (originally published 1975).

Piaget, J. (1987). *Possibility and necessity*, Vols. 1 and 2 (H. Feider, Trans.). Minneapolis: University of Minnesota Press (original work published 1983).

Piaget, J. (2001). *Studies in reflecting abstraction* (Campbell, Ed. & Trans.). Philadelphia, PA: Taylor & Francis (original work published 1977).

Popkin, R. H. (1979). *The history of scepticism*. Berkeley: University of California Press.

Popper, K. (1959). *The logic of scientific discovery*. New York: Harper.

Popper, K. (1965). *Conjectures and refutations*. New York: Harper.

Popper, K. (1972). *Objective knowledge*. London: Oxford Press.

Popper, K. (1985). The problem of induction. In D. Miller (Ed.), *Popper selections*. Princeton: Princeton University Press.

Psillos, S. (2000). The present state of the scientific realism debate. *The British Journal for the Philosophy of Science*, *51*, 705–728.

Radnitzky, G., & Bartley, W. W. (1987). *Evolutionary epistemology, theory of rationality, and the sociology of knowledge*. La Salle: Open Court.

Rescher, N. (1980). *Scepticism*. Totowa, NJ: Rowman and Littlefield.

Riordan, M. (1992). The discovery of quarks. *Science*, *256*(5061), 1287–1293.

Rosenthal, S. B. (1983). Meaning as habit: Some systematic implications of Peirce's pragmatism. In E. Freeman (Ed.), *The relevance of Charles Peirce* (pp. 312–327). La Salle, IL: Open Court.

Ryder, L. H. (1985). *Quantum field theory*. Cambridge: Cambridge University Press.

Schacht, R. (1984). *Classical modern philosophers*. London: Routledge.

Shapere, D. (1984). *Reason and the search for knowledge*. Dordrecht: Reidel.

Sher, G. Y. (1991). *The bounds of logic*. Cambridge, MA: MIT.

Sher, G. Y. (1996a). Did Tarski commit "Tarski's Fallacy"? *The Journal of Symbolic Logic*, *61*(2), 653–686.

Sher, G. Y. (1996b). Semantics and logic. In S. Lappin (Ed.), *The handbook of contemporary semantic theory* (pp. 511–537). Oxford: Blackwell.

Sher, G. Y. (1998/99). On the possibility of a substantive theory of truth. *Synthese*, *117*, 133–172.

Sher, G. Y. (1999a). What is Tarski's theory of truth? *Topoi*, *18*, 149–166.

Sher, G. Y. (1999b). Is logic a theory of the obvious? *The Nature of Logic*. *European Review of Philosophy*, *4*, 207–238.

Smith, B. C. (1987). *The correspondence continuum*. Stanford, CA: Center for the Study of Language and Information, CSLI-87-71.

Stroud, B. (1977). *Hume*. London: Routledge.

Sudbery, A. (1986). *Quantum mechanics and the particles of nature*. Cambridge: Cambridge University Press.

Tarski, A. (1986). What are logical notions? Text of a 1966 lecture. In J. Corcoran (Ed.), *History and philosophy of logic*, Vol. 7 (pp. 143–154).

Torretti, R. (1983). *Relativity and geometry*. Elmsford, NY: Pergamon Press.

Vuyk, R. (1981). *Piaget's genetic epistemology 1965–1980*, Vols. I and II. New York: Academic.

Wald, R. M. (1984). *General relativity*. Chicago: University of Chicago Press.

Weinberg, S. (1972). *Gravitation and cosmology*. New York: Wiley.

Wimsatt, W. C. (1972). Teleology and the logical structure of function statements. *Studies in the History and Philosophy of Science*, *3*, 1–80.

Wuketits, F. M. (1990). *Evolutionary epistemology and its implications*. Albany, NY: SUNY Press.