Ethic Theory Moral Prac (2009) 12:25–51
DOI 10.1007/s10677-008-9142-6

What Does the Frame Problem Tell us About Moral Normativity?

Terry Horgan · Mark Timmons

Accepted: 11 November 2008 / Published online: 10 December 2008

Abstract Within cognitive science, mental processing is often construed as computation over mental representations—i.e., as the manipulation and transformation of mental representations in accordance with rules of the kind expressible in the form of a computer program. This foundational approach has encountered a long-standing, persistently recalcitrant, problem often called the frame problem; it is sometimes called the relevance problem. In this paper we describe the frame problem and certain of its apparent morals concerning human cognition, and we argue that these morals have significant import regarding both the nature of moral normativity and the human capacity for mastering moral normativity. The morals of the frame problem bode well, we argue, for the claim that moral normativity is not fully systematizable by exceptionless general principles, and for the correlative claim that such systematizability is not required in order for humans to master moral normativity.

Keywords Frame problem · Relevance problem · Computational cognitive science · Particularism · Generalism · Moral normativity

Within cognitive science, one important approach to modeling human mental processes has been to construe such processes as computation over mental representations—i.e., as the manipulation and transformation of mental representations in accordance with rules of the kind expressible in the form of a computer program. This foundational approach goes by various labels—e.g., the rules and representations approach, the computational theory of mind, classical cognitive science (or classicism). There is a long-standing, persistently recalcitrant, problem that has been encountered within classicism. This conundrum is often called the frame problem; it is sometimes called the relevance problem.[1] In this paper we describe the frame problem and certain of its apparent morals concerning human cognition,

[1]The more specific difficulty originally called the frame problem in early logic-based artificial intelligence is described by Shanahan (2004) as "the challenge of representing the effects of action in logic without having to represent explicitly a large number of intuitively obvious non-effects" (p. 1). This was just one instance of the generic problem that nowadays is usually called the frame problem.

T. Horgan (✉) · M. Timmons
University of Arizona, Tucson, AZ, USA
e-mail: thorgan@email.arizona.edu

M. Timmons
e-mail: mtimmons@email.arizona.edu

and we argue that these morals have significant import regarding both the nature of moral normativity and the human capacity for mastering moral normativity.

Ethical theorists have sometimes sought to discover a set of exceptionless general moral principles that would fully systematize moral normativity—that is, the principles would connect descriptive features to moral-normative features in such a way that for any potential circumstance C, all true moral-normative statements about C (e.g., statements about what would be morally just in C, what would be morally obligatory in C, what would be morally wrong in C, etc.) are logically entailed by a conjunction of (i) a sufficiently detailed characterization of C in descriptive, non-normative, terms and (ii) the sought-for principles.[2] The morals of the frame problem bode well, we will argue, for the claim that moral normativity is not fully systematizable by exceptionless general principles, and for the correlative claim that such systematizability is not required in order for humans to master moral normativity.

These two claims constitute an important element of the recently influential "particularist" approach to moral normativity—an element we call *core particularism*. We will distinguish core particularism from other, more expansive, versions of particularism that incorporate various further negative claims about general moral principles. We will explain why the morals of the frame problem that we describe in the first part of the paper have no obvious import, one way or the other, with respect to these additional negative claims sometimes espoused by particularists.

The primary considerations from cognitive science that we will be appealing to here are not ones involving the empirical-psychological investigation of moral-judgment formation per se; that we have done elsewhere (Horgan and Timmons 2007). Rather, we mainly will be drawing upon an ongoing foundational discussion/debate in cognitive science, one that resides at the interface of empirical science and philosophy (rather than falling more squarely within empirical cognitive science itself). That debate concerns the viability of classicism in light of the frame problem, and the prospects for a non-classical, non-computational, foundational framework for cognitive science.

1 The Frame Problem and Computational Cognitive Science

In the closing pages of Fodor (1983), it is argued that certain persistent and recalcitrant difficulties in classical cognitive science look to be in-principle obstacles, and hence that the prospects for understanding human cognitive processes like belief generation within the framework of classical cognitive science are very bleak indeed.[3] These obstacles continue to plague the computational approach to the mind (as stressed, for instance, by Fodor 2001), and they suggest the need for a radically different approach. The obstacles in question constitute the frame problem.

The main claim of Fodor (1983) is that the human cognitive system possesses a number of important subsystems that are *modular*: domain specific, mandatory, limited in their access to other parts of the larger cognitive system, fast, and informationally encapsulated. There is good evidence, Fodor argues, that human input systems, including those that

[2] We speak as though there are two distinctive kinds of features—descriptive and moral-normative—merely for convenience. We mean to be neutral in this paper about whether there are moral properties at all, and also about whether, if there are such properties, they are identical to certain properties characterizable in non-normative language. Also, it should be noted that according to the conception of moral normativity just described, although an overall system of moral normativity might include certain general principles linking so-called 'thick' moral features like moral justice to 'thin' moral features like moral obligation—e.g., W. D. Ross's principle that there is a prima facie duty to behave in a manner that accords with moral justice—nonetheless any such moral principles would have to be derivable themselves from more fundamental, exceptionless, principles that link purely descriptive features to (thick and/or thin) moral features.

[3] This section draws freely on material in Horgan and Tienson (1994, 1996) and Horgan (1997a).

mediate speech perception, exhibit modularity. Where the classical computational approach has gotten somewhere, he says, is in understanding such modular subsystems, which by their nature delimit the class of relevant information.

Classical computationalism has made very little progress in understanding *central* processes, however. Belief fixation—the generation of new beliefs on the basis of current input together with other beliefs—is a paradigmatic example. Updating of the overall belief system in light of currently available new sensory input is a closely related example. Fodor argues convincingly that these processes are non-modular: they need to have access to a wide range of cognitive subsystems, and to information on an indefinitely wide range of topics. And the very considerations that point to non-modularity, he maintains, also constitute grounds for extreme pessimism about the prospects of explaining central processes within the framework of classical computational cognitive science.

Fodor articulates these considerations in terms of the analogy between belief fixation in human cognition and scientific confirmation. Concerning central cognitive processes like belief fixation, he says, "it seems reasonable enough that something can be inferred about them from what we know about *explicit* processes of nondemonstrative inference—viz., what we know about empirical inference in science" (104). Scientific confirmation, "the nondemonstrative fixation of belief in science," has two crucial features. It is (in Fodor's terminology) *isotropic* and *Quineian*. He says:

> By saying that confirmation is isotropic, I mean that the facts relevant to the confirmation of a scientific hypothesis may be drawn from anywhere in the field of previously established empirical (or, of course, demonstrative) truths. Crudely: everything that the scientist knows is, in principle, relevant to determining what else he ought to believe.... (p. 105)

> By saying that scientific confirmation is Quineian, I mean that the degree of confirmation assigned to any given hypothesis is sensitive to properties of the entire belief system; as it were, the shape of our whole science bears on the epistemic status of each scientific hypothesis (p. 107).

Isotropy brings in the whole of current theory and background information: any bit of actual or potential information from any portion of the belief system might, in some circumstances, be evidentially relevant to any other. Being Quineian makes confirmation holistic in a deeper way: confirmation depends upon "such considerations as simplicity, plausibility, and conservatism" (108), which are determined by the global *structure* of the whole of the current belief system and of potential successor systems.

Since belief fixation in human cognition is a matter of nondemonstrative inference from the information provided by input systems and the information in memory, evidently it too must be isotropic and Quineian. Fodor concludes that it must be non-modular. He also stresses that these global aspects of belief fixation look to be at the very heart of the persistent difficulties that classicism has encountered in attempting to understand such central processes, linked to what has been dubbed the "frame problem":

> The difficulties we encounter when we try to construct theories of central processes are just the sort we would expect to encounter if such processes are, in essential respects, Quineian/isotropic.... The crux in the construction of such theories is that there seems to be no way to delimit the sorts of informational resources which may affect, or be affected by, central processes of problem-solving. We can't, that is to say, plausibly view the fixation of belief as effected by computations over bounded, local information structures. A graphic example of this sort of difficulty arises in AI, where it has come to be known as the "frame problem" (i.e., the problem of putting a "frame" around the set of beliefs that may need to be revised in light of specified newly available information) (pp. 112–3).

When one considers the sorry history of attempts in philosophy of science to construct a theory of confirmation, the prospects for understanding central processing within the classical computational paradigm look very discouraging indeed. The problem is that in general, a "frame" *cannot* be put around the set of beliefs that may need to be revised given newly available evidence. Thus Fodor:

> Consider...the situation in the philosophy of science, where we can see the issues about fixation of belief writ large. Here an interesting contrast is between deductive logic—the history of which is, surely, one of the great success stories of human history—and confirmation theory which, by fairly general consensus, is a field that mostly does not exist. My point is that this asymmetry, too, is likely no accident. Deductive logic is the theory of validity, and validity is a *local* property.... In this respect,...validity contrasts starkly with the level of confirmation, since the latter...is highly sensitive to global properties of belief systems.... The problem in both cases [belief fixation and confirmation] is to get the structure of the entire belief system to bear on individual occasions of belief fixation. We have, to put it bluntly, no computational formalisms that show us how to do this, and we have no idea how such formalisms might be developed (pp. 128–9).

Within philosophy, efforts by Carnap and his successors to develop a rigorous confirmation theory—for instance, a theory in which in which relations of high "logical probability" among propositions operate analogously to the relation of logical entailment within deductive logic—nowadays are widely regarded as having been unsuccessful. Hence Fodor's remark that the field of confirmation theory "mostly does not exist." And, although the pursuit of a rigorous confirmation theory still continues—typically within artificial intelligence rather than philosophy, and often within the framework of a Bayesian conception of epistemic probability rather than the "logical" conception of thinkers like Keynes and Carnap—the problem underscored by Fodor in the lately quoted passage remains very much alive. How can the structure of the entire belief system be brought to bear content-appropriately on individual occasions of belief fixation?

Let us underscore this problem by dwelling just a bit on the depth of, and the apparently in-principle nature of, the difficulties encountered by attempts to model global cognitive processes computationally. These difficulties, which involve the holistic aspects of belief fixation that Fodor labels the Quineian and isotropic features, are at the heart of the frame problem. Take, first, the Quineian aspect of belief systems. Simplicity and conservatism are properties of (or relations between) belief systems that depend upon the formal, semantic, and evidential relations among *all* of the beliefs in the system(s). A computational system would have to somehow survey the entire stock of beliefs, in a manner that tracks all the multifarious interconnections among the beliefs, and somehow derive a *measure* of net overall simplicity and net overall conservatism from these local features. In addition, when new information comes in from the input modules, the central system would have to find, from among the vastly many competing, incompatible ways of revising the whole system to accommodate this new information, a mode of revision that maintains overall simplicity and conservatism better than most of the others. All this would have to done via *tractable* computation, executable quite rapidly.[4] As Fodor said, "We have...no computational formalisms that show us how to do this, and...no idea how such formalisms might be developed."

[4] The expression 'tractably computable' is a term of art in fields like computer science, artificial intelligence, and cognitive science. It means 'computable by a physical computational device of the relevant kind'. This makes the expression both somewhat context-dependent (since the relevant kind of computational device varies with context) and vague (since in general it is not clear exactly what functions can be computed by a particular kind of computational device). In the present context, tractable computability essentially means computability by a physical device with roughly the same physical structure and roughly the same resource-capacity as a human brain.

Now consider isotropy—the potential relevance of anything to anything. One problem is that there is *too much* that may need to be considered, and that there is no known manageable way to bring what is needed into consideration. But there is actually an even deeper problem, as suggested by Fodor's talk of local versus global properties and by his comparison with the historical failure of confirmation theory: viz., that *confirmatory relevance itself is a Quineian as well as an isotropic property.* For instance, the belief that many black crows have been observed and that no non-black ones have been observed does not, by itself, support an inference concerning crows in general or concerning any particular crow. In order for such an inference to be appropriate, the predicate 'black' must be *projectible* for the system relative to the predicate 'crow'. But co-projectability of predicates is a global feature of a cognitive system, depending upon features of a great many other relevantly (!) similar predicates within the cognitive system. Fodor, in the passages bemoaning the lack of an available computational formalism, is telling us that human central processing evidently does not operate via any kinds of computation we currently know about or can even contemplate. He comes right to the brink of saying that it does not operate via computation—i.e., information processing in accordance with programmable rules—at all.

In Fodor (2001) these same themes are reiterated, with an eye on the ensuing two decades of work in cognitive science since Fodor (1983). On one hand, Fodor underscores the excellent reasons to believe that important central processes, e.g., belief-fixation and rational action-planning, are isotropic and Quineian, and thus are non-modular—and hence that the idea that such processes are "massively modular," though recently popular in some cognitive-science circles, is massively mistaken. And on the other hand, now he explicitly crosses the brink he reached in Fodor (1983), by asserting in effect that for agents whose background epistemic commitments are as extensive as those of humans, cognitive processes like belief fixation and planning just cannot operate via computation. He says:

> Computational nativism is clearly the best theory of the cognitive mind that anyone has thought of so far (vastly better than, for example, the associationistic empiricism that is the main alternative); and there may indeed be aspects of cognition about which computational nativism has got the story more or less right. But it's nonetheless quite plausible that computational nativism is, in large part, not true. (2001, p. 3)

The problem isn't nativism (however nativism might fare vis-à-vis empiricism in other respects), but rather the idea that the cognitive mind operates by *computation*—i.e., operates in accordance with exceptionless rules that could constitute a computer program. What seems needed is a fundamentally different theory of the cognitive mind, in place of the computational theory. Well, since the mind doesn't work *that* way—viz., via the *computational* manipulation of mental representations, then how must it work instead in order to overcome frame-type difficulties?

2 Morals of the Frame Problem

The frame problem arises largely because of the apparent computational intractability of managing all relevant information, insofar as that information gets *explicitly represented* in the course of cognitive processing. What this suggests is that somehow, belief fixation and related cognitive processes operate in a way that accommodates much relevant information *automatically* and *implicitly.* The suggestion is that the holistic aspects of belief fixation—the isotropic and Quineian aspects—do not involve the finding and fetching of relevant representations from memory-banks where they are stored in explicit form (because that could not tractably accommodate isotropy), and also do not involve the overt representation and comparative evaluation of large-scale alternative belief-systems (because that could not tractably accommodate the Quineian dimension). Rather, these holistic aspects are somehow implicit in the structure of the cognitive system, in such a way that temporal transitions from one occurrent cognitive state to another accommodate the holistic aspects automatically. In the terminology of Horgan and Tienson (1994, 1996), the holistic informational content is *morphological*, rather than occurrent. Morphological content is information that (i) is implicit in the standing structure of the cognitive system (rather than explicitly represented in the system's occurrent cognitive states or explicitly stored in memory), and (ii) gets accommodated in cognitive processing without getting explicitly represented in occurrent cognitive states, either conscious or unconscious.[5]

So one apparent moral of the frame problem is that in general, human belief fixation *must* operate in a way that draws heavily upon morphological content, in order to avoid computational intractability. As we will put it, these processes are *essentially* morphological. In general, belief fixation is not accomplished by computationally manipulating explicit, occurrent, representations of all relevant information. Essentially morphological processing is a fundamentally different way of accommodating the holistic aspects of belief fixation.[6]

There is a second moral of the frame problem, complementary to first one about the need for essentially morphological processing: viz., that cognitive processing that exhibits the Quineian and isotropic features, and does so with respect to a range of potential total cognitive states as rich as the range of such states instantiable by humans, is too complex and subtle to conform to *programmable rules* over mental representations. Another way to put this is that Quineian/isotropic human cognition *is not computation*; computation is too brittle, too frame-problem susceptible, to accommodate the Quineian and isotropic aspects of non-demonstrative belief-fixation.

Why exactly is it a moral of the frame problem that key aspects of human cognition are not computation? After all, within the confines of the computational theory of mind, there certainly is *some* room for information to be accommodated implicitly and automatically, rather than by being explicitly represented during processing. This happens when the information gets "proceduralized" through the deployment, in the computational system, of certain *special-purpose* rules that operate in a way that is automatically appropriate to that information. In executing such rules, the system accomplishes in a "shorthand" way what instead could be accomplished "longhand": it applies the special-purpose rules to a set of explicit representations that does not include representations of the background information, as a shorthand alternative to applying general-purpose rules to a larger set of explicit representations that *does* include representations of the background information.

A simple example is a computer program for assigning classes to classrooms that contains a shorthand rule to the effect that any input with the prefix PHIL should be paired with an output with prefix DERRIDA HALL, and also contains a similar rule for every other departmental prefix in the university. The information about which department has which building(s), the information that each department's classes are required to be paired

[5] For further elaboration of the notion of morphological content, and for discussion of how such content could get structurally embodied within cognitive systems whose temporal evolution is described formally in terms of the high-dimensional mathematics of dynamical systems theory, see Horgan and Tienson (1996).

[6] This is not to deny that within classical computational systems too, information can get implicitly accommodated rather than being explicitly represented. Indeed, at least some of the rules executed by a computational system need to be "hardwired in," even though usually various specific items of information (often including certain more specific rules) do get explicitly represented.

with rooms in that department's building(s), and the information that all and only philosophy classes have the prefix PHIL, is accommodated procedurally in these rules, rather than being explicitly represented. Contrast this with a computer program that assigns classes to classrooms by doing all the pertinent information-processing "longhand" by applying general-purpose rules to a set of representations that includes explicit representations of all such information.[7]

But there is excellent reason to think that special-purpose rules cannot be of much use in addressing the frame problem. The argument for this negative conclusion goes as follows. In order for computational processes employing special-purpose rules to be implementable in human neurobiological "wetware," the number of special-purpose rules executed by the computational system cannot be intractably huge. (For instance, if each "rule" is just a totally specific entry $[S_i(t) \rightarrow S_i(t')]$ in a lookup table, an entry stipulating exactly what state-transition should transpire at the next moment t' when the specific, explicit, representation S_i is instantiated at t—then the number of special-purpose rules would grossly outrun neurobiological implementability.) Hence, tractably executable special-purpose rules must still retain a significant aspect of *generality*, even though some of the background information that is accommodated via the execution of such rules is implicitly embodied in the rules themselves rather than being explicitly represented during processing. I.e., such a rule must apply to a reasonably wide range of explicit representations that can potentially occur in the cognitive system, thus generating a reasonably wide range of specific cognitive state-transitions. But such a special-purpose rule will be a rule that implicitly accommodates some *specific* body of background information; it will effect transitions from one explicit representation to another that are appropriate to the content of *that* implicit information. (In the example lately given, for instance, it would be the information that all and only philosophy classes have the prefix PHIL, and the information that all and only philosophy classes are to be assigned classrooms in Derrida Hall.) Change the background information, and you need a new special-purpose rule that is appropriate to the new background information rather than to the original background information. Thus, suitably general special-purpose rules will have to be *framed* rules, in order for the number of such rules to remain small enough for them to be tractably executable by a physical device with roughly the physical resources of a human brain; i.e., for any such rule there will have to be a specific, delimited, body of background information that the rule implicitly accommodates. But the moral of the frame problem, remember, is this: because of the holistic Quineian/isotropic ways that background information is relevant to belief fixation, one *cannot* delimit ("put a frame around") the background information that *might become relevant when some new explicit representation arises in the system*. Thus, the frame problem arises all over again, for shorthand rules. Tractably implementable special-purpose rules must be framed rules, but those can't accommodate the Quineian/isotropic aspects of belief fixation.

In short, one can only expect there to be tractably implementable computational rules of a special-purpose, shorthand, kind for accomplishing a cognitive task when there are also tractably implementable computational rules for accomplishing a task by longhand. To the extent that the Quineian and isotropic aspects of background information cannot be

[7] The distinction between general-purpose rules and special-purpose rules is orthogonal to the distinction between rules that are explicitly represented and rules that are automatically executed by virtue of a system's architecture (its "hard-wiring"). *Some* rules need to be executed architecturally in a computational system, whereas others can be explicitly represented and then explicitly consulted during processing. But wired-in rules can be special-purpose (as they often are in special-purpose devices like hand-held calculators) or general-purpose, and explicitly represented rules also can be either special-purpose or general-purpose.

accommodated longhand via general-purpose computational rules operating on explicit representations of all pertinent information, they also cannot be accommodated shorthand via special-purpose computational rules in which background information is embodied procedurally.

The persistence of the frame problem in computational cognitive science, along with the above considerations to the effect that the Quineian and isotropic features of central cognitive processes are computationally intractable for cognitive systems on the scale of real humans, together provide strong abductive grounds for concluding that processes like belief-fixation must somehow deploy very extensive morphological content, and must somehow operate in ways more complex and subtle than does any device that manipulates mental representations via computation (i.e., via exceptionless rules of the kind that could constitute a computer.

It bears emphasis that both claims we are treating as morals of the frame problem are *abductively* warranted by the existence of the problem, via inference to the best available explanation—rather than being outright entailed. First, consider the non-computation moral. Although nobody has any idea how the Quineian/isotropic features of human cognition could be managed via tractable computation, it remains possible nonetheless that this kind of cognition deploys some not-yet-discovered form of tractable computation anyway. We grant the point. Still, the best available explanation for the fact that nobody has any idea how the Quineian/isotropic features of human cognition could be managed via tractable computation is that these features *cannot* be so managed—and hence *are* not so managed, in humans.

Second, take the claim that Quineian/isotropic cognition of the kind exhibited by humans must accommodate lots of background information in an essentially morphological way. Even if the first moral is correct, i.e., human Quineian/istotropic processes cannot operate via tractable computation, it remains possible nonetheless that such processes always deploy explicit, stored, representations of all relevant information—representations that are fetched non-computationally, and perhaps are then manipulated non-computationally. Still, the only available explanation of how these features could be non-computationally managed—and hence the *best* available explanation, despite being somewhat sketchy—invokes the idea that much relevant background information must be accommodated implicitly and automatically by the operative non-computational processes. Moreover, the abductive case for the claim that extensive morphological content is needed is considerably strengthened, in our view, by these three facts: (1) there is a credible-looking nonclassical framework for cognitive science—briefly described in section 4 below—that explains in principle how non-computational Quineian/isotropic processing could be subserved by certain kinds of high-dimensional dynamical systems that are implementable by brain-style neural networks, (2) this framework affords a central role to essentially morphological processing, involving extensive information implicitly embodied "in the weights" of a neural network, and (3) there are not, as far as we know, any credible ideas in cognitive science for how non-computational Quineian/isotropic processing could be accomplished in any *other* way.

Another point worth stressing, in addition to the non-demonstrative, abductive, nature of our preceding argumentation, is this: the claim that central cognitive processes are *non-computational* does the key dialectical work in the later part of the present paper. Strictly speaking, our overall argument does not need to mention morphological content at all. (Thus, one could accept our subsequent arguments even if one repudiates the idea of morphological content altogether, as long as one accepts that Quineian/isotropic processing is non-computational.) Nevertheless, the reason we are stressing the idea of essentially morphological processing is that this notion seems so central to any remotely plausible conception of non-computational Quineian/isotropic cognition, and because it also is central to the non-computational framework for cognitive science briefly described in section 4.

3 An Optimistic Reply and a Pessimistic Rejoinder

At this juncture, a fan of the computational theory of mind might reply as follows. First of all, recent approaches in cognitive science to what was *originally* dubbed 'the frame problem' typically invoke *default* rules, not exceptionless rules. (The original frame problem arose in the context of AI systems that operated via logical deduction from exceptionless axioms.) Secondly, default rules can be implemented computationally—something that is often done in AI modeling.[8] Hence, it is just a mistake to claim that the computational theory of mind requires cognitive processing to conform to exceptionless rules. Thus, the frame problem does not really tell against the computational theory of mind at all.

Although this reply sounds initially plausible, we maintain that it actually rests on a subtle confusion involving the widely deployed term 'implementation'. In order to clarify the dialectical situation, let us make a distinction. On one hand is what we will call *direct* computational implementation: the rules to be implemented are themselves programmable (i.e., they collectively constitute a computer program), and hence they are exceptionless.[9] These exceptionless programmable rules then get implemented by *other* exceptionless programmable rules in a "lower-level" language—and so on perhaps, down through an implementational hierarchy terminating in exceptionless machine-language rules that get directly implemented physically in the computational device.

On the other hand is what we will call computational implementation by *classical refinement*. Here the set of to-be-implemented rules includes one or more default rules (e.g., rules containing a 'ceteris paribus' clause), and the set does not fully specify what are the exceptions to these default rules. Implementation now involves two components. The first component is a refinement of the initial set of rules into a set of exceptionless programmable rules. This refinement is a new set of rules with these two features: (i) the new rules fully specify all potential exceptions vis-à-vis the default clauses of the original rules, and (ii) the new rules fully specify what the system is to do in any given situation (including any situation involving an exception to one of the original default rules). The second component of computational implementation by classical refinement is a *direct* implementation of the classically refined rule-set—which thereby amounts in effect to an *indirect* implementation of the original default rules.

The classicist framework, in artificial intelligence and in cognitive science, is committed to the idea that cognitive processing is computation. More specifically, the idea is that cognition conforms to rules with these three features: (i) they are expressible in the format of a computer program, which means that they must be precise and exceptionless, (ii) they operate directly over representations themselves, and (iii) they advert solely to formal-

[8] The term 'implementation' is used in computer science, artificial intelligence, and cognitive science in much the same way that the term 'realization' is used in philosophy of mind. The idea is that normally there are multiple distinct ways in which states or processes, as characterized at some "higher" level of description, can be carried out by states or processes characterized at some "lower" descriptive level.

[9] Note that rules can be exceptionless without being deterministic. For instance, the rules could instruct a system to "throw the dice" in certain specific situations (e.g., by consulting a table of random of numbers), and could specify different follow-up actions for different potential outcomes of the dice-throw.

structural features of the representations that play an information-encoding role. These are what Horgan and Tienson (1996) call *programmable representation-level rules* (PRL rules).[10] This commitment of classicism has the following consequence: whenever a system of default rules gets computationally implemented, that implementation proceeds via a classical refinement of the initial rule-set. According to the computational theory of mind, the default rules *must* be implementable this way, because classicism presupposes that cognition operates by executing PRL rules. Typically, the theoretician who is modeling the use of default rules produces such an implementation by formulating the replacement-rules themselves, in the format of a computer program. But in principle, an implementation by classical refinement could be produced in other ways too—e.g., by providing suitable training-input to a self-modifying system that employs meta-level PRL rules to progressively alter and update its own first-order PRL rules.

It is easy not to notice or appreciate the difference between direct implementation and implementation via classical refinement. Once this distinction is thematized, however, it becomes evident that there is confusion residing in the optimistic idea that classicism can overcome the frame problem by "computationally implementing" default rules. The pessimistic rejoinder goes as follows. In order to overcome computational intractability, the PRL rules that collectively constitute a classical refinement of a given set of default rules will have to be *framed* rules, in the sense explained earlier—i.e., rules that delimit the range of potentially relevant information that they take into account. But framed PRL rules run afoul of the isotropic and Quineian aspects of evidential holism. Thus, we get the result that for any set of PRL rules that constitutes a classical refinement of an initial set of default rules, the replacement rules will not come anywhere close to approximating the isotropic and Quinean aspects of genuine human belief fixation. So the frame problem arises all over again, this way: when default rules get computationally implemented via classical refinement into PRL rules, these replacement rules must be framed rules (on pain of computational intractability); but framed PRL rules, for reasons emphasized earlier, evidently cannot accommodate the holistic aspects of belief fixation.

At this point, some who profess to be fans of both default rules and a computational conception of cognition might protest that they themselves do not assume that cognition must conform to *representation-level* programmable rules. They might say that the following scenario is entirely consistent with what they themselves understand as the computational theory of mind: (1) there are default rules, governing human mental representations themselves, that cannot be refined into exceptionless *representation-level* rules of the kind that would constitute a computer program (i.e., cannot be refined into PRL rules); yet (2) at one or more sub-representational levels of description, the brain-processes that subsume human cognition do constitute computation (and hence conform to exceptionless, programmable, rules).

We have several remarks to make, concerning such a stance. First, we believe that the scenario just described is intelligible: a representation-processing system could conform to exceptionless, programmable, sub-representational rules while yet failing to conform to exceptionless, programmable rules over representations themselves. (See section 4 below, and especially note 12.) Second, for present purposes we need not insist that labels like 'the computational theory of mind' are rightly applicable only to views committed to PRL rules;

[10] For further elaboration and defense of the claim that classicism presupposes such PRL rules, see Horgan and Tienson (1996), Chapter 2, and Horgan (1997b).

rather, how to apply such labels is really a terminological matter, not a substantive one.[11] However, third, it appears to us that a de facto commitment to PRL rules really is virtually always at work in the computational modeling that goes on in AI and in cognitive science, and thus that the kind of implementation of default rules provided in such models is virtually always implementation by classical refinement—the kind of implementation for which the frame problem arises all over again. Fourth, the task of seeking out some alternative kind of implementation of default rules, perhaps one that deploys sub-representational computation without deploying representation-level computation at all, is certainly a worthy goal for cognitive science; and meeting that goal may well be an important prerequisite for coming to terms with the frame problem. But fifth, it is important to appreciate how radically different such a form of implementation would be from implementation by classical refinement; this alternative would be implementation—perhaps subrepresentational *computational* implementation—of cognitive processing that does not itself constitute representation-level computation at all. Finally, sixth, it is also important to appreciate that so far, computational modeling in artificial intelligence and in cognitive science has not delivered any clear way of implementing default rules *except* for implementation by classical refinement.

4 Dialectical Interlude: A Non-Classical Framework for Cognitive Science

As a free-standing piece of abductive reasoning, the argument set forth in sections 2 and 3 is imposing. Yet it also leads directly to the following question. Is there work in cognitive science suggesting how some meat could be put on the bones of the non-computational, massively morphological, conception of human cognition described in section 2? Some philosophers and cognitive scientists will say yes, and point to *connectionism*—the use of neural-network models. Caution is required here, however, because extant neural-network models applied to simple "toy problems" face scale-up issues just as much as do classical computational models, and the frame problem kicks in with a vengeance when one seeks to scale up from toy problems to human-style cognition. Nevertheless, there is at least one potentially promising non-computational framework for cognitive science: the "dynamical cognition" framework described in Horgan and Tienson (1994, 1996) and in Horgan (1997a). This framework is inspired partly by the persistence of the frame problem as an in-principle looking problem for the computational theory of mind, partly by the emergence of the connectionist movement in cognitive science in the mid-1980's, and partly by the natural links between neural networks and the branch of mathematics called dynamical systems theory. The dynamical cognition framework jettisons the classical assumption that human cognitive state-transitions must conform to a tractably computable transition-

[11] Larger issues loom in the vicinity of this one, however, and are not merely verbal. Among them are these: (1) Do those who embrace labels like "the computational theory of mind" make the foundational assumption, at least implicitly, that human cognition conforms to PRL rules? (2) How good is the evidence for the sociological claim that they do make this assumption? Our own view is that the actual practice of researchers in cognitive science and in artificial intelligence is ultimately more important in addressing such questions than are the answers the researchers might give if queried about their own foundational assumptions—and that actual practice reflects a strong (if sometimes only implicit) commitment to PRL rules. Suppose, however, that many practicing researchers doing computational modeling would say "We don't assume that cognition conforms to PRL rules, because our models deploy default rules that then get implemented computationally." This would only mean that many researchers are somewhat confused about the underlying assumptions that inform their actual model-constructing practice.

function over mental representations, and it puts heavy emphasis on morphological content.[12]

Fodor, as is well known, is a philosophical arch-enemy of connectionism, which he regards as a recidivist return to the dark days of associationism in psychology (Fodor and Pylyshyn, 1988). Here we will make just a few quick observations about this matter. First, no doubt many extant connectionist models really do operate via statistical/associative information-processing; and to the extent that they do, there is little reason to expect these models to scale up. But second, the dynamical cognition framework is not fundamentally associationist; on the contrary, it gives pride of place to mental representations with very rich structure (including language-like structure, construed non-classically) and to structure-sensitive processing. Third, even if one agrees with Fodor that connectionism is really a step backward in cognitive science, and even if one repudiates the dynamical cognition framework (say, because allegedly it is a "degenerating research program"), the fact remains that there are very strong abductive grounds for holding that central cognitive processes are both non-computational and essentially morphological. Fodor himself has been rubbing people's faces in this fact for many years, and rightly so.

This brief section has figured merely as a dialectical interlude because of the third point lately mentioned. Although it is certainly worth noting that there is at least one potential foundational framework for cognitive science that has some promise for explaining non-mysteriously how human cognition could outstrip the limits of computationalism and could accommodate the holistic aspects of belief fixation—viz., the dynamical cognition framework—the overall line of argumentation in the present paper is independent of this fact.

5 Epistemic Competence and Epistemic Normativity: Beyond Exceptionless General Rules

It will be useful to elaborate in the following way the idea that human cognitive transitions involved in belief fixation are non-computational and essentially morphological (cf. Horgan and Tienson, 1996). Let a *total cognitive state* (a TCS) be a state of a cognitive agent A, at a time t, that incorporates not only all mental states occurrently tokened by A at t, but also all background information that is morphologically embodied at t by A's cognitive architecture. A TCS can be thought of as a pair $<B_i, S_j>$, where B_i is the full body of background information possessed by an agent at a time, and S_j is the full set of mental states occurrently tokened by A at t. Let the *epistemic competence profile* (the ECP) for a given kind of agent (e.g., the kind *human*) be an enormous list of completely specific cognitive transitions from one potential TCS to another:

$$< B_i, S_j > (t) \rightarrow < B_{i^*}, S_{j^*} > (t')$$

$$< B_h, S_k > (t) \rightarrow < B_{h^*}, S_{k^*} > (t')$$

.
.
.

[12] One might think that the kinds of connectionist networks that are simulable by standard computers—networks that update in discrete time-steps, have discrete node-activation values, and are discrete in other pertinent ways too—could not implement non-computational cognitive processing. But that would be a mistake, because conformity to programmable rules is a property that need not "transfer upward" from the subcognitive level to the cognitive level; cf. Horgan and Tienson (1996), Chapter 4, section 4.4, and also Horgan (1997b).

where t' is the next moment in time after t, and where this list embodies all the cognitive transitions that would accord with an ideally complete and correct cognitive-scientific account of *cognitive competence* for agents of the kind in question. The ECP abstracts away from potential "performance errors" due to tiredness, emotional distress, inattention, and the like. Thus, the cognitive total-state transitions on the list are the ones that the cognitive agent would exhibit insofar as it alters and updates its beliefs by deploying its belief-forming cognitive mechanisms as those mechanisms are *designed* to operate. (The relevant cognitive-architectural design was installed evolutionarily, presumably.)[13] The first moral of the frame problem, viz., that central cognitive processes are essentially morphological, can be put by saying that for those processes, the B_i component is (and must be) both (i) very extensive and (ii) very largely morphological.

The second moral, viz., that central cognitive processes are non-computational (and thus that the ECP is not a tractably computable function), can be put by saying that the individual state-transitions collectively constituting the ECP *cannot be fully systematized by exceptionless general principles expressible as programmable rules*.[14] This means that there is no set of exceptionless, programmable, general rules to which the individual state-transitions collectively conform.[15]

A third moral now emerges too, a non-demonstrative extrapolation from the second one: viz., that the individual state-transitions collectively constituting the ECP *cannot be fully systematized by exceptionless general principles at all*. This third moral does not follow logically from the second. One reason why not is that there can be phenomena that conform to exceptionless general principles without conforming to principles that constitute *programmable rules*—e.g., temporally continuous physical phenomena that (i) conform to exceptionless general dynamical laws, and yet (ii) are such that the temporal trajectories of these phenomena are not always computable. But the Quineian/isotropic aspects of belief fixation tell against full systematizability by exceptionless general rules of *any* kind—not merely against full systematizability via tractable, programmable rules (i.e., rules of the

[13] One could allow for different forms of cognitive competence, capturable by different ECP's; and one could allow the ECP to be indeterministic, with some items on the right of the arrow being disjunctions of TCS's; but we ignore such complications for simplicity.

[14] Actually, the moral as so expressed is logically stronger than the earlier moral that human cognitive processes are non-computational. For, the earlier moral amounts to the assertion that the ECP cannot be fully systematized by any *tractably executable* exceptionless general principles expressible as a computer program. This assertion is logically consistent with the thesis that the ECP nonetheless can be fully systematized by certain exceptionless general rules collectively constituting a computer program that is *not* tractably executable. However, the stronger moral too seems well supported abductively by the frame problem, as follows. First, nobody has any idea how the Quineian/isotropic features exhibited in human cognition could be managed via *any* kinds of exceptionless, general, programmable rules—regardless of whether or not the rules in question are tractably implementable. Second, the best available explanation for this fact is that the Quineian/isotropic features exhibited in human cognition *cannot* be so managed—not by general programmable rules that are tractably implementable, and not even by general programmable rules that are *not* tractably implementable.

[15] It is worth noting that the ECP could be a computable function, in the mathematical sense described by Church and Turing, even if the ECP is not *tractably* computable—i.e., even if it does not conform to any set of rules which both (i) are expressible as a computer program, and (ii) can be executed by a physical system with physical resources roughly comparable to those of a human brain. For instance, the ECP would be a computable function in the Church/Turing sense as long as it is *finite*, even if the number of state-transitions it includes is gargantuan. This is because the full set of concrete state-transitions constituting the ECP would itself count as a set of programmable rules (by virtue of being a finite set). Yet, since each of these rules would be completely specific in content—would be a single entry in a "lookup table"—these kinds of rules would lack any form of generality. Such non-general transition-rules therefore would not *systematize* the ECP, but instead would merely collectively *constitute* the ECP.

kind that can be implemented by a digital computer, or by a neural network with physical resources comparable to those of a human brain). For, insofar as anyone has clearly envisioned general psychological laws that would be analogous to general dynamical laws in physics—e.g., psychological laws expressible as universally quantified partial differential equations involving quantitative psychological magnitudes—such laws would describe interactions among psychological properties or magnitudes that depend only on the *quantitative strengths* of the properties, rather than depending on their intentional contents together with the Quinean/isotropic features of background morphological content. (A classic example is Hume, who envisioned dynamical laws of association for psychological magnitudes that would be analogous to Newton's laws of motion for physical magnitudes. Humean association, involving "ideas" with varying degrees of "force" and "vivacity," presumably would involve additive or quasi-additive interactions of all the simultaneously present ideas, as determined by their respective degrees of force and vivacity—interactions roughly analogous to physical interactions describable mathematically via a "parallelogram of forces.")

So in light of the frame problem, it is very likely that the ECP for humans is not fully systematizable by exceptionless general rules or laws. It bears emphasis, however, that the ECP still might be *partially* systematizable via some set of general principles—e.g., by principles that have ineliminable *ceteris paribus* clauses that advert not merely to potential exceptions due to physical breakdown or external interference (e.g., having a stroke or being hit by a bus) but also to exceptions within the domain of competent cognition itself. Such principles are what Horgan and Tienson (1990, 1996) call *soft laws*. Horgan and Tienson argue that soft laws, which can be thought of as expressing certain cognitive *tendencies*, figure importantly in psychological explanation, even though they cannot be refined into exceptionless generalizations.[16]

An epistemic competence profile specifies, in list form, all the potential belief-updating cognitive transitions that a given kind of cognitive agent *would* make, insofar as the agent avoids performance errors and deploys its belief-forming cognitive architecture in the manner in which it was evolutionarily designed to be deployed. Consider now another kind of list, which we will call an *epistemic normativity profile* (ENP). This is an enormous list of completely specific cognitive transitions from one specific potential TCS to another that the given kind of agent *ought epistemically* to make—i.e., cognitive transitions that are evidentially appropriate, given the agent's total available information. The list will have the following form, with 'O_e' symbolizing '$\text{Ought}_{\text{epistemically}}$':

$$< B_i, S_j > (t) \rightarrow O_e\left[< B_{i^*}, S_{j^*} > \left(t'\right)\right]$$
$$< B_h, S_k > \rightarrow O_e\left[< B_{h^*}, S_{k^*} > \left(t'\right)\right]$$
.

.

.

The ENP for a given kind of agent might, or might not, fully coincide with the agent's ECP—where coinciding means that the two are identical apart from the presence or absence of the 'ought'-operator in the right-side entries. This would depend primarily on whether or not there are certain respects in which the agent is disposed, as a matter of its cognitive-architectural design, to form certain beliefs that are inappropriate from the perspective of epistemic normativity. But in any case, for humans the ECP certainly can be expected to coincide at least

[16] These remarks about soft laws are closely related to the points we made in the final paragraph of section 3 above. If a cognitive system executes "default rules" *that are not implemented in the system by classical refinement*, then these rules are soft in the sense just specified.

approximately with the ENP, for at least two reasons. First, otherwise the human race presumably would not have emerged and persisted in the course of evolution. Second, *human* epistemic normativity is surely constrained by an "ought implies can" principle—or, more accurately, a principle of the form "ought implies can, *near enough*." (The "near enough" qualifier allows for the possibility that the demands of human epistemic normativity *modestly* exceed human capacities to meet those demands—e.g., because of innate limitations due to factors like the size of working memory, or because of innate cognitive processes that are "fast and frugal" at the cost of sometimes leading to epistemically unwarranted beliefs.[17])

There are three further morals of the frame problem, pertaining specifically to human epistemic normativity—i.e., to *evidentially appropriate* cognitive-state transitions, for humans.[18] (They are parallel to the morals already mentioned concerning human epistemic competence, and they arise because the human ENP must closely approximate the human ECP.) First, because human epistemic normativity extensively involves holistic evidential factors with the Quinean and isotropic features, the B_i component in the ENP is (and must be) very extensive. Second, human epistemic normativity very probably does not conform to exceptionless general normative principles of belief-updating that are expressible as programmable rules; i.e., the human ENP is not fully systematizable by such principles. Third, by non-demonstrative extrapolation from the first two morals, the individual transitions collectively constituting the ENP very probably cannot be fully systematized by any general principles at all.

Further reinforcing the claim that the ENP is not fully systematizable by exceptionless general rules or laws is the sorry history of attempts over the years to discover and articulate exceptionless general normative-epistemic principles that govern cognitive tasks like belief fixation and rational planning. This fact combines with the persistence of the frame problem to strongly suggest not only that human epistemic competence is not fully systematizable by exceptionless general principles, but also that epistemic normativity itself is not thus systematizable.

6 The Import for Moral Normativity

Turn now to moral normativity. Here it will be easiest to discuss the import of the frame problem disjunctively. On one hand, perhaps the right metaethical position is some version of what we will call *robust objectivism* about moral normativity. Roughly, robust objectivism asserts one or both of the following: (i) that the right ontology includes moral facts and moral properties (with the terms 'fact' and 'property' being employed in an ontologically committal, non-deflationary way), or (ii) that some single system of moral normativity uniquely counts as fully rational, fully in accord with reason itself. On the other hand, perhaps the right metaethical position is one that denies that there is robust objectivity in morals—e.g., some version of expressivism.[19] We will consider these two possibilities in turn.

[17] Cognitive architecture might, for instance, be evolutionarily designed to deploy certain heuristics that sometimes generate biased judgments that fail to be epistemically appropriate. Work in cognitive science that is relevant to this possibility includes Kahneman and Tversky (1972), Kahneman, Slovic, and Tversky (1982), Kahneman and Frederick (2005), Gigerenzer and Selten (2002), Gigerenzer (2007).

[18] We speak of *human* epistemic normativity because we find it plausible that what counts as epistemically appropriate belief-formation for a given kind of creature depends in part on what kinds of belief-forming processes that creature is capable of executing. Epistemic competence constrains epistemic normativity, whether or not the two coincide.

[19] See, for example, Horgan and Timmons (2006a, 2006b).

6.1 Robust Objectivism and the Frame Problem

Suppose, first (and for the duration of the present subsection), that robust objectivism is true—i.e., there is some single system of moral normativity that is objectively correct. Think of this system as given by a gigantic list of highly specific conditionals of the form

$$\textit{If } D_i \textit{ then } E_j,$$

where D_i is a (possibly very complex) descriptively characterized state of affairs, and E_j is a specific moral-normative claim that is true in D_i. We will call this list the *objective moral normativity profile* (OMNP). Human beings presumably are capable of forming moral judgments, and then acting on those judgments, in ways that closely conform to the OMNP itself; for, objective moral normativity applicable to humans is presumably constrained by an "ought implies can" principle—or, more accurately, a principle of the form "ought implies can, *near enough*." Thus, if robust objectivism is true then humans presumably have a form of judgment-forming competence, vis-à-vis objective moral normativity. There will be an *objective moral competence profile* (OMCP) for humans. It too will be given by a gigantic list of conditionals of the form

$$\textit{If } D_i \textit{ then } E_j,$$

and the OMCP either will be outright identical to the OMNP or at least will closely approximate it. (The OMCP will be identical to the OMNP if any failures to conform to the OMNP would be performance errors. The OMCP will be non-identical to the OMNP if certain failures to conform to the OMNP would be products of the evolutionary design of human cognitive architecture.[20])

The lessons of the frame problem have import for (putative) objective moral normativity much like their import for epistemic normativity. First, given that central cognitive processes like belief fixation and rational planning do not conform to a tractably computable function, and thus that the human *epistemic* competence profile is not fully systematizable by exceptionless general rules or laws, it is entirely possible that the OMCP and the OMNP are not tractably computable either, and that moral normativity is not fully systematizable by general rules or laws. Since humans are capable of mastering epistemic normative standards that are too subtle and complex to conform to exceptionless general principles (because otherwise central processes could not be Quineian and isotropic), presumably they are also capable of mastering moral normative standards that are comparably subtle and complex.

Second, the lessons of the frame problem provide positive grounds for holding that objective moral normativity (if there is such a thing) is *in fact* not systematizable by exceptionless general principles. For, prima facie it is just enormously plausible that competent moral judgment-formation and moral decision-making are isotropic and Quineian, and that this is because moral normativity itself exhibits these same features.

[20] For possible examples of such failures, see Haidt (2001) and Greene (2007). Haidt points to cases in which subjects are morally 'dumbfounded' in not being able to give reasons for the putative wrongness of certain actions that they immediately judge to be wrong (e.g., harmless brother-sister incest). Greene argues that moral intuitions associated with deontology—e.g., retributive responses to wrong-doing—are (contrary to what defenders of this view suppose) largely a matter of emotion-based gut responses that can be traced to humans' evolutionary past, and that reflect morally arbitrary differences. For a response to Greene's critique of deontology, see Timmons (2007).

They are isotropic because virtually anything the agent knows or believes, and virtually any piece of information that is morphologically embodied in the agent's cognitive system either innately or through learning, can have a bearing on what is good, or right, or obligatory, etc. Moral judgment-formation and decision-making are Quineian because certain moral judgments and decisions often involve all-things-considered moral features that are structurally no less holistic than are features like *simplicity* in non-demonstrative theoretical inference—features like *net overall fairness*.

Several additional points bear emphasis, concerning these two lessons of the frame problem. First, these lessons undercut one line of thought that often seems to motivate the belief that moral normativity must conform to exceptionless general principles—viz., the idea that otherwise humans would be incapable of mastering moral normativity and developing the capacity for accurate moral-judgment formation. Humans are capable, somehow, of *epistemically* appropriate belief-formation—even though the Quineian/isotropic holistic aspects of evidential relevance mean that epistemic normativity outstrips conformity to exceptionless general principles. So humans should be capable of mastering moral normativity as well, even if it too outstrips conformity to exceptionless general principles. Thus, whether or not moral normativity is *in fact* fully systematizable by such principles, the first lesson stands: it *need not* be, in order to be masterable and deployable by humans. (We return to this theme in section 10.)

Second, although we ourselves find it overwhelmingly plausible that robustly objective moral normativity (if there is such a thing) must conform to the principle "ought implies can, near enough," suppose one were to deny this claim, or at any rate call it into question. It nonetheless would remain very likely, given that robustly objective moral normativity exists, that it is not fully systematizable by exceptionless general principles. For, it would remain very likely that Quineian/isotropic factors holistically determine what is all-in morally good, all-in morally obligatory, etc., and that these holistic factors operate in a manner too subtle and too complex to conform to exceptionless general principles.

Third, certain kinds of value monists—e.g., utilitarians of the kind who construe social utility as a descriptive feature rather than a moral-normative one—might want to argue that moral normativity is, in fact, fully systematizable by an exceptionless general principle linking the descriptive to the normative—e.g., the principle that acts that maximize overall social utility are morally right. But this view really only *re-locates* the import of the frame problem for moral normativity, rather than eliminating it. For, holistic Quineian/isotropic factors would be among the determinants of overall social utility—and hence there would be no exceptionless general principles that systematize how the putatively *descriptive* attribute of overall social utility would be determined by various other descriptive features. The frame problem would arise all over again—now as having the moral that there are no exceptionless general rules that express how descriptive features combine and interact with one another to yield some specific degree of overall social utility.[21]

6.2 Non-objectivism and the Frame Problem

By *non-objectivism* we mean any metaethical position that denies the existence of robustly objective moral normativity. Under non-objectivist views, an agent with a developed

[21] Value pluralists, on the other hand, can plausibly maintain that moral properties like net overall goodness/rightness are often subject to holistic, Quineian/isotropic, determinants—and thus that moral normativity itself is not fully systematizable by exceptionless descriptive-to-normative principles. Certain value monists can plausibly maintain this too—e.g., utilitarians of the kind who construe social utility as itself a moral-normative feature rather than a purely descriptive one.

capacity for moral-judgment formation will have internalized some specific, subjective, form of moral normativity. Maybe there is just one kind of subjective moral normativity internalized by all humans—say, because it was instilled into humans via evolution as a "moral grammar," analogous to the kind of "universal grammar" that Chomskian linguists suppose was evolutionarily instilled in humans.[22] Or maybe there are multiple different kinds of subjective moral normativity, with different ones being internalized by different individuals and/or different social groups. In any event, the specific, subjective, version of moral normativity that has been internalized by a particular agent will be given by a *subjective moral normativity profile* (SMNP) for that agent, which can be construed as a gigantic list of highly specific conditionals of the form

$$\textit{If } D_i \textit{ then } E_j,$$

where D_i is a (possibly very complex) descriptively characterized state of affairs, and E_j is a specific moral-normative claim that is true in D_i. Likewise, the agent will have developed a specific cognitive competence vis-à-vis this SMNP—i.e., a specific *subjective moral competence profile* (SMCP)—which also is construable as given by a gigantic list of such conditionals. The agent's SMCP either will be outright identical to the SMNP or at least will closely approximate it. (The SMCP will be identical to the SMNP if any failures to conform to the SMNP would be performance errors. The SMCP will be non-identical to the SMNP if certain failures to conform to the SMNP would be products of the evolutionary design of human cognitive architecture.[23])

For virtually any potential form of subjective moral normativity that is even a moderately realistic candidate for internalization by a human agent, we submit, the various remarks in section 6.1 all carry over, mutatis mutandis. Real-life moral judgments and decisions, after all, often involve all-things-considered moral features that are structurally no less holistic than are features like *simplicity* in non-demonstrative theoretical inference—features like *net overall fairness*. So, although no doubt some forms of subjective moral normativity internalized by humans are more crude and more rigid than others, few (if any) are apt to conform to exceptionless general principles.

7 Partial Systematizability of Moral Normativity via Moral Principles?

Historically, philosophers have often hankered after exceptionless principles that would fully systematize moral normativity. This is an itch that probably cannot be scratched, because very likely no such principles exist. Even so, however, it is important to recognize several (mutually compatible) ways that certain kinds of general principles might still *partially* systematize moral normativity.

First, there might be exceptionless general principles to the effect that certain descriptive features are *contributory* to morality in a certain way. A candidate such principle, for example is that *causing pleasure is always a good-contributing characteristic of an action*.

[22] For a cognitive-scientific defense of the idea of a universal "moral grammar," see Hauser (2006). Of course, this idea is also compatible with (without presupposing) robustly objective moral normativity.

[23] There might be certain innate cognitive-architectural features that influence moral-judgment formation in a manner that the agent could not reflectively endorse (and do not fit well with rest of the agent's subjective moral normativity profile), but persist anyway. Examples might include judgments of the kind mentioned in note 20 above, such as the intuitive judgment that harmless brother-sister incest is morally wrong, or certain types of deontological judgments reflecting retributivist reactions to wrong-doing.

One could hold that moral normativity conforms to principles of this kind, even if one denied that there are exceptionless general principles that invariably determine, for instance, how various good-making and bad-making characteristics combine to yield *net overall* moral goodness or badness.

Second, there might be general principles that are *soft*, in the sense that they have *ceteris paribus* clauses that cannot be eliminated—i.e., the principles cannot be replaced by refined principles that lack *ceteris paribus* clauses but still are interestingly general rather than being highly specific. A candidate such principle is that *ceteris paribus, lying is wrong*. One could hold that moral normativity conforms to such principles, while denying that it is fully systematizable.[24]

Third, there might be general *contributory* principles that are soft rather than exceptionless. A candidate such principle is that *ceteris paribus, causing pleasure is a good-contributing characteristic of an action*. Someone might hold, for instance, that certain descriptive characteristics have the default status of being good-making, while also allowing that in certain circumstances this status gets "silenced" or even "reversed."[25] (Such a person might claim, for instance, that in a circumstance when a sadist derives pleasure from a sadistic act, the fact that the act causes such pleasure counts as a *bad*-contributing characteristic.)

[24] Both W. D. Ross and A. C. Ewing accepted soft principles of this sort, though they used the term 'prima facie' instead of 'ceteris paribus' in articulating them. In addition, for each basic soft principle of rightness or duty, they accepted a corresponding exceptionless contributory principle. In discussing the soft principle of fidelity, Ross remarks: "It remains a *hard fact* that an act of promise-breaking is morally unsuitable in so far as it is an act of promise-breaking, even when we decide that in spite of this it is the act that we ought to do" (1939: 86, our emphasis; see also 134). Ewing is even clearer on the matter: "'It is a prima facie duty not to tell lies' does not mean 'we ought usually to avoid telling lies', but 'that X would be a lie is always a valid reason, though not necessarily a conclusive reason, against saying X'" (1959: 110); "*prima facie* principles are still in a sense universal: the fact, e.g., that something gives pain is *always* a reason against doing it…" (1959: 126). One point we are calling attention to here is that embracing soft ceteris paribus principles does not commit one to corresponding exceptionless contributory principles; with a soft principle of rightness, one way in which cetera may fail to be paria is that the right-making feature mentioned in the principle fails in some instance to be relevant.

[25] See, for instance, the "threshold default theory" of default reasons described in Horty (2007). This approach posits default generalizations with variable priorities assigned to them, where the priorities themselves have default values. In the context of a specific situation, the default priorities first get adjusted in light of the details of the situation, and then any default generalization whose post-adjustment priority is above a certain threshold gets "triggered." The reason-providing default generalizations, in a specific situation, are the triggered ones. As Horty points out, on this account, cases of silencing can be accommodated despite the central reason-providing role played by default generalizations. Silencing occurs when an otherwise applicable default generalization (e.g., "Ceteris paribus, causing pleasure is a good-making characteristic of an action") with an above-threshold *default* priority gets a below-threshold *contextual* priority, and hence does not get triggered. (Horty contests the putative cases of reversal advanced by Jonathan Dancy (1993, 2004). But genuine cases of reversal, if such there be, presumably would be a matter of a default generalization getting thus silenced, together with some additional attribute of the situation. Such an attribute might be the triggering of some other default generalization involving the feature in question, e.g., "Ceteris paribus, pleasure obtained from causing someone else pain is a bad-making characteristic." Or the additional attribute might be the presence of a "reverse default generalization" that has a below-threshold default priority but in certain special contexts gets an above-threshold priority—e.g., "Lying is a good-making characteristic," which could get an above-threshold priority in unusual situations such as games where lying is both acceptable and good strategy.) Jackson, Pettit, and Smith (2000) also purport to show how a generalist moral theory—specifically, a utilitarian theory—can accommodate silencing and reversal, but unfortunately their putative examples turn on a failure to distinguish fundamental principles from derivative ones. (In their account, features like the pleasure that an action causes the agent can sometimes get silenced or reversed, but only because in the specific situation, the presence of such features diminishes the net overall utility of the action. Maximal overall utility is here being treated as a descriptive feature that (i) is normatively basic rather than derivative, and (ii) is not itself subject to silencing or reversal.)

Fourth, there might even be some utterly exceptionless principles about all-in moral features like overall goodness, overall rightness, or overall moral permissibility. Candidate such principles are that *abortion is always morally wrong*, or that *abortion to save a mother's life is always morally permissible*.[26] Here the point to appreciate is that even if there are *some* exceptionless all-in moral principles, the full set of such principles could yet fall short—maybe very far short—of fully systematizing all of moral normativity. Even if there are such principles, cases that fall under them might be few and far between.

As far as we can see, the morals of the frame problem leave open, and leave viable, the possibility that moral normativity is partially systematizable by generalizations of any of these four kinds. Giving up full systematizability of moral normativity hardly means turning one's back altogether on a significant—albeit partial—systematizing role for generalizations.[27]

8 Moral Principles and Moral Authority

An important distinction to recognize, largely orthogonal to the distinctions canvassed in section 7, is between *summary* generalizations and *authoritative* generalizations. An authoritative generalization is one that is more fundamental, ontologically or justificationally, than concrete moral claims that are instances to which the generalization applies. Thus, for concrete moral claims that are true, their truth is explained by and thus justifiable by appeal to ('groundable by') the generalization.[28] A summary generalization, on the other hand, lacks any such ontological or justificational fundamentality, vis-à-vis the concrete instances that it subsumes. All it does is report a de facto distributional connection, across instances, between a certain descriptive characteristic and a certain moral-normative one.

The traditional hankering for exceptionless general principles that would fully systematize moral normativity has been, surely, a hankering for principles that not only would do that but also would be authoritative. The sought-for principles would be the

[26] Aristotle, who arguably repudiated the idea that morality is fully systematizable by exceptionless general principles, held that adultery is always morally vicious (*Nichomachean Ethics* 1107a 10-20).

[27] Partially systematizing generalizations also could play important roles in moral judgment-formation, including in ways other than by being consciously rehearsed and consciously applied to concrete cases; cf. Horgan and Timmons (2007). In this connection, we note that Hubert and Stuart Dreyfus (1990) argue on phenomenological grounds that much spontaneous moral behavior ("comportment," as they call it) occurs without any use of moral principles by the agent. But in our view this phenomenology-based conclusion is too quick, because (as argued in Horgan and Timmons 2007) there are good reasons to hold that spontaneous moral judgments often rest in part on moral principles that operate morphologically and implicitly in the cognitive system, rather than by being consciously rehearsed by the agent. (We call this view *morphological rationalism*.) Indeed, some of our arguments in support of morphological rationalism appeal explicitly to certain features of moral phenomenology itself.

[28] McKeever and Ridge (2006, pp. 12-14) call this the *truth-maker* conception of moral principles. Note that the truth-maker conception is orthogonal to the dispute between moral realists and moral irrealists: an irrealist, using the truth predicate disquotationally and deflationarily, can endorse the truth-maker conception—i.e., can maintain that whenever a concrete moral statement is true, the explanation for its truth is that it follows from some true conjunction of moral principles and descriptive claims. Note too that this sort of authority should be distinguished from questions about the *epistemic* priority ('authority') of moral principles vis-à-vis particular cases in the order of justification or knowledge. This epistemic issue, as we are here understanding it, has to do with whether being justified in believing, or knowing, a specific moral claim requires being justified in believing, or knowing, a moral principle, which then serves as (or is part of) an epistemically justificatory basis for belief in the specific moral claim. It might be the case, for example, that one can be epistemically justified in holding a specific moral claim on the basis of concrete moral experience (and thus without one's belief being based on moral principles), even if moral principles are what 'ground' in the sense of *make true* (together with relevant nonmoral facts) specific moral claims.

ontological and justificatory foundation of morals, with specific moral claims about concrete situations always being groundable and justifiable on the basis of these principles.[29] (The moral claims about a concrete situation would be logically entailed by the conjunction of the principles and sufficiently detailed descriptive characterizations of the concrete situation.) Call this conception of moral normativity *strong principle-based moral foundationalism*: it involves exceptionless generalizations as a foundation, and it conceives of the grounding relation as one of entailment. But of course, if there are no exceptionless moral principles that fully systematize moral normativity at all—as seems both possible and fairly likely, given the morals of the frame problem—then *a fortiori*, there are no such systematizing principles that are also authoritative. So strong, principle-based moral foundationalism is seriously called into question by the morals of the frame problem.

On the other hand, not only could there be general principles of various kinds (as canvassed in section 7) that partially systematize moral normativity, but there also could be principles of any of these kinds that also are authoritative. Such principles could still play a justificatory/grounding role in relation to concrete moral claims and judgments, even though collectively the principles would fall short of fully systematizing moral normativity and of constituting a full foundational basis for morals. A host of intriguing questions arise, in this connection, including these. (1) Is it the case that for any concrete moral claim that is true, the claim is groundable/justifiable by appeal to some principle(s), or is this only the case for some concrete moral claims? (2) What is the structure of principle-based grounding/justification, in cases where the principle is soft?[30] (3) Can a soft principle ground/justify a concrete moral claim even in a circumstance where that soft principle comes into conflict with other soft principles?

Addressing such questions is beyond the scope of the present paper. The main point we want to stress here is this: even if moral normativity is not fully systematizable by exceptionless general principles, and hence even if principle-based moral foundationalism is false, there might yet be plenty of grounding/justifictory work for moral principles to do anyway.[31]

9 Moral Particularism

The philosophical movement calling itself moral particularism has been quite influential in recent moral philosophy.[32] A variety of themes and doctrines have been espoused and defended under the banner of particularism, and several kinds of vagueness obscure the intellectual landscape. First, there is a certain degree of vagueness about which of various announced theses are supposed to be partly constitutive of particularism, and which are not. Second, there is vagueness about how various pertinent theses are supposed to be related to one another, logically or conceptually. Third, individual theses often are themselves formulated somewhat vaguely or metaphorically. The region of logical space called particularism is a dense and tangled thicket.[33]

[29] The principles might be self-evident truths or they might themselves be groundable by some further non-moral (though not necessarily non-normative) considerations.

[30] One structural possibility concerning soft principles, analogous to the structure proposed by Horgan and Tienson (1990) for deductive explanations adverting to soft laws, is a deductive justificatory argument containing an explicit premise to the effect that *cetera* are *paria* vis-à-vis the soft principle(s) invoked in the argument.

[31] We ourselves would maintain that there is plenty of grounding/justificatory work that moral principles *in fact* do, but that is not a claim we seek to defend here. For some defense, see Horgan and Timmons (2007).

[32] See, for instance, McDowell (1979), Dancy (1982, 1993, 2004), McNaughton (1988), and Little (2000).

[33] For some useful clearing of the conceptual terrain by way of helpful distinctions, see Chapter 1 of McKeever and Ridge (2006).

It does seem indisputable, however, that any position worthy of the label ‘particularism’ will at least embrace this thesis: There are no exceptionless general principles that fully systematize moral normativity. A corollary is that there are no principles that do so and also are morally authoritative; i.e., principle-based moral foundationalism is false. We will call the thesis, together with the corollary, *core moral particularism*. The morals of the frame problem, we take it, provide substantial support for core particularism. There *need not be* principles of the kind repudiated by core particularism, in order for humans to master moral normativity. And there *probably are not* such principles, since moral normativity very likely exhibits Quineian/isotropic holistic features.

A variety of intriguing, but somewhat obscure, articulations of particularist doctrine(s) offered by particularists themselves can be plausibly (if not mandatorily) construed as expressions of core particularism. This interepretation seems to fit reasonably well, for instance, with several ways that Little (2000) puts what she takes to be the basic idea of particularism: “There is no way of cashing out propositionally the ways in which non-evaluative properties contribute to the evaluative natures of situations, actions, characters” (p. 283). “The particularist’s claim is that the *good-making relation* cannot be cashed out in propositional form” (p. 285). “[Particularists] share the intuition that moral properties are, to use Simon Blackburn’s felicitous phrase, ‘shapeless’ with respect to the nonmoral” (p. 279). “To understand the real lesson of particularism is to understand that there is reason to doubt the existence of *any* codifiable generalities linking moral to nonmoral properties” (p. 288). On this reading, the propositional “cashing out” that Little speaks of is the full systematization of moral normativity via exceptionless general principles; likewise, “shapelessness”—the lack of “codifiable generalities”—is the absence of such principles.

On the other hand, particularists’ various articulations of their doctrine(s), including the articulations just cited from Little, also can be read as asserting not only core particularism but also various stronger negative claims as well—e.g., (1) that there are no exceptionless moral generalizations at all (not even contributory ones), and/or (2) that there are no authoritative moral generalizations. Stronger claims like these receive no direct support from the morals of the frame problem, as far as we can tell. And because these claims are stronger, they are thereby both more tendentious and harder to defend than is core particularism.

10 The Masterability Argument and Moral Particularism

We turn now to a well known critique of moral particularism: Jackson et al. (2000). Here are their most central remarks concerning the nature of particularism:

> The contention of the particularists is that, when given a list of conditionals of the form
>
> $$If\ D_i\ then\ E_j,$$
>
> where the D_i are various descriptive states of affairs in which some particular moral claim E is true, no matter how long and varied the list may be, we can never say ‘and so on’. The problem is not the term ‘and so on’ as such…. The problem, according to the particularists, is that there is no projectible pattern in the D_is to latch on to. There is no pattern *in the* D_i*s*, the grasping of which would enable you to write down new members of the list. (p. 83)

The heart of their objection is that if particularism were true, then moral normativity would not be masterable by finite creatures like humans. They say:

> There must be some commonality in the sense of a pattern that allows projection from some sufficiently large subset of the D_is to new members. If there isn’t, we finite

creatures could not have grasped through a finite learning process (the only sort there is) the predicate 'is right'. So, there must be a pattern or commonality—in the weak sense…of that which enables projection—uniting the set of right acts. (p. 87)

What should one make of this objection, given the morals of the frame problem? First a preliminary point. It is possible that there is a moral "universal grammar" that is innate in humans, and that moral normativity would be otherwise unlearnable. (Hauser (2006), and compare Chomsky's familiar "poverty of the stimulus" arguments in favor of innate universal grammar as a condition for the possibility of mastering the grammar of any extant human language.) But this possibility is really beside the point, because the core objection could be modified as follows so as to be still applicable, *mutatis mutandis*:

There must be some commonality in the sense of a pattern that allows projection from some sufficiently large subset of the D_is to new members. If there isn't, we finite creatures could not have evolved a finite cognitive architecture (the only kind there is) that subserves innate dispositions to form moral judgments employing the concept *rightness* in conformity to the full list of conditionals $D_i \rightarrow E_j$. (An innate *lookup table* would be out of the question, given the size of the list, because the architectural resource-requirements would be orders of magnitude too huge.) So, there must be a pattern or commonality—in the weak sense of that which enables projection—uniting the set of right acts.

How, then, to assess the objection, in light of the morals of the frame problem? We have two points to make about it. First, if the objection is construed as directed at *core* particularism, then it does not succeed. For, projection from some sufficiently large subset of the D_is to new members could well be possible, and hence moral normativity could well be masterable, even if moral normativity does not conform to exceptionless general principles. And, since moral normativity is very likely determined in part by holistic Quineian/isotropic features, core moral particularism is very likely true.

Second, however, there is a sound and important point at the heart of the objection—viz., that moral normativity must have some kind of structure that enables finite creatures like humans to project from some sufficiently large subset of the D_i's to new members. (Whether one calls any such structure a 'pattern' or not is a terminological issue, not a substantive one.[34]) It is an important—and empirically open—question what such structure might be. More specifically, it is an important—and empirically open—question whether, and to what extent, such structure needs to involve partial systematizability of moral normativity by generalizations of various kinds. Perhaps, in the end, the objection will prove fatal to certain radical versions of particularism—versions that deny even any significant *partial* systematizing-role for generalizations linking descriptive features to normative ones. But that remains to be seen, because in principle there might be projectible structure in moral normativity that is not a matter of systematizability by generalizations at all, not even partial systematizability.

[34] Dancy (1999) says, in the course of replying (promptly!) to Jackson, Pettit, and Smith (2000), "Particularism is the claim that there neither is nor needs to be such a pattern, not the claim that every such pattern is uncodifiable" (p. 60). Others might choose to use 'pattern' in such a way that any basis for projectibility counts as a pattern.

11 Classification and Reclassification of Moral Cases: A Neural-Network Model

Let us mention a recent piece of empirical research that bears upon this issue. Guarini (2006) describes experimental results obtained from a simple neural-network model of case-based moral reasoning. Guarini poses three questions for particularism:

- In those cases where particularism is alleged to be true, how do we learn to classify cases without grouping them under common principles?
- How do we generalize from cases we've learned in ways that let us classify new cases?
- How do we know when our initial classification of cases needs revision; and if it does, how do we do this? (2006, p. 22)

Guarini first experimented with a simple recurrent network for single-case classification. He reports:

> I trained it on cases that described either killing or allowing to die. Every case involves an actor (the person doing the killing or allowing the dying), an action (killing or allowing to die), and a recipient (the person being killed or allowed to die). The initial training set contained cases with at most one motive or one consequence…. Initially, I trained the network on 22 cases and tested it on 62 cases…. I arrived at the classification of those training and testing cases by polling 60 students in an ethics course.... Although the training set didn't include any suicide cases, the network generalized and replied plausibly to suicide and other cases.... When it comes to cases that combine an unacceptable motive with an acceptable consequence, the…network has problems…. This isn't surprising because the original set didn't include any cases that combined an unacceptable motive with an acceptable consequent…. When I expanded the training set to 28 cases (the original set plus six cases that included combinations of motives and consequences), the network was able to generalize to more cases. (pp. 23–26)

He concludes from these results that "classifying cases without executing or consulting moral principles is more plausible than it first appears" (p. 27). Thus, he takes the model to provide some support for particularism with respect to the first two of the three above-mentioned questions.

Guarini investigated the third question by focusing on the use of "contrast cases," as he calls them, in revising one's moral beliefs. He writes (p. 25):

> Humans use various strategies to examine their views on moral cases to consider the possibility of revision. One strategy uses contrast cases. For example, consider a… version of a case provided by Judith Thompson [viz., the famous case of the famous violinist]…. Thomson suggests that the case is analogus to a woman who becomes pregnant as the result of rape and is seeking an abortion…. Certainly, there are ways to challenge this analogy. One strategy is to find differences between the cases…. We could find two cases that are identical except for the feature we wish to test and then see if varying that feature makes a difference. For example, consider the following cases:
>
> - Case A. Input: Jill kills Jack in self-defense. Output: Acceptable…
> - Case B. Input: Jill allows Jack to die in self-defense. Output: Acceptable…
>
> …. Let's say we want to model the views of an individual who believes that it's morally permissible to walk away from the violinist but not to have an abortion in

cases of rape where the pregnancy results from rape. The respective representations for these cases are as follows:

- Case C. Input: Jill allows Jack to die; freedom from an imposed burden results. Output: Acceptable…
- Case D. Input: Jill kills Jack; freedom from an imposed burden results. Output: Unacceptable.

He then sought to see whether the original simple network could be embedded in a larger system, consisting of the original network paired with a "metanet," where the metanet's task is to identify the relevant contrast pairs (i.e., the ones that would be judged relevantly different by someone with the moral views being modeled). Here the results were discouraging. He reports:

> I trained the simple recurrent network so that the inputs for cases A through D yielded the stated outputs. I trained a second network to take as its inputs the first network's outputs and inputs. This second net (or metanet) takes pairs of cases together with their classifications (acceptable or unacceptable) as inputs. If the two cases are identical in all respects except one, and if the initial classification of these cases differ [sic], the output neuron fires to indicate that this case pair is a contrast pair of the desired type. The metanet flags case pairs where one feature appears to make a relevant difference between the two cases. I've trained and tested several configurations for the metanet, but I've had difficulty getting it to flag the desired contrast pairs with better than 60% success. (p. 25)

So his overall conclusion about the potential lessons of his network models is mixed: although classifying new cases without executing or representing moral principles is more plausible than it first appears, reclassification may very well require general rules—perhaps "contributory" rules such as this: "Killing always counts against doing an act, but letting someone die doesn"t count against doing an act" (p. 27). Commenting on the likely need to use contributory rules as a guide to potential reclassification, he says:

> [E]ven if we could scale up artifical neural networks and have them continue to learn and classify moral situations without consulting or executing moral rules, the process of reflecting on initial case classifications might involve such rules…. The challenge [for the use of contrast cases] is to find cases that are different in only one relevant respect. The question then arises: how do we determine what is relevant? For the method of contrast cases to work, we must be able to find out what is relevant. A generalist appears to have a straightforward answer: use any one of several contributory principles that state what is generally relevant for assessing moral situations…. If we bar principles from our account of moral cognition (as a thoroughgoing particularist would), then it's unclear how we could settle on a reasonable set of contrast cases. (p. 27)

We will conclude this section with a few brief thoughts about Guarini's model, his conclusions, and the pertinence of his results to the question (broached at the end of section 10) whether general principles of some sort are required in order for moral normativity to be masterable. First, caution is always needed in extrapolating from simple neural-network models—inter alia, because (a) a simple model might not scale up smoothly to more complex, human-like, processing, and (b) because a simple model might be operating just on the basis of statistical frequencies in its training set, rather than in a way that has anything much to do with the contents that the modeler has assigned (by fiat) to its input and output vectors. Second, although Guarini *asserts* that the simple recurrent network doesn't execute rules when it generates classifications, he has not precluded the possibility that it is actually executing rules procedurally. (Rules might end up morphologically embodied in the network's weighted connections, as the result of training—so that the network executes the rules without explicitly representing them.) Third, even with these first two caveats duly noted, it can be granted that his modeling results do provide *some* non-negligible evidence in support the conclusions he draws. Fourth, his conclusions appear not to conflict with core particularism, because they are consistent with the denial that moral normativity is fully systematizable by exceptionless general principles. However, fifth, his conclusions do lend some support to a key idea behind the masterability argument discussed in section 10—the idea that moral normativity, in order to be masterable by humans, must be at least partially systematizable by general moral principles.

12 Conclusion

In light of the morals of the frame problem, it is entirely possible—and also quite likely—that human moral normativity, although masterable by humans, nevertheless is too complex and nuanced to be fully systematizable by exceptionless general principles. Hence it is quite likely that strong principle-based moral foundationalism is false. Deployment of moral concepts thus requires certain skills at categorization and projection that outstrip full systematizability by exceptionless principles, and thereby have a particularistic aspect. Nonetheless, general moral principles might still have important roles to play, both in justifying moral judgments and in the psychology of moral belief-formation.[35]

References

Aristotle (2000) Nichomachean Ethics. Roger Crisp (transl and ed). Cambridge, Cambridge University Press

Dancy J (1982) Ethical Particularism and Morally Relevant Properties. Mind 92:530–547

Dancy J (1993) Moral Reasons. Oxford, Basil Blackwell

Dancy J (1999) 'Can the Particularist Learn the Difference Between Right and Wrong?'. In: Brinkmann K (ed) The Proceedings of the Twentieth World Congress of Philosophy, vol 1: Ethics. Bowling Green, OH, Philosophy Documentation Center

Dancy J (2004) Ethics without Principles. Oxford, Oxford University Press

Dreyfus H, Dreyfus S (1990) 'What Is Morality? A Phenomenological Account of the Development of Ethics Expertise'. In: Rassumssen D (ed) Universalism vs. Communitarianism. Cambridge, MA, MIT Press

Ewing AC (1959) Second Thoughts in Moral Philosophy. London, Routledge & Kegan Paul, LTD

Fodor J (1983) The Modularity of Mind: An Essay in Faculty Psychology. Cambridge, MA, MIT Press

Fodor J (2001) The Mind Doesn't Work That Way: The Scope and Limits of Computational Psychology. Cambridge, MA, MIT Press

Fodor J, Pylyshyn Z (1988) 'Connectionism and Cognitive Architecture'. In: Pinker S, Mehler J (eds) Connections and Symbols. Cambridge, MA, MIT Press

Gigerenzer G (2007) Gut Feeling: The Intelligence of the Unconscious. London, Viking Publishing

Gigerenzer G, Selten R (eds) 2002 Bounded Rationality: The Adaptive Toolbox. Cambridge, MA, MIT Press

Greene J (2007) 'The Secret Joke of Kant's Soul'. In: Sinnott-Armstrong W (ed) Moral Psychology, vol. 3: The Neuroscience of Morality: Emotion, Brain Disorders, and Development. Cambridge, MA, The MIT Press

[35] Thanks to Michael Gill, Neil Levy, Shaun Nichols, and two anonymous referees for helpful comments.

Guarini M (2006) Particularism and the Classification and Reclassification of Moral Cases. IEEE Intell Syst Mag 24:22–28 doi:10.1109/MIS.2006.76
Haidt J (2001) 'The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment'. Psychol Rev 108:814–834 doi:10.1037/0033-295X.108.4.814
Hauser MD (2006) Moral Minds. New York, Harper Collins Publishers
Horgan T (1997a) Connectionism and the Philosophical Foundations of Cognitive Science. Metaphilosophy 28:1–30 doi:10.1111/1467-9973.00039
Horgan T (1997b) Modelling the Non-computational Mind: Reply to Litch. Philos Psychol 10:365–371 doi:10.1080/09515089708573226
Horgan T, Tienson J (1990) Soft Laws. Midwest Stud Philos 15:256–279 doi:10.1111/j.1475-4975.1990.tb00217.x
Horgan T, Tienson J (1994) A Nonclassical Framework for Cognitive Science. Synthese 101:304–345 doi:10.1007/BF01063893
Horgan T, Tienson J (1996) Connectionism and the Philosophy of Psychology. Cambridge, MA, MIT Press
Horgan T, Timmons M (2006a) 'Cognitivist Expressivism'. In: Horgan T, Timmons M (eds) Metaethics after Moore. Oxford, Oxford University Press
Horgan T Timmons M (2006b) 'Expressivism, Yes! Relativism, No!'. In: Shafer-Landau R (ed) Oxford Studies in Metaethics, Vol. 1. Oxford and New York, Oxford University Press
Horgan T, Timmons M (2007) Morphological Rationalism and the Psychology of Moral Judgment. Ethical Theory Pract 10:279–295 doi:10.1007/s10677-007-9068-4
Horty J (2007) 'Reasons as Defaults', *Philosophers'*. Imprint 7(3):1–28
Jackson F, Pettit P, Smith M (2000) 'Ethical Particularism and Patterns'. In: Hooker B, Little M (eds) Moral Particularism. Oxford, Oxford University Press
Kahneman D, Tversky A (1972) Subjective Probability: A Judgment of Representativeness. Cognit Psychol 3:430–454 doi:10.1016/0010-0285(72)90016-3
Kahneman D, Frederick S (2005) 'A Model of Heuristic Judgment'. In: Holyoak K, Morrison R (eds) The Cambridge Handbook of Thinking and Reasoning. Cambridge, UK, Cambridge University Press
Kahneman D, Slovic P, Tversky A (1982) Judgment Under Uncertainty: Heuristics and Biases. Cambridge, UK, Cambridge University Press
Little MO (2000) 'Moral Generalities Revisited'. In: Hooker B, Little M (eds) Moral Particularism. Oxford, Oxford University Press
McDowell J (1979) Virtue and Reason. Monist 62:331–350
McKeever S, Ridge M (2006) Principled Ethics: Generalism as a Regulative Ideal. Oxford, Oxford University Press
McNaughton D (1988) Moral Vision. Oxford, Basil Blackwell
Ross WD (1939) Foundations of Ethics. Oxford, Oxford University Press
Shanahan M (2004) 'The Frame Problem', Stanford Encyclopedia of Philosophy (http://plato.stanford.edu/entries/frame-problem/)
Timmons M (2007) 'Toward a Sentimentalist Deontology'. In: Sinnott-Armstrong W (ed) Moral Psychology, vol. 3: The Neuroscience of Morality: Emotion, Brain Disorders, and Development. Cambridge, MA, The MIT Press