Practical Reasoning in a Modular Mind

PETER CARRUTHERS

Abstract: This paper starts from an assumption defended in the author's previous work. This is that distinctively-human flexible and creative *theoretical* thinking can be explained in terms of the interactions of a variety of modular systems, with the addition of just a few a-modular components and dispositions. On the basis of that assumption it is argued that distinctively human *practical* reasoning, too, can be understood in modular terms. The upshot is that there is nothing in the human psyche that requires any significant retreat from a thesis of massively modular mental organization.

1. Introduction

Can we explain, or make sense of, distinctively-human practical reasoning within a modular mental architecture? The question is worth asking, since there are good reasons for thinking that the human mind is massively modular in its organization. First, there is comparative, developmental, and neuro-pathological evidence that this is so (Shallice, 1988; Gallistel, 1990; Sperber *et al.*, 1995; Tager-Flusberg, 1999). Second, there are strong evolutionary and biological arguments for the conclusion that we should *expect* the mind to be modular (Tooby and Cosmides, 1992). And third (and most importantly), minds *must* be modularly organized if they are to be computationally realized, since computations have to be modular if they are to be tractable (Fodor, 1983, 2000; Bryson, 2000; McDermott, 2001). Since the computationalist assumption is easily our best (and perhaps our only) hope for understanding how minds can be realized in physical brains (Rey, 1997), this gives us a powerful motive for believing in massive modularity. Rather than giving up on computational psychology for the foreseeable future (as Fodor, 2000, urges us to do in the light of the alleged 'holism' of the mental), we should see just how far we can get in developing and defending the modularity hypothesis.

This paper is the second stage in a two-part investigation. In the first set of installments (Carruthers, 2002a, 2002b, 2003b) I investigated the extent to which we can make sense of distinctively-human *theoretical* thinking in modular terms,

I am grateful to the following for discussion of earlier drafts of this paper: Patricia Greenspan, Scott James, Christopher Morris, and an anonymous referee for this journal. I am also grateful to all those who participated in the discussion of this material following its presentation to the conference on *Innateness and Culture*, in Sheffield in July 2003, with special thanks to Richard Samuels.

Address for correspondence: Department of Philosophy, University of Maryland, College Park, MD 20742, USA.
E-mail: pcarruth@umd.edu

Mind & Language, Vol. 19 No. 3 June 2004, pp. 259–278.

arguing — albeit tentatively and speculatively — for a positive outcome. (I shall return to sketch some of the ingredients of that account below.) In the present discussion I turn to consider our capacity for *practical* reasoning, arguing that it, too, can be understood in terms of the interaction of a number of modular systems, together with a minimum in the way of further a-modular apparatus.

1.1 On Modularity

An initial problem that arises for any attempt to explain practical reasoning in modular terms is easily dealt with. This is the objection that practical reasoning cannot be modular, because if an organism possesses just a single practical reasoning system (as opposed distinct systems for distinct domains), then such a system obviously cannot be domain-specific in its input-conditions. For it will have to be capable of receiving as input beliefs and desires concerning all of the various domains that the animal is capable of representing and taking account of in its practical reasoning. Such a system could, nevertheless, be highly restricted in terms of its processing data-base, however; and this is all that is actually needed to secure its modular status in the sense that matters (Sperber, 2002; Carruthers, 2003a). Let me elaborate.

Those who believe in cognitive modules are apt to characterize them as *domain-specific systems* of a certain sort. And developmental psychologists, too, are now united in claiming (in contrast with the earlier views of Piaget and his followers) that human development is domain-specific rather than domain-general in character. This coincidence in terminology is unfortunate, and has encouraged a mistaken reading of the sense in which modules are necessarily domain-specific. When developmental psychologists talk about domains, they have in mind a domain of facts of a certain kind, or a certain class of contents. Thus mind-reading is a domain (dealing with the mental states of conspecifics), folk-physics is a domain (dealing with the mechanical properties of middle-sized objects and substances), folk-biology is a domain (dealing with generic relationships amongst living things), and so on. And modularists believe, indeed, that each these early emerging competencies is underpinned by a distinct cognitive module. But this is by no means essential to modularity as such.

The best way to understand the notion of modularity is by looking at the main arguments that have been offered in support of modular accounts of mind (Carruthers, 2003b, 2004). A module will be whatever those arguments would warrant. When we adopt this perspective, what we get is that a module is a processing system that might be innate or innately channeled in its development, that is targeted on some specific adaptive problem or task (its *domain*), and that is encapsulated *in its processing* from most of the information contained elsewhere in the mind/brain (thus enabling its operations to be computationally tractable). Sometimes the adaptive problem in question will correspond to a domain in the developmental psychologist's sense, such as mind-reading or folk-biology. But often it will not. And certainly there is nothing contradictory in the idea of a module that can take all sorts of different kinds of content as input.[1]

1.2 A Very Simple Practical Reasoning System

Seen in this light, then, there is nothing incoherent in the idea of a modular practical reasoning faculty. For example, we can imagine the following very simple practical reasoning module (some non-human animals might have a system of this sort). It takes as input whatever is the currently strongest desire, *P*. It then initiates a search for beliefs of the form, $Q \supset P$, cueing a search of memory for beliefs of this form and/or keying into action a suite of belief-forming modules to attempt to generate beliefs of this form.[2] When it receives one, it checks a motor-schema database to see whether Q is something for which an existing motor-schema exists. And if so, it initiates a search of the contents of current perception to see if the circumstances required to bring about Q are actual (i.e. to see, not only whether Q is something doable, but doable here and now). If so, it goes ahead and does it. If not, it initiates a further search for beliefs of the form, $R \supset Q$, or of the form Q (for if Q is something that is already happening or about to happen, then the animal just has to wait in order to get what it wants, it doesn't need to do anything more); and so on. Perhaps the system also has a simple stopping rule: if you have to go more than *n* number of conditionals deep, stop and move on to the next strongest desire.

Although I described this practical reasoning module as 'very simple' (in relation to the sorts of reasoning of which humans are capable), its algorithms would not, by any means, be computationally trivial ones. But each of the component tasks should be computationally *tractable* — at least, so far as I can see. (Intuitions of computational tractability are notoriously unreliable, however; so there is a promissory note here that eventually needs to be cashed.)[3] And we know that architectures of this sort are of very ancient ancestry.

[1] Sperber (1996) proposes, for example, that there is a dedicated *logic module*, whose job is to derive some of the more accessible logical consequences from any given set of beliefs as input. Such a system might be innate, and would be focused on a specific adaptive problem: namely, how to derive some of the useful logical consequences of what you already know. And it could be fully encapsulated in its processing. For in order to derive Q from *P* and $P \supset Q$ as inputs, such a system would need to consult no other beliefs whatsoever.

[2] The system might also employ a suite of heuristics (such as: if you want something, first approach it), as well as heuristics for information-search. For example, much of the literature on navigation suggests that children and other animals operate with a nested set of heuristics when disoriented. The sequence appears to be something like this: if you don't know where you are, seek a directional beacon (e.g. a distant landmark, or the position of the sun); if there is no directional beacon, seek to match the geometric properties of the environment; if geometric information is of no help, then seek a local landmark (Shusterman and Spelke, forthcoming).

[3] Note, too, that the notion of 'tractability' in question has to be relativized to brain processing-speeds, memory power, and the real-world time-scales in which computations have to be effected. Many problems that *would not* be characterized by computer scientists as 'NP hard' — i.e. intractable — will nevertheless be intractable in this sense.

The desert ant, for example, uses dead reckoning to calculate the direction of a food source in relation to its nest (integrating distance traveled in each direction with each angle of turn, and using information about the time of day and year, and the angle of the sun in the sky), and puts that information together with its current goals (to carry some food home, or to return to a previously discovered food source) in order to determine an appropriate course of action (Gallistel, 1990, 2000). Bees perform similar calculations, and integrate the resulting information, not only with their goals of returning to the hive or returning to a food source, but also with the goal of communicating that information to other bees (Gould and Gould, 1995).[4] There would therefore have been a very long period of time for computationally sophisticated, but nevertheless relatively simple, systems of practical reasoning to evolve (Carruthers, forthcoming).

Note that the sort of module described above would be input-unrestricted. Since almost anything can in principle be the antecedent of a conditional whose consequent is something desired (or whose consequent is the antecedent of a further conditional whose consequent. . .etc.), any belief can in principle be taken as input by the module. But what the module can *do with* such inputs is, I am supposing, extremely limited. All it can do is the practical reasoning equivalent of modus ponens (*I want P*; *if Q then P*; *Q is something I can do here-and-now*; *so I'll do Q)*, as well as collapsing conditionals ($R \supset Q$, $Q \supset P$, *so* $R \supset P$), and initiating searches for information of a certain sort by other systems. It can't even do conjunction of inputs, I am supposing, let alone anything fancier.

We can easily imagine a minor elaboration of such a system that would allow for the formation and execution of intentions for the future.[5] When the system reaches a doable Q for which there isn't presently an opportunity, but where it is known that such opportunities often arise, or where it is predictable that such an opportunity *will* arise, it then stores the conclusion, 'I'll do Q' in memory, together with some specification of the conditions necessary to do Q ('In circumstance C, Q is doable'). Then it later reasons: *In circumstance C, Q is doable*; *circumstance C obtains*; *so I'll do Q* (reactivating the previous decision). Again, the computations involved are by no means trivial; but so far as I can see, they should be tractable ones.

[4] Bees also integrate the information received from the dances of other bees with an existing knowledge-base; sometimes to the extent of *rejecting* information that they judge to be implausible. When one set of bees have been trained to fly to a food-source on a boat in the middle of a lake, other bees reject the information contained in their dances in favor of an indicated food-source that is an equal distance away on the *edge* of the lake. And it isn't that bees are simply reluctant to fly over water, either. When the indicated food source is on the other side of the lake, many bees will accept the information and choose to fly to it (Gould and Gould, 1995).

[5] Bratman (1987) argues convincingly that intentions are not reducible to combinations of belief and desire; and he shows how the adaptive value of intentions lies in the way that they facilitate planning in agents with limited cognitive resources, by reducing the computational demands on decision-making.

Note that, as a result of the limited processing of which the module is capable, there *is* a kind of de facto input-encapsulation here too. For although the system can receive any arbitrary belief as input, it can't actually do anything with that input unless it is a conditional belief with the right consequent, or a belief that is the antecedent of an existing conditional (or a belief that the circumstances necessary for the truth of an existing antecedent are likely to obtain in the future).

Would such a system deserve to be called a 'module', despite its lack of input-encapsulation? It seems to me plain that it would. For it could be a dissociable system of the mind, selected for in evolution to fulfill a specific function, genetically channeled in development, and with a distinct neural realization. And because of its processing-encapsulation, its implementation ought to be computationally tractable. In my view, this is all that any reasonable notion of 'modularity' should require (Carruthers, 2003a, 2004).

1.3 Can We Get From There to Here?

It is plain that human practical reasoning isn't at all like this, however. There seem to be no specific limits on the kinds of reasoning in which we can engage while thinking about what to do. We can reason conjunctively, disjunctively, to and from universal or existential claims, and so forth. And contents from all of the various allegedly-modular content-domains can be combined together in the course of such reasoning. This makes the practical reasoning system look like an archetypal holistic, a-modular (and hence computationally *in*tractable) central system, of just the sort that Fodor thinks makes the prospects for a worked-out computational psychology exceedingly dim (Fodor, 1983, 2000).

However, I have argued elsewhere that seemingly a-modular creative *theoretical* thinking might be constructable out of modular components with minimal further additions (Carruthers, 2002a, 2002b, 2003b). On the proposed account, a modular language faculty serves to link together the outputs of various central/conceptual modules, and makes possible cycles of processing activity by making non-domain-specific linguistic contents available to the central modules once again, in 'inner speech' (Carruthers, 1998, 2002b; Hermer-Vazquez *et al.*, 1999; Spelke, 2002). A computationally-simple supposition generator is built on the back of the language faculty, generating new sentences in ways that pick up on weak similarities and analogies, past associations, and so on (Carruthers, 2002a). And a sort of virtual faculty of inference to the best explanation can be constructed from principles involved in the assessment of linguistic testimony and the interpretation of speech, leading to a preference for internally generated sentences that are consistent, coherent, and fit the data, as well as being simple, fruitful, and unifying (Carruthers, 2003b).

Might human practical reasoning, too, co-opt the resources of this language-involving reasoning system? One fact that is especially suggestive, in this regard, is our tendency to convert desiderative contents into seemingly-descriptive ones. Thus instead of simply expressing a desire for some object or situation ('If only P!', or 'Would that P were the case!'), we tend to say (and think) that P would be *good*, that P is *important*, that *I want* P, or whatever. Instead of expressing desires in sentences with the same kind of world-to-mind direction of fit of desires themselves, we use indicative sentences with the sort of mind-to-world direction of fit appropriate for belief.

One plausible explanation of this otherwise-puzzling tendency — and the main hypothesis to be explored in this paper — is that by enabling motivational states to be re-represented as theoretical ones, it enables those states to be reasoned with using the resources of the language-involving theoretical reasoning system. I shall return to this suggestion in section 4 below.

2. Pre-linguistic Practical Reasoning

Let me back up a bit, however, and first ask what further assumptions can be made about the nature and powers of a modular practical reasoning faculty, prior to the advent of natural language and the language-dependent supposition-generator. In addition to the capacities described above – initiating searches for conditional beliefs about actions that would lead to the satisfaction of desires, as well as being on the look-out for circumstances that would activate previous decisions – what other powers might it have had?

2.1 Mental Rehearsal

One thing that immediately-ancestral practical reasoning systems would have had, surely, is the power to initiate episodes of *mental rehearsal*. This can probably come in two distinct varieties: simpler and more sophisticated.

The simplest form of mental rehearsal would be this. Once some sort of initial plan has been hit upon — *I want P*; *if I do Q I'll get P*; *Q is something I can do* — there would be considerable value in feeding the supposition that I do Q back through the various conceptual modular systems once again as input, to see if such an action would have other as-yet-unforeseen consequences, whether beneficial or harmful. This doesn't seem to require anything a-modular to be built into the system yet; for both the desire for *P*, and the belief that if Q then *P*, can be the product of individual modules.

What *is* presupposed here, however, is that representations within the practical reasoning system are in the right format for use in generating inputs to the various central-modular conceptual systems. This will be so if the practical reasoning system itself operates through the manipulation of sensory images; or if it has some way of mapping motor schemata onto such images, so as to generate visual images with appropriate contents that can then be consumed by the central modules in the usual way. (The 'mirror neurons' described by Rizzolatti *et al.*, 2000, might play such a role. These neurons in monkeys respond both when the monkey makes a given movement, *and* when the monkey sees that very same movement performed by another.) For central modules would, of course, already have been set up in such a way as to feed off perceptual outputs.

Given this sort of limited capacity to reason with suppositions, in the form of mental rehearsals of action-plans, we might expect that some animals would evolve a capacity for *creative generation* of such suppositions. With *P* desired and no obvious way to achieve it, the animal might try supposing that it does *R*, or that it does *S*, in the hope that one of these might bring about a doable *Q*, such that $Q \supset P$. (*R* and *S* might be elements from the motor-schema database, cued by aspects of the context, for example.) Such a system would still be fully modular, exploiting cycles of processing of existing modules.

A good deal of the evidence that has been cited — controversially — in support of chimpanzee theory of mind can also be used — much *less* controversially — to support the claim that this species of great ape, at least, engages in this sort of creative mental rehearsal. (I am quite prepared to believe that there is evidence of mental rehearsal from outside the great-ape lineage, too. But I shall not pursue the point here.) For example: a subordinate ape knows the location of some hidden food within an enclosure, and from previous experience expects to be followed by a dominant who will then take the food. So the subordinate heads off in the other direction and begins to dig. When the dominant pushes her aside and takes over the spot, she doubles back and retrieves and quickly eats the food.

Such examples are generally discussed as providing evidence that chimps can engage in genuine (theory-of-mind involving) *deception* — that is, as showing that the chimp is intending to induce a false belief in the mind of another (Byrne and Whiten, 1988; Byrne, 1995) — whereas critics have responded that chimpanzees may just be very smart behaviorists (Smith, 1996; Povinelli, 2000). But either way, it seems that the chimp must engage in mental rehearsal: trying out an alternative action in imagination (walking in the wrong direction and beginning to dig), predicting its effects (the dominant will follow and take over the digging), and discerning the opportunities for hunger-satisfaction that will then be afforded.

There is also direct evidence (of two distinct but related kinds) for a capacity for mental rehearsal in hominids dating from at least 400,000 years ago, at a stage when the language faculty had presumably not yet made its appearance. First, we know quite a bit about the cognitive requirements of stone knapping, both from painstaking reconstructions of the sequence of flakes (in those cases where not only a completed tool but also all the flakes that were by-products of its production have been found), and from the direct experience of contemporary knappers. And what we know is that stone knappers have to plan several strikes ahead, preparing a striking platform and so on, using variable and imperfectly predictable materials (Pelegrin, 1993; Mithen, 1996). Knapping cannot be routinized, and seems to require mental rehearsals of the form, 'If I hit it just *so*, then *this* will be the result; *that* would then enable me to strike the result *thus*, which would give me the desired edge'. The second source of knowledge is provided by the fine three-dimensional symmetries that begin to be produced at about this time. Wynn (1998) has argued convincingly that these require the knapper to visualize the results of a planned strike, and then to rotate the image mentally in such a way as to predict how the stone will then appear from the other side.

We know that members of these sub-species of *Homo* were pretty smart, colonizing much of the globe and thriving even in sub-arctic environments, while their brain-sizes approached the modern range (Mithen, 1996). And we can predict that their capacity for classifying items in the world must have burgeoned at around this time. So it may be that it was the increase in the number and sophistication of the concepts they employed that strengthened the back-projecting neural pathways in the visual system, that are used in normal vision to 'ask questions' of degraded or ambiguous input, and that are also the source of visual imagery. (On the latter, see Kosslyn, 1994.) In any case, whatever the explanation, we can be confident that these earlier hominids were engaging in sophisticated forms of practical reasoning involving mental rehearsal.

2.2 Somasensory Monitoring

It is also presupposed by the account I am sketching, of course, that the results of mental rehearsals can have further effects upon the practical reasoning system, leading to strengthenings or weakenings of desire. But this is very plausible. Besides engaging with the belief-generating modules, imagined scenarios will engage with the various motivational systems. Plausibly what happens next, is that the animal should monitor its somasensory responses to the imagined outcomes. All the animal then has to do is to monitor, not just its reactions to the thought of the desired state of affairs, *P*, but also to the various other states of affairs that it now foresees as consequences of the plan to bring about Q. A very simple mechanism can then be imagined that would sum across these various responses, either diminishing or extinguishing the attractiveness of the original goal, or strengthening it still further.

There is independent reason the believe that monitoring our own emotional and bodily reactions to imagined possibilities plays a crucial part in human practical reasoning. In the model provided by Damasio (1994), for example, we convert theoretical judgments of desirability into motivational ones by monitoring our own reactions to the thought of the circumstance described. His frontally-damaged patients can generally *reason* perfectly well on practical matters, making sensible considered judgments about what ought to be done, and when. But they fail to act on those judgments; and their lives as practical agents are a terrible mess. In Damasio's view, what has gone wrong is something to do with the somasensory self-monitoring system. (See also Rolls, 1999, for a related theory.)

Here is how one might flesh out this account a bit in modularist terms. Suppose that there are a whole suite of desire-generating as well as belief-generating modular systems. The former have been designed to take both perceptual and conceptual inputs of various sorts, combining them with knowledge stored in module-specific data-bases, to generate appropriate motivational states of varying strengths. When the supposition, 'Suppose I do *Q*', is received as input by these systems, they are keyed into action; and if Q is itself something desirable, an appropriate desire, conditional on the original supposition, will be generated. But the supposition that I do Q is also taken as input by the various belief-forming modules, some of which may generate further predictions of the form, $Q \supset R$. In which case R, too, can in turn be taken as input by the desire-forming modules, issuing in positive or negative (or neutral) motivational states, again still conditional on the supposition that I do Q.

Essentially what one might have, then, in the absence of a language faculty, are cycles of module-specific activity which support practical reasoning by mental rehearsal. We can imagine a variety of different algorithms for integrating the motivational results of such reasoning. The simplest would be summing the somasensory responses to the various possibilities considered. But one can also easily imagine a variety of more complex calculations, such as one in which a somasensory response is multiplied by the number of predicted items to be gained (e.g. five yams as against two).[6] And one can imagine a variety of ways in which considerations of time might be factored into such calculations (e.g. discounting distant satisfactions to some degree, or devaluing desires caused by currently perceived opportunities in such a way as to compensate for the relative vividness of perception as against mere imagination).

In what follows I shall be supposing, then, that prior to the evolution of a modular language faculty we might have had in place, both a capacity for mental rehearsal, and a disposition to monitor our own responses to the predicted results of such rehearsal, adjusting the level of our motivation to perform the action in question accordingly.

3. Adding Language and Normative Belief

It is then easy to see how things might go (and improve) when we add to the whole arrangement both a natural-language faculty and a language-dependent cross-modular *supposer*, or supposition-generating system (Carruthers, 2002a, 2002b, 2003b). The latter would vastly extend the range and creativeness of the suppositions whose implementation can then be mentally rehearsed.[7] And our reactions to those predicted effects could still be monitored, and would still have a tendency to create or diminish desires. We would be able to mentally rehearse scenarios that link together concepts or propositions drawn from a variety of modular domains, as well as genuinely-novel scenarios not inferable from previous

[6] We know that many species of animal can do approximate addition, subtraction and multiplication; see Gallistel, 1990.

[7] Since linguistic utterances are actions, we can think of the suppositional system as deploying the very same action-rehearsal mechanisms that were available pre-linguistically in the way described in section 2.1. But since sentences (when interpreted) have contents, rehearsing the utterance of a sentence will also be an entertaining of its content, and might have many of the same effects as supposing its content to be true.

experience, but perhaps suggested by weak analogies or similarities from the past, or from other domains.

I am suggesting, then, that we can come quite far in approaching distinctively human practical reasoning within a basically modular cognitive framework. There can be an encapsulated practical reasoning system that draws its input from a variety of belief-producing and desire-producing modules, and that can pass its output back as input to those modules once again in the form of suppositions, for purposes of mental rehearsal. And once a theoretical reasoning system is added to our cognitive architecture (consisting *inter alia* of a language module and a language-dependent creative supposition-generator), such mental rehearsals can become immeasurably more flexible and creative.

3.1 Weighing Goals

All of the above has been concerned with reasoning about *means*, however, not reasoning about *ends*. Cycles of mental rehearsal begin with the supposition, 'Suppose I did *Q*', and adjust motivations and plans accordingly. Yet humans don't *just* reason in this fashion. We also reason about ends. We reason about whether one end is *better* or *more important* than another, which ends *ought to be pursued* in a given context, and so forth. (And it is hardly very plausible that all of this should really be covert reasoning about means, where the overarching end is happiness, or maximum desire-satisfaction, or something of the kind.) How is such reasoning about ends to be provided for within a modular framework, without radically changing the powers of the practical reasoning system as such?

One obvious thing that humans can do in the course of practical reasoning that hasn't yet been allowed for in the simple model sketched above, indeed, is to weigh one goal against another. Is there any way in which a language-involving theoretical system could help with this? Well yes, there is. One way in which theoretical thinking might help me to adjudicate between the goal for *P* and the goal for *G* is by enabling me to think in more detail about what getting each would involve, and about their further consequences. Such thinking would provide me with more detailed *P*-involving and *G*-involving scenarios to feed forward through the motivational systems, and by monitoring my own reactions I can find myself responding to one scenario more favorably than the other.

Another possibility would be this. Suppose that I have two active goals in the present context, for *P* and for *G*. I can form from these the quasi-descriptive thoughts (in language), 'Getting *P* would be good', and, 'Getting *G* would be good'. I might then already have a stored belief that enables me to adjudicate between them, of the form, 'Getting *P*-type things is better than getting *G*-type things'. Or I might believe things from which I can infer something that adjudicates between them. (Such beliefs might be acquired by testimony from others, inculcated by moral teaching, or learned from previous experience.) This would then lead me to focus exclusively on the thought that getting *P* would be good. Imagining *P* and monitoring my own reaction, the desire for *P* is reactivated and now presented to practical reasoning as the only candidate for action. The search for ways of achieving *P* can then go on as before. Here the effects of theoretical thinking on practical reasoning would be by way of manipulating *attention*.

We have made a start on our problem, then. But other possibilities remain to be explored. One of these is that theoretical thinking about *norms* can lead to the creation of a new desire.

3.2 Normative Modules

A number of the modules postulated by evolutionary psychologists (and to some degree confirmed by later experimental work) are concerned with normative issues. Thinking about permissions, obligations, and prohibitions develops very early in young children, for example (Cummins, 1996; Núñez and Harris, 1998).[8] Moreover, they understand these notions, not just insofar as they pertain to themselves (in which case one might have postulated some sort of direct connection to the motivational system or the will), but also as applying to third parties. It seems inevitable, then, that we should think of the module in question as delivering *beliefs about* obligations and prohibitions. But these would be beliefs that would be frequently accompanied by the associated desires; and presumably the whole system would have been designed in this way. So when I believe that I am obliged to do something, I generally have a desire to do that thing in consequence. And when I believe that I am forbidden from doing something, I generally have an accompanying desire *not* to do that thing.

Similar points hold in connection with the social-contracts system — or 'cheater detection module' — proposed and investigated by Cosmides and Tooby (1992; Fiddick *et al.*, 2000; Stone *et al.*, 2002). This, too, seems designed as a belief-generating module, which operates not just in the first person, but also in the third person and from the perspective of another (Gigerenzer and Hug, 1992). Its job is to reason about social contracts in terms of cost/benefit structures, and in terms of who owes what to whom; one central concept of the system being that of a *cheat* — someone who takes the benefit in an exchange without paying the cost. And here, too, it only really makes sense that such a system should evolve, if it were to co-evolve with adaptations on the desiderative side of the mind, in such a way that one generally also has a *desire* to do one's bit in an agreed exchange, as well as a desire to punish or avoid those who have cheated on a contract (whether with oneself or with others).

From a modularist perspective it seems likely, then, that we have one or more modules designed for normative issues, which can straddle the belief/desire divide.[9]

[8] Núñez and Harris don't believe in an obligations module themselves. But if massively modular models of mind are taken for granted (in the way that I am doing in this paper) then their work is quite naturally seen as providing support for the existence of such a module.

[9] I shall leave it open whether the social-contracts system is a sub-module within a larger obligations/prohibitions system, or whether there are two distinct modules here dealing with closely related types of content.

It might be that such modular systems pre-existed the language module, or co-evolved with it, or both. But the selection pressures that led to the evolution of these systems would surely have been long-standing. We know that exchange and trading networks long pre-existed the appearance of modern *Homo sapiens*, for example, and we also know that these earlier hominids would have led complex social lives, in which the co-ordination of plans and activities would have been crucial (Mithen, 1996). Indeed, as Gibbard (1990 ch.4) emphasizes, it is problems of inter-personal co-ordination that create the main pressure for systems of normative thinking and speaking to evolve.

Moral beliefs form one sub-class of normative belief, of course. And a variety of different accounts are possible of the relationship between moral thinking and the sorts of modular systems envisaged above. On one view, for example, morality might have an independent source, being grounded in our natural feelings of *sympathy* when constrained by considerations of consistency (Singer, 1979). On another view, morality might be what you get when you combine an idea drawn from one of the above normative modules — that of *fairness* in an exchange — with thinking about the general systems of rules that should regulate human conduct (Rawls, 1972). And on yet another view, morality may result when more general normative thinking is combined with a certain distinctive sort of *affect* (Nichols, 2002). Moreover, once conducted in language, of course, or in cycles of linguistically-formulated 'inner speech', such thinking would be capable of uniting concepts across modular domains, as well as generating novel norms for consideration and evaluation through the activity of the suppositional system.

I should emphasize that the ideas sketched here are consistent with a variety of different positions concerning the nature of moral belief itself. Some such account as this ought to be acceptable to those who defend a sort of quasi-realism, or covert expressivism, about moral discourse, say (Gibbard, 1990; Blackburn, 1993). But it ought also to be acceptable to those who think that morality is more properly cognitive in character, and who perhaps see moral truths as *constructions*, grounded in the idea of a set of rules that no one could reasonably reject who shared the aim of free and unforced general agreement, for example (Rawls, 1972, 1980; Scanlon, 1982, 1999).

4. Theoretical Reasoning about Desires and Goods

The proposals on the table so far, then, include first, a flexible and creative theoretical reasoning system built out of modular components. (These would comprise a suite of belief-generating and desire-generating conceptual modules, a language module capable of integrating the outputs of the conceptual modules, a supposition-generator built on the back of the language system, a disposition to cycle natural language representations back through the whole arrangement, in 'inner speech', as well as dispositions to accept such sentences under conditions of 'best explanation'.) And second, the proposal includes one or more modules for generating normative beliefs, which can interact in complex ways with the theoretical reasoning system, and which also tend to generate the corresponding desires. So in these respects, at least, our theoretical thinking can become practical.

This cannot, by any means, be the whole story, however. For it is not just in respect of moral matters, or with regard to obligations and prohibitions more generally, that we are capable of practical reasoning of an unlimitedly flexible and creative sort. We can also reason in the same kind of flexible and creative way concerning things that we *want*. How is this to be accommodated within a modularist perspective, without conceding the existence of an a-modular, holistic, practical reasoning arena?

4.1 The Problem of Descriptive Goals

Let us return to the suggestion briefly mooted at the end of section 1. This is that one way in which language might be implicated in an enhanced (but still basically modular) practical reasoning faculty, could result from our disposition to express desires and intentions in descriptive form. This would enable them to be processed in the manner of theoretical ones, hence harnessing the resources of the language-involving theoretical reasoning system. Thus instead of just thinking longingly of some desired state of affairs, *P*, we are often disposed to think in descriptive mode, 'I want P', or, 'Getting P would be good'. These thoughts are then in the right format to be treated by a flexible and creative theoretical reasoning faculty. (See Gibbard, 1990 ch.5, for a related proposal.)

One significant problem for such an account, however, is the following. Systems of theoretical reasoning will be constructed so that *conviction* is transferred from premises through to conclusions. If we start from initial propositions that we believe to be true, and reason theoretically, then the result will be a further proposition in whose truth we have some tendency to believe. Now, the proposal under consideration here is that we can expand the powers of our limited-channel practical reasoning module by being disposed to convert motivational propositions into descriptive/theoretical ones, thereby harnessing the non-domain-specific powers of our theoretical reasoning system. But how are we to guarantee that transfer of *conviction* in an argument involving such covertly-desiderative propositions will also deliver a transfer of *motivation*?

Suppose that I start from some desired state of affairs, *P*. I then transform this into the descriptive statement, *P is good*, reason with it theoretically, and derive a further conclusion that I then have some disposition to believe, *Q is good*. But what then ensures that I translate this new descriptive belief back into a *desire* for Q? Without such a 'translation', the augmenting of the practical reasoning module by the resources of theoretical reason would be without any practical pay-off.

It looks, then, as if the proposal for the use of non-domain-specific theoretical reasoning to augment the powers of a limited-channel practical reasoning module might require a corresponding adaptation to the practical reasoning system.

Namely, whenever the latter module receives a descriptive-evaluative, but covertly desiderative, belief as input, it should generate the corresponding desire, and reason accordingly. While such an adaptation is no doubt possible, it would require a complex and messy interface between the two systems. For some way would have to be devised of identifying, from amongst the wider class of descriptive propositions, those that are covertly desiderative in content. Given the extensive range of evaluative predicates humans employ, this would be by no means easy. Moreover, such complexity would partly undermine the attractiveness of the original proposal, which was to explain how human practical reasoning can become non-domain-specific and inferentially flexible by co-opting resources that are already available.

4.2 Desiring to Do What is Best

More plausible, and more minimal, might be the following. We can postulate that an adaptation subsequent to the appearance of theoretically-augmented practical reason is an innate desire to do what one judges that it would be *best* to do. For in that case, when I conclude my theoretical reasoning about value with the belief that, all things considered, doing Q would be best, then this would create in me the corresponding desire to bring about Q.

We would still need a story about how such a desire could be selected for, however. So we need a story about how the use of theoretical reason to augment practical reason would have some advantages in the absence of such a desire, and yet still more advantages *with* such a desire. Telling such a story isn't trivial, by any means. But we have already made a start on it, in fact, through some of the proposals sketched in section 3.

Consider the attention-manipulating use of theoretical reason, for example. One might expect that the effect of such attention manipulation would be less than perfect. For often the original desire may remain salient, and hence continue to intrude in the process of practical reasoning. If our judgments of what is better than what are generally reliable (aligning themselves with goals of greater adaptive significance), then there might be pressure for the theoretical system to have yet more influence on the practical one. An obvious way to do that, would be to fix in place an innate desire to do what one judges to be the best.

Note that some such proposal can be rendered independently plausible through its capacity to handle the traditional philosophical problem of weakness of will. Sometimes we judge that, all things considered, it would be best to do Q, but we then go and do *P* instead. How is this possible? The answer comes readily to hand if a judgment of what it would be best to do is a mere belief, albeit one that is innately liable to generate a corresponding desire. For sometimes this tendency might fail; or the desire created might be of insufficient strength to defeat the desire for *P* in the competition to control the resources of the practical reasoning module.

4.3 From the Good to the Obligatory

There is another way in which theoretical reasoning about goods might become practical. As we noted in section 3, a number of belief-forming modules would seem to have — associated with them or built into them — connections to desire, in such a way that certain kinds of belief will normally give rise to a corresponding desire. It is possible, therefore, that as our theoretical reasoning abilities become enriched through the addition of language and language-based creative thinking, thereafter some of the theoretical reasoning that resulted could at the same time have become covertly practical, piggy-backing on existing connections between belief-forming modules and desire. To see how this can happen, we need to draw a distinction between the *actual* and *proper* domains of a module (Sperber, 1996).

The *proper* domain of a module is the task or tasks for which the system in question evolved. But the *actual* domain is the set of concepts/conditions that happen to meet the modular system's input conditions. For example, the proper domain of the system in human males that generates sexual desire from visual inputs would presumably have been the presence of a sexually receptive female. But the actual domain now includes paintings, photographs, videos, and much else besides.

In the case of the obligations/prohibitions module, the proper domain would have been the task of learning, conforming to, enforcing, and exploiting the norms that are prevalent in one's society or social group. But the actual domain might be much wider. So it could be that language enables us to feed additional creative or inter-modular contents to the obligations/prohibitions module, in such a way as to generate a desire out of a theoretical belief where none existed previously. Specifically, if we were disposed to convert evaluative statements of what it would be *good* to do into statements of what one *must* do, what one is *obliged* to do, or of what one *should* do, then this would meet the input-conditions of the obligations/prohibitions module in such a way as to generate the appropriate desire. And surely we do have just such a disposition. When reasoning about what it would be good for me to do, and reaching the conclusion that doing *P* would be best, it is entirely natural to frame such a conclusion in the format, 'So that's what I *must* do, then', or in the thought, 'So that's what I *should* do.'

A disposition of this sort would have the same sort of evolutionary rationale as the previous proposal concerning a desire to do what one judges best. For if our theoretical reasonings about value are reliable enough in evolutionary terms, then one might expect pressure for an evolved disposition to convert judgments of what it would be *good* to do or *best* to do into judgments of what one *must* do. This would enable them to harness the motivational powers of the normative reasoning system, in such away — again — that our theoretical reasoning can become practical. And it would be the functional equivalent of a disposition to desire what one judges to be best.

It begins to look, then, as if much that is distinctive of human practical reasoning might be explicable in modular terms. Such reasoning can happen via cycles of mental rehearsal and self-monitoring, and through a disposition to cast practical issues in theoretical language, hence harnessing the powers of our theoretical reasoning abilities (which are assumed, for these purposes, to be themselves constituted out of a suite of interacting modules). This would require the evolution of a further disposition, either to desire what one judges to be best, or to convert theoretical judgments of value into judgments of what one *must* do or *is obliged* to do — hence piggy-backing on the simultaneous theoretical/practical functions of a normative reasoning module — or both.

5. Coda: Desires *versus* Reasons

All of the above has been premised on the assumption that human practical reasoning has a belief/desire structure, however. I have assumed that human practical reasoning involves the integration of beliefs with desires to generate intentions or actions. But some have argued that human theoretical judgment can itself be directly practical. Some think that beliefs about what it would be *good* to do are directly motivating, without requiring the intervention of any desire (Dancy, 1993). Others have argued that it is our perception of *reasons* for action, rather than our desires, that provides the motivational element in practical reasoning — with desires themselves only being motivating to the extent that they involve the perception that we have a reason for acting (Scanlon, 1999).

One thing that is correct about such claims from a modularist perspective, of course, is that beliefs will generally be included amongst the inputs to our desire-generating modules. For example, coming to believe for the first time that a particular woman is my mother might cause in me feelings of sexual revulsion; and one might say, in consequence, that it is the fact that she is my mother that provides my *reason* for avoiding sex with her. So reasons for action can often be beliefs. But it is quite another matter to say that reasons can motivate *independently of desires*.

Moreover, from the perspective that we have adopted in this paper it is easy to see how someone might be tempted to think that beliefs can be directly motivating. For all that need be manifest to me through introspection is that I reach a theoretically-framed conclusion — 'Doing *P* would be best, all things considered' or 'I ought to do *P*' — and then I act or form an intention to act. We aren't normally aware of any separate desire to do what is best, or to do what we ought. But such a desire may be operative nonetheless. And something of this sort *must* be the case if we are to share the same basic belief/desire cognitive architecture that is common to the rest of the animal kingdom.

This is not the place to discuss the alternatives to belief/desire architectures in any detail. Let me just say that I regard them as ill-motivated. The considerations that are alleged to support them do not, in reality, do so (Brink, 1994; Copp and Sobel, 2002). And the authors of such proposals display an unfortunate disregard for the constraints on theorizing that should be provided by comparative and evolutionary psychology. Since we have good reason to think that the minds of animals have a belief/desire architecture, it ought to require some showing how a reasons architecture can be grafted onto the back of, or in place of, that. And there should be discussion of the evolutionary pressures that would have necessitated such wholesale changes in our mental organization. Not only is none of this to be found in the writings of the authors in question, but it is very hard to see how any such story might go.

On any account, of course, there are significant cognitive differences between ourselves and other animals. But any acceptable explanation of those differences should also preserve the extensive commonalities between us. And other things being equal, an account that preserves a basic common cognitive architecture while adding differences of detail (such as additional belief-forming or desire-forming modules) is surely preferable to one suggesting that the evolutionary transition from other animals to ourselves involved the creation of a whole new — reasons employing — cognitive architecture.

6. Conclusion

Massive modularity of mind is now routinely assumed by just about everyone working in the artificial intelligence community (Bryson, 2000; McDermott, 2001), and more generally amongst most of those who have confronted the question of how intelligent functions can be computationally realized. It is also assumed by many who come at the question of human cognition from a biological, neurological, comparative-psychological, or evolutionary-psychological perspective. But very few philosophers, and not many developmental or cognitive psychologists, are inclined to think about the architecture of the mind in such terms. I hypothesize that this is because the distinctive flexibility and creativity of the human mind, which is manifest both to ordinary introspection and to common sense, seems to resist explanation within a massively modular framework (Fodor, 2000).

What I have done in this paper is to make a start on the task of showing that distinctively human practical reasoning can be accounted for within a modular mental architecture. (This is provided, of course, that I am granted the assumption I have defended elsewhere, that distinctively human *theoretical* thinking is consistent with massive modularity.) But this has only been a beginning, and I have been able to provide only the merest sketch of an account. I certainly don't expect to have converted anyone to the modularist cause. In fact, I shall be satisfied if I have succeeded only in convincing you of the following: that it can't any longer be taken to be a mere truism that the human mind *isn't* massively modular.

Department of Philosophy
University of Maryland

References

Blackburn, S. 1993: *Essays in Quasi-Realism*. Oxford: Oxford University Press.

Bratman, M. 1987: *Intentions, Plans, and Practical Reason*. Cambridge, MA: Harvard University Press.

Brink, D. 1994: A reasonable morality. *Ethics*, 104, 593–619.

Bryson, J. 2000: Cross-paradigm analysis of autonomous agent architecture. *Journal of Experimental and Theoretical Artificial Intelligence*, 12:2, 165–190.

Byrne, R. and Whiten, A. (eds) 1988: *Machiavellian Intelligence*. Oxford: Oxford University Press.

Byrne, R. 1995: *The Thinking Ape*. Oxford: Oxford University Press.

Carruthers, P. 1998: Thinking in language?: evolution and a modularist possibility. In P. Carruthers and J. Boucher (eds.), *Language and Thought*. Cambridge: Cambridge University Press.

Carruthers, P. 2002a: Human creativity: its evolution, its cognitive basis, and its connections with childhood pretence. *British Journal for the Philosophy of Science*, 53, 1–25.

Carruthers, P. 2002b: The cognitive functions of language. And author's response: Modularity, language, and the flexibility of thought. *Behavioral and Brain Sciences*, 25, 657–719.

Carruthers, P. 2003a: Moderately massive modularity. In A. O'Hear (ed.), *Mind and Persons*. Cambridge: Cambridge University Press.

Carruthers, P. 2003b: On Fodor's problem. *Mind & Language*, 18, 502–523.

Caruthers, P. 2004: The mind is a system of modules shaped by natural selection. In C. Hitchcock (ed.), *Contemporary Debates in the Philosophy of Science*. Oxford: Blackwell.

Carruthers, P. forthcoming: On being simple minded. *American Philosophical Quarterly*.

Copp, D. and Sobel, D. 2002: Desires, motives and reasons. *Social Theory and Practice*, 28, 243–278.

Cummins, D. 1996: Evidence for the innateness of deontic reasoning. *Mind & Language*, 11, 160–190.

Damasio, A. 1994: *Descarte's Error*. London: Picador Press.

Dancy, J. 1993: *Moral Reasons*. Cambridge: Cambridge University Press.

Fodor, J. 1983: *The Modularity of Mind*. Cambridge, MA: MIT Press.

Fodor, J. 2000: *The Mind Doesn't Work that Way*. Cambridge, MA: MIT Press.

Gallistel, R. 1990: *The Organization of Learning*. Cambridge, MA: MIT Press.

Gallistel, R. 2000: The replacement of general-purpose learning models with adaptively specialized learning modules. In M. Gazzaniga (ed.), *The New Cognitive Neurosciences* (second edition). Cambridge, MA: MIT Press.

Gibbard, A. 1990: *Wise Choices, Apt Feelings*. Oxford: Oxford University Press.

Gigerenzer, G. and Hug, K. 1992: Domain-specific reasoning: social contracts, cheating and perspective change. *Cognition*, 43, 127–171.

Gould, J. and Gould. C. 1995: *The Honey Bee*. New York: Scientific American Library.

Hermer-Vazquez, L., Spelke, E., and Katsnelson, A. 1999: Sources of flexibility in human cognition. *Cognitive Psychology*, 39, 3–36.

Kosslyn, S. 1994: *Image and Brain*. Cambridge, MA: MIT Press.

Mithen, S. 1996: *The Prehistory of the Mind*. London: Thames Hudson.

McDermott, D. 2001: *Mind and Mechanism*. Cambridge, MA: MIT Press.

Nichols, S. 2002. Norms with feeling: towards a psychological account of moral judgment. *Cognition*, 84, 221–236.

Núñez, M. and Harris, P. 1998: Psychological and deontic concepts. *Mind & Language*, 13, 153–170.

Pelegrin, J. 1993: A framework for analyzing prehistoric stone tool manufacture and a tentative application of some early stone industries. In A. Berthelet and J. Chavaillon (eds.), *The Use of Tools by Human and Non-human Primates*. Oxford: Oxford University Press.

Povinelli, D. 2000: *Folk Physics for Apes*. Oxford: Oxford University Press.

Rawls, J. 1972: *A Theory of Justice*. Oxford: Oxford University Press.

Rawls, J. 1980: Kantian constructivism in moral theory. *Journal of Philosophy*, 77, 515–572.

Rey, G. 1997: *Contemporary Philosophy of Mind*. Oxford: Blackwell.

Rizzolatti, G., Fogassi, L. and Gallese, V. 2000: Cortical mechanisms subserving object grasping and action recognition. In M. Gazzaniga (ed.), *The New Cognitive Neurosciences*, 2nd Edition. Cambridge, MA: MIT Press.

Rolls, E. 1999: *The Brain and Emotion*. Oxford: Oxford University Press.

Scanlon, T. 1982: Utilitarianism and contractualism. In A. Sen and B. Williams (eds.), *Utilitarianism and Beyond*. Cambridge: Cambridge University Press.

Scanlon, T. 1999: *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.

Shallice, T. 1988: *From Neuropsychology to Mental Structure*. Cambridge: Cambridge University Press.

Shusterman, A. and Spelke, E. forthcoming: Investigations in the development of spatial reasoning. In P. Carruthers, S. Laurence and S. Stich (eds.), *The Innate Mind: Structure and Contents*. Oxford: Oxford University Press.

Singer, P. 1989: *Practical Ethics*. Cambridge: Cambridge University Press.

Smith, P. 1996: Language and the evolution of mind-reading. In P. Carruthers and P. Smith (eds.), *Theories of Theories of Mind*. Cambridge: Cambridge University Press.

Spelke, E. 2002: Developing knowledge of space: core systems and new combinations. In S. Kosslyn and A. Galaburda (eds.), *Languages of the Brain*. Cambridge, MA: Harvard University Press.

Sperber, D., Premack, D., and Premack, A. (eds) 1995: *Causal Cognition*. Oxford: Oxford University Press.

Sperber, D. 1996: *Explaining Culture*. Oxford: Blackwell.

Sperber, D. 2002: In defense of massive modularity. In I. Dupoux (ed.), *Language, Brain and Cognitive Development*. Cambridge, MA: MIT Press.

Stone, V., Cosmides, L., Tooby, J., Kroll, N. and Wright, R. 2002: Selective impairment of reasoning about social exchange in a patient with bilateral limbic system damage. *Proceedings of the National Academy of Science*, 99, 11531–11536.

Tager-Flusberg, H. (ed) 1999: *Neurodevelopmental Disorders*. Cambridge, MA: MIT Press.

Tooby, J. and Cosmides, L. 1992: The psychological foundations of culture. In J. Barkow, L. Cosmides and J. Tooby (eds.), *The Adapted Mind*. Oxford: Oxford University Press.

Wynn, T. 2000: Symmetry and the evolution of the modular linguistic mind. In P. Carruthers and A. Chamberlain (eds.), *Evolution and the Human Mind*. Cambridge: Cambridge University Press.