

Individual Minds as Groups, Group Minds as Individuals
Robert D. Rupert (University of Colorado, Boulder)

To appear in B. Kaldis (Ed.), *Mind and Society: Cognitive Science Meets the Philosophy of the Social Sciences*, Synthese Library Special Volume

I. Introduction

This essay addresses the following question: “What is required in order that a collection of humans, as a single entity, be in a mental state?” The approach taken is mostly exploratory and theoretical; I spend little time surveying empirical facts about groups, although these are by no means irrelevant. Instead, I focus largely on questions about mentality itself, or rather, in light of the contested nature of the category of the mental, questions about cognition. I ask “What makes a state or process cognitive?”, “What is the role of the individual subject in standard cases of cognitive processing – ones that involve only a single human?”, and “How could a group take the form of a single cognitive subject?”

The upshot is hardly decisive, but here is a sketch of the reasoning to follow. We discover what is distinctively cognitive by successfully modeling paradigmatic cases of intelligent behavior (*modulo* some methodological complications to be acknowledged below). Paradigmatic cases of intelligent behavior involve individual human organisms. The range of thus-far successful approaches to modeling in cognitive science (e.g., computationalist, connectionist, evolutionary-robotics based, and dynamicist approaches) consistently deliver one central distinction, between causal contributors to the production of intelligent behavior that lie within boundary of the relatively stable, integrated, persisting architecture and those that lie beyond that boundary. Thus, it appears that the boundary between the relatively persisting and integrated cognitive system and what lies

beyond that system offers the most plausible grounds for the distinction between what is genuinely cognitive and what is not.

The preceding line of thought demands that there be an individual of sorts – a self or subject – in order that there be cognitive states. What sort of self or subject? The relatively persisting, integrated collection of mechanisms that constitutes the cognitive system. This construct packs much less metaphysical punch than philosophers have typically associated with the self: there's no commitment to the self's possession of a single, rationally coherent set of beliefs or to a unified, consciousness-based perspective that plays a privileged role in the life of the self. In a way, then, my approach to individual cognition weakens the demands on groups; it renders more plausible the thought that a group might be an individual of the necessary sort to have cognitive states. What are the chances that groups do constitute such individuals, while also meeting other relevant conditions for being in cognitive states? How different might a group's structure be from the individual human's cognitive architecture while still engaging in cognitive processing? Given how likely it is that there be differences, it seems worthwhile to explore the idea that there are different *kinds* of cognition, and ask what it would take for a group to be a different *kind* of cognitive individual, but a cognitive one nevertheless. In the service of this discussion, I speculate about the nature of overarching or generic kinds, proposing a "tweak-and-extend" modeling-based criterion for the identification of different species of such overarching, or generic, kinds as *cognition*.

Here is one final preliminary point, a point that concerns consciousness, mentality, and my focus on cognition. Phenomenal experience is widely thought to be a central aspect of our mental lives (Chalmers 1996). There is little reason to think that groups

have such experiences, and thus, the question of group mental states might be quickly put to rest (Rupert 2005). Then again, the nature of such experiences is poorly understood, and, thus, it is not clear what methodology should be used to investigate group mental states of the sort that seem to presuppose robust conscious experience (cf. Goldstone and Gureckis 2009). Given this uncertainty, one way forward would be to establish group-level cognitive states and thereby show that at least one necessary condition for mentality in groups is met; and if it turns out that certain forms of cognition is all there is to consciousness – and that the so-called hard problem of consciousness is an illusion (Dennett 1988, 1991, Rupert forthcoming *a*) – this route to full mentality for groups seems especially promising (cf. Tollefsen 2006, 147). Alternatively, if it is shown that groups do not, as single units, instantiate cognitive states, then, on the plausible assumption that things with mental states almost certainly have cognitive states as well, we will have in hand a strong argument against group mentality. Consider, too, that if one inclines toward eliminativism about mental states (Churchland 1981), or even to more moderate positions, such as a revisionary reductionism or psychofunctionalism (Block 1980), investigation of cognition ascends to the paramount position – either because there are no mental states to investigate or because our scientific description of cognition will play a privileged role in the empirically based refinement of folk conceptions of mental states.

II. What Is Cognition?

Whether groups have cognitive states depends partly, of course, on the nature of cognition itself. A reasonable strategy for those interested in the possibility of group cognition recommends that we attempt to identify the nature of cognition, then let the

empirical chips fall where they may with respect to groups. Cognition, however, is a scientific kind, hypothesized by the relevant sciences to explain (by scientific, not folk, standards) a particular domain of phenomena; thus, when I talk of finding the nature of cognition, this should be understood in a metaphysically and methodologically minimal way: I am out to find the most promising theoretical characterization of it – most promising in light of our overall package of experimental results, successful models, intuitions about theoretical glosses of those models, results in closely related sciences, etc.; this is not an attempt to identify a metaphysical essence of cognition codified in a set of necessary and sufficient conditions that themselves constitute the everyday concept of cognition.

Given this naturalistic framework, let us ask a general methodological question: How do the sciences home in on nature's kinds and properties or select the language to be used in the formulation of their most successful theories?¹ The naturalist's way (Quine 1969, Quine and Ullian 1970) to answer this question would be to examine what appear to be our most successful scientific theories, checking to see how they were formulated and chosen over competitors. Here we might consider the circumstances surrounding the emergence of a new science, when theorists are faced with what they take to be a distinct domain of phenomena the investigation of which demands introduction of new vocabulary, concepts, principles, and laws.

¹ Hereafter, I omit qualifications meant to emphasize this paper's neutrality in respect of the issue of scientific realism and competing, more procedurally oriented interpretations of scientific enquiry. I ask instead that the reader take this neutrality as read, even when the particular choice of words suggests a thoroughgoing realism.

Successful scientific (sub)disciplines take shape largely in response to some range of observable phenomena that (a) are unexplained (not subject to prediction/retrodiction or manipulation and not subsumed under a general theory the principles of which have garnered scientific support from circumstances of successful prediction/retrodiction and manipulation) and (b) seem similar enough such that when they are ultimately accounted for, we suspect it will be by some common set of principles, theories, or models. What makes us think that given collection of phenomena are, in this way, of a piece with each other? It may be partly because we deploy the same everyday terms (e.g., ‘remembering’ or ‘intelligence’) in our descriptions of them or, in some cases, because we have had limited success predicting and explaining those phenomena using the same proto-theory or folk theory (for discussion of one pertinent case, that of a folk theory of mind, see Gopnik and Wellman 1992).

In response to such unaccounted-for phenomena, early thinkers in the field propose models of the mechanisms at work or hypothesize properties the interactions among instances of which account for the observed phenomena. The target of such modeling is likely, especially in the early days, to be a small subset of the phenomena in the presumed domain. Depending on the degree of success such modeling meets with, a given model, theory, or collection of models might be generalized to account for a wide range of the initial phenomena and, perhaps, after a flash of insight, to account for phenomena that weren’t thought to be of the same type as the phenomena to which the model in question was, initially, successfully applied; additionally, if an otherwise powerful model or family of models does not account for one of the phenomena originally thought to be in the relevant domain, we may expel that phenomenon from the domain. (Here’s a toy

example: if two chemicals feel to humans to be of different temperatures, but turn out to have such effect because of their chemical interactions with human skin – not because of differences in their mean kinetic energy – we will cease treating the difference between these chemicals as something to be explained by the same sort of models that explain the difference between warm water and cold water).

Thus begins a reciprocal dance: we recategorize phenomena into new similarity-groups based on the similarity of the models (the properties invoked by the models and actual relations claimed to hold between elements in the models) that can successfully account for at least some of them, while the constitution of what seem to us to be current similarity groups (including second-order similarity groups) guides our search for more general models that unify a broader range of phenomena. Along the way, relations of overlap between domains are discovered, single domains split into multiple ones, and the place in the enterprise of everything from individual bits of data to entire phenomena can be reconceived. (In the present context, it's important to keep in mind that the naturalistic project proceeds in this relatively holistic way, by the balancing of various considerations – theoretical, mathematical, empirical – against one another, with sometimes surprising results, for instance, that whales aren't fish. This is partly to say that whether groups have cognitive states cannot be decided just by noting that group activity doesn't [or does?] appear among cognitive science's central, motivating *explananda*.)

In the case of cognition, then, we should ask about the paradigmatic examples of the observable phenomena the collecting together of which gives rise to the very idea that there should be a science of cognition. Such paradigmatic phenomena appear primarily at

the level of the individual human organism: verbal communication, the writing of books, the production of art works, the formulation and testing of scientific theories, and so on.

What, then, is cognition in the individual case? In other work (2004, 2009, 2010), I have argued (a) that many successful research programs treat organisms as containing cognitive systems and (b) that a striking fact about cognitive modeling provides both the best explanation of (a) and grounds a theoretically based location-neutral account of cognition. The striking fact in question is that virtually all successful forms of cognitive modeling distinguish between, on the one hand, the persisting architecture, which is taken to have a relatively fixed number of elements (e.g., connectionist units) and stable relations among them (e.g., degrees of inhibition or ways in which the degrees of inhibition change over time), and on the other, more transient causal contributors that, together with aspects of the persisting architecture, produce intelligent behavior (cf. Wilson's [2002] distinction between obligate and facultative systems). Think of this as an inference to the best (available) explanation, twice over. First, the fact that the distinctive and central aspect of cognitive modeling – the persisting architecture – is typically instantiated within the organism explains why there's been as much success as there has been in doing "organism-oriented" cognitive science. Second, that the persisting architecture is the distinctively cognitive thing best explains why it runs through the various forms of successful modeling.

The question arises what sort of persisting integrated set of mechanisms is required for cognition. I've suggested a general characterization: that the relatively persisting, integrated cognitive system is the collection of mechanisms that co-contribute in overlapping subsets to the production of a wide range of forms of intelligent behavior.

Moreover, I've made a fairly concrete proposal about the sort of co-contribution that that serves as a measure of systemic integration among those mechanisms (Rupert 2009), which I summarize here:

1. For a subject at a time, form each non-singleton subset of the mechanisms that have distinctively causally contributed to the production of any form of intelligent behavior.
2. For each subset, relative to each form of intelligent behavior, there is, for each of its proper subsets, a probability of its being a causal contributor to the production of that form of behavior conditional on the complement of that set's being a causal contributor. (Here's an illustration: Take an edge-detection mechanism. It causally contributes to the avoidance of obstacles. So does a mechanism that detects distance from retinal disparity – Marr 1982 – and so does a mechanism that calculates shape from detected shading. Relative to this form of intelligent behavior – obstacle avoidance – each of these three mechanisms can appear in two different two-member sets, and each can appear in one three-membered set. For each two-membered set, two conditional probabilities are relevant: the first-mechanism's contributing conditional on the second's, and vice versa. For the three-membered set, there are six relevant conditional probabilities: each single mechanism's contributing conditional on the other two's, and each combination of two's contributing conditional on the third's. Now go through this procedure – in principle! – for every grouping of all causally contributing mechanisms relative to each form of intelligent behavior that's been exhibited by the subject in question [so long as the subject has exhibited a reasonably wide range of forms of intelligent behavior – for this very

richness of repertoire is one of the central features of the *explananda* of cognitive science].)

3. These conditional probabilities can be rank ordered.

4. Take the natural cut-off between the higher probabilities and lower ones. (If something's being an integrated system is a natural kind, we should expect such a statistically significant gap.)

5. For each mechanism appearing on the list of sets with higher conditional probabilities (that is, the sets above the gap), count the number of times it appears and rank order mechanisms (individually) accordingly.

6. A statistically significant gap separates those mechanisms that appear frequently on the first list from those that don't.

7. The integrated cognitive system comprises all and only those mechanisms the sum-scores of which appear above the gap.²

² In this context, it might be worth revisiting a worry about such entities as the sun; on the measure I've proposed to diagnose the scope of the persisting integrated architecture (2009, 42–43), something that consistently contributes, along with other mechanisms, to the production of a wide range of forms of intelligent behavior is almost certain to qualify as part of the persisting cognitive system. Objection: doesn't the sun fit that category?

I don't intend to defend this view; it may well be wrong, in that it might not capture what is common to the architectures presupposed by the various forms of modeling, but the point is that, from an ecumenical perspective, we should want to identify some common thread among the relatively persisting architectures – some abstract feature that they share – which can then be identified as the nature of cognition (or as a necessary feature

Doesn't the big bang? Yes, and although I've tried to avoid these consequences in various plausible ways, it may be best to rely on a "common element" strategy. All forms of models leave, for instance, the sun out of the architecture, and this gives us reason to toss out (by reflective equilibrium, one might say) some mechanisms that might otherwise seem to contribute causally in the way deemed adequate by my formal measure. The models all treat the sun as a background condition and that alone justifies treating it as such.

It might also be worth pointing out that my measure of the clustering of mechanisms (that is, of the scope of the integrated architecture) is consistent with a modular architecture (*contra* the suggestion made by Clark – 2011, 456 – and others). The measure is sensitive to the way in which factors co-contribute to the causal production of intelligent behavior, not to whether the factors causally interact with each other when producing that behavior. (One provocative way to think about the mechanisms in question is as local neural mechanisms that perform simple computational functions and contribute to the production of various forms of intelligent behavior partly by being assembled in different combinations for different purposes [see Anderson 2010], which might or might not be done in a functionally modular way.)

for something's being cognitive, *viz.* that be a state or activity of a part of a system that has the abstract feature identified).

Setting aside the formal measure of integration, I offer some clarificatory remarks and note some of the virtues of my systems-based approach:

First, as suggested above, I take it that, for most individual human subjects at most times, the persisting system is housed within the organism. Note that Otto and his trusty notebook (Clark and Chalmers 1998) can be described so that, by hypothesis, the notebook becomes part of Otto's cognitive architecture; that being said, bear in mind that Otto's case is make-believe, and so we should be careful not to let it – and other hypothetical cases like it – unduly influence our understanding of actual human cognition. The inferences to the best explanation offered above pertain to the explanation of actual data, results, and modeling practices.

Second, my strategy is to find a common thread among successful model-types, issuing from the entire range of orientations in cognitive science (including modeling that has sometimes been interpreted as extension-friendly): connectionist, computationalist, dynamicist, brute biological, robotics-based, and artificial-agent-based. If this “supervaluation” strategy is right, it provides powerful evidence in favor of the persisting-system-based view of cognition. Moreover, this evidence is of precisely the kind that Ross and Ladyman (2010) rightly demand. The evidence has nothing to do with intuitions about causality and constitution or everyday examples involving air conditioning or stereo systems, all of which they eschew. My reasoning appeals to real, scientific modeling, on the ground. Note, too, that contrary to what Ross and Ladyman suggest, the supervaluation strategy yields a fairly definite result: when one considers the

full range of successful models, one does not find a widespread, context-sensitive shift that sometimes places the privileged elements (the architectural elements) inside the organism (or agent) and sometimes out, when applied to actual cases.

Consideration of models and the relations between them raises an important concern about my strategy, about which I digress briefly. Consider the possibility that a coarse-grained model relate only a collective variable of a group to observable outcomes (e.g., the price at which company x's stock is being traded on Wall Street to the near-future profitability of company x – cf. Huebner, forthcoming), with some success. Why not think that here, we encounter a successful form of the modeling of intelligent behavior that draws a paramount distinction between the group and what is beyond the group's boundary? And, if there are such cases, why doesn't their existence establish that a group sometimes constitutes the cognitive system, at least if we're reading off the location of cognitive systems from the properties of our successful models?

In response, consider the following argument (cf. Giere 2006, 717–718): When a given *explanandum* can be successfully accounted for by both a less-articulated model and a more articulated one and the more articulated model accounts for more of the variance in the relevant behavioral data, we should derive our ontological conclusions (or the procedurally oriented equivalent) from the more articulated model. Our current situation would seem to satisfy the compound antecedent of the preceding conditional. On any particular occasion of successful prediction stock market's behavior is subject to two explanations: one is the group-level cognitivist explanation according to which the stock market – as single entity – sizes up the prospects of company x and makes a prediction, represented as the company's stock's selling price; the second explanation

appeals only to the various individuals who made decisions whether to buy and at what price and their reasons for doing so. In at least some cases, one would expect to account for more variance in performance of companies by appeal to the decisions of the best informed individuals – it is because *they* saw, say, the likelihood of the increased profitability of company x, and because others followed along with this judgment – than one does by appealing to a collective state of the group. (Even when this isn't the case, one can capture at least as much variance by attributing cognitive states to individuals and aggregating them – and the prices each pays – as one does when one, in contrast, takes the aggregated value – the selling price – to be the content of a group-level cognitive state; in which case, Occam's razor cuts away the group-level cognitive states, a matter discussed in more detail below.)³

Third, my approach is locationally neutral. I attempt to find a central and pervasive distinction present in a wide range of successful cognitive models, then check, afterwards, to see where the elements on either side of this divide appear. The division is not sought with malice aforethought (my initial reaction to the extended view was actually sympathetic, even though I distinguished it from the embedded view – see Rupert 2001, 505, n7); in fact, the distinction that seems to be most central to cognitive modeling in itself leaves wide open the possibility that elements on either side of the division fall on either side of the organismic boundary.

Fourth, the present exercise does not presuppose that a mark of the cognitive (Adams and Aizawa 2001) is needed to do cognitive science. Successful work in cognitive science helps us to see what the mark of the cognitive might be (or what might be

³ Thanks to Sydney Shoemaker and Peter Menzies for pressing me on this point.

necessary conditions for something's being cognitive), but we definitely do not need a mark of the cognitive in order to *do* cognitive science. The present issue is how best to interpret the cognitive science we have, that is, what theoretically or philosophically interesting gloss to give it.

Fifth, my proposal fulfills Mark Rowlands's demand for an owner of cognitive states (Rowlands 2009). Rowlands realizes that his various conditions for extended cognition lead to unacceptable results (cognitive bloat, in particular) absent a further constraint on cognitive extension: to be part of a cognitive process, it must be owned by a subject. More generally, philosophers of cognitive science will want a theory of the self, and my emphasis on cognitive systems suggests an empirically motivated one: the self is the cognitive architecture, or at least, a state's being a state of the self requires that it be a (nonbackground) state of one of the architecture's component mechanisms.

III. Cognition as a Generic Kind

On one reading of the preceding discussion, I've delivered all the self there need be for cognition to occur. And, given how minimal this self would seem to be, the proponents of group cognition might take heart. On the systems-based view, the individual human self would seem to be little more than a collection of "demons," all doing their individual bits to produce behavior, with no central controller or Cartesian theatre (Dennett 1991). Surely, in many extant groups, we find something like this – perhaps even a relatively persisting collection of mechanisms that contribute in overlapping subsets to the production of a wide variety of forms of intelligent behavior. Isn't that just what we see in, for example, ship navigation (Hutchins 1995)? Thus, some – perhaps many – groups

should meet the systems-based necessary (perhaps even sufficient) condition for group-level cognition.

Some readers may wonder, though, whether I've made matters too easy for the proponent of group cognition. My abstract characterization of an integrated cognitive system (including the formal measure of integration, or not) might seem dialectically well placed in the context of the debate about individual extended cognition: many of the external resources that contribute to the production of an individual's behavior probably do not, as mechanisms, score highly enough to count as components of that individual's cognitive system, nor are they, more generally speaking, part of the relatively persisting package of resources that contribute in overlapping subsets to the production of a wide variety of forms of intelligent behavior. Thus, thinking of inclusion in that set as a necessary condition does all the work it need as a prophylactic against (what might be perceived as) exaggerated claims to extended cognition.

Ultimately, though, much is missing from my systems-based characterization of the cognition (thus, my inclination to treat it as a characterization of only a necessary condition for something's being cognitive). What must be added? We simply don't know yet, because we do not know what precise shape the actual cognitive architecture takes. So, when we turn to groups, we find ourselves in an awkward situation: we might be confident that the group must meet the systems-based necessary condition toward which I'm inclined, but we also know what it should meet some as-yet-unspecified richer conditions – at least if our reasoning is to proceed from an identification of cognitive processing in the case of individual humans to the group case by the identification there of the same kind of process.

Let me illustrate how difficult it might be to establish that groups have cognitive states just by showing that they have the same architecture as exists in individuals. Fodor (1983) suggests that the human architecture consists of transducers together with computation-driven modules together with a central-processing unit or area that reasons holistically in a way that's sensitive to global properties of the set of the subject's belief set together with output systems. Now add some details to that characterization of the architecture. It has been claimed, for example, that the architecture of the human motor system includes emulator circuits; part of such a circuit receives an efferent copy of outgoing motor commands and runs a very fast simulation that, in effect, predicts whether the appendage in question will land where the motor command was meant to send it (Grush 1997). As such detail piles up, one begins to chafe at the chauvinism; perhaps the human architecture suffices for cognition, but must a group really meet all of these conditions in order to be cognizing? What if groups seem to engage in a sufficiently wide range of forms of intelligent behavior, but their group architecture doesn't match the individual human's? Must we say that groups don't have cognitive states solely on grounds of that mismatch? (And this gets especially tricky if group-level behavior doesn't actually seem so obviously of a piece with intelligent behavior in humans.) Why not say, instead, that the groups have cognitive states but that groups have a different kind of cognitive state from the ones humans have? (In this respect, the present discussion dovetails a long-running debate about functionalism and chauvinism at the individual level – Block 1978, Sprevak 2009, Rupert 2013.) Is this option available?

Tentatively accept that many aspects of group processing will not resemble processes that take produce intelligent behavior in individual humans. In the face of such an

assumption, I ask now what it would take for them to cognize nonetheless. What I'm after, then, more generally, are conditions for identifying an overarching natural kind – in this case, *cognition* – that has different species.

Some of Andy Clark's remarks, as well as some of John Sutton's, bear on this issue. On their view, the fact that human memories comprise a motley at the fine-grained level does not preclude the development of a proper science of memory (although note that Clark has since distanced himself from the underlying assumption of a motley: 2011, 452). When discussing the motley of causal processes involved in the production of intelligent behavior, Clark expresses hope for a science of cognition regardless of such disunity: "The study of mind might...need to embrace a variety of different explanatory paradigms whose point of convergence lies in the production of intelligent behavior" (2008, 95; see also, 2010, 64). And in advocating for an overarching, interdisciplinary approach to memory, Sutton suggests that we pursue all manner of memory-related phenomena, looking for "higher-level accounts which do find commonalities" (2010, 214) in spite of disunity at the nitty-gritty levels.

We should resist a purely outcomes-oriented characterization of cognition, one that simply lumps together all causal contributors to the production of any form of intelligent behavior. Not only is this theoretically unsatisfying and at odds with standard scientific practice (although one wonders about such cases as *reproduction* – does biology treat all processes that lead to reproduction as being of the same natural kind, regardless of how different they might otherwise be?); but it also seems ill-advised in the present context, given that typical groups engage in behavior that seems questionably intelligent (at least not clearly intelligent) and thus that we may need to rely on considerations of theoretical

unity (among the processes that produce clearly intelligent behavior in individuals and whatever processes produce not clearly intelligent behavior of groups) to decide the question whether the groups engage in intelligent behavior at all.

What, then, distinguishes a generic but still natural kind from a merely nominal one (even though the terms referring to such merely nominal kinds may nevertheless serve a useful purpose in scientific discourse)? Minimally, various species of natural, generic kinds must bear some kind of family resemblance to each other (cf. Wheeler 2011 on *cognition*), but I propose that it is not just any family resemblance: it is a family resemblance determined by the causal-explanatory roles of the components of the generic kind's instances; there must be a unity to various instances that is legitimated by theoretically important aspects of the relevant disciplines themselves. Each instance (or kind of instance) of a generic kind is, I maintain, constituted by a cluster of mechanisms; variations in the components of this cluster, from one species of the generic kind to another, and variations in the relations between these components, determine the fine-grained differences in the causal profiles of various species of the generic kind. Think in terms of partially overlapping models. The models of the way in which various species produce instances of the relevant explananda must have significantly overlapping elements and relations among them. This would seem to be the order of the day in most sciences; an initial (typically simple) model of some paradigmatic phenomenon succeeds (well enough), then related phenomena are modeled by the "tweaking" of the initial models—terms are added, parameter values adjusted, etc. If a phenomenon that might have been thought to be of a piece with the others turns out not to be amenable to this "tweak and extend" treatment, it is, and should be, treated as a different kind of

phenomenon after all; and this is when we say that original, full range of phenomena weren't all of the same kind—that is, there is no generic kind that subsumes them all.⁴

What might the tweak-and-extend in cognitive scientific practice? Consider, for example, Rumelhart, Hinton, and McClelland's (1986) now-standard presentation of the idea of a connectionist model (N.B. this is *not* meant to be an illustration of various models of *cognition* itself that might be produced, one from the other, by tweak-and-extend, but, rather, an illustration of how a family of scientifically important models are related to each other by tweak-and-extend). The authors lay out a number basic elements of a connectionist model and with respect to each of the elements, they describe ways in which related parameters can be varied, with the possibility that such tweaks will produce different kinds of behavior in the resulting network with the possibility that the connectionist modeling will cover (that is, be extended to) a wider range of forms of intelligent behavior.

⁴ Compare Fodor (1983, 101): "I have been arguing that the psychological systems whose operations "present the world to thought" constitute a natural kind by criteria independent of their similarity of function; there appears to be a cluster of properties that they have in common but which, *qua* input analyzers, they might perfectly well not have shared."

Translated to the current context: Various kinds of cognitive systems constitute a natural kind by criteria independent of the fact that they produce what appears to be intelligent behavior; there appears to be a cluster of properties that they have in common – revealed by the success in accounting for the relevant activities of the different systems by appeal to models that bear tweak-and-extend relations to each other – but which, *qua* producers of behavior that appears intelligent, they perfectly well might not have shared.

This rules out certain models from the connectionist category: one cannot without excessive contortions take the basic PDP framework and tweak it so that it becomes a look up table (one can simulate or implement a look-up table but that's another matter; in that case one is exploiting the resources of the connectionist model to create the effects that one would get from a system that could be more directly modeled with a different formalism). The boundary between models that are related by tweak-and-extend and those that are not is somewhat loose, admittedly, but bear in mind that my purpose is not to define what it is to be a generic kind. Rather it is to characterize diagnostic tools for identifying the presence of different species of a single generic natural kind as well as cases in which two different processes are not species of the same generic natural kind.

How might this criterion apply when anchored to a particular theory of the human cognitive architecture? Bryce Huebner (2008) has argued that the activity of Hutchins's sailors, who work together to take a large navy vessel through port waters, should be seen as the contribution of many different mechanisms each dedicated to the production of a particular kind of representation that (typically) varies along only one dimension, a representation which is then integrated with others to produce a final "authoritative" representation; and that this counts as the activity of a single group cognitive system; the sailors individually do work analogous to the work done by individual subpersonal components of the human cognitive architecture (such as when various subpersonal components of the visual system contribute to the recognition of a familiar face).

Huebner's general strategy seems clear enough and his analogy compelling. Assume now (as our anchor) the cognitive architecture of Fodor's *Modularity of Mind*. The boat-system surely has transducers, even by their strict definition, as mechanisms that convert

analogue signals to digital representations. But, when we reach modular input systems, one has to wonder whether enough of the features of the Fodorian construct will be satisfied. If, for instance, sailors can radically change their roles or their behavior at a single command, then they are, so to speak, cognitively transparent – quite the opposite of Fodor’s informational encapsulation – and their operation not at all mandatory. Moreover, if they can very well explain what they’re doing to the captain, then they don’t provide only limited access to their internal goings-on. One could claim that the system’s behavior can be accounted for by a model that bears the tweak-and-extend relation to a Fodorian model of human intelligent behavior, with the relevant parameter settings at zero in the case of input systems, but this looks disingenuous; modules as characterized by Fodor have some significant number of his nine traits to some significant extent, by definition, and a model without modules isn’t a Fodorian model – as a model without passing activation just wouldn’t be a connectionist model.

Now, I do not mean to endorse Fodor’s approach. Still, the discussion of Fodorian architecture should make clear what sort of demands must be met, according to my proposal concerning generic kinds, for a “nonstandard” system to exhibit a kind of cognition that’s different from human cognition. And, note that similar remarks apply to kinds of visual perception or motor control or any other overarching category relevant to the production of intelligent behavior.

IV. Group Cognitive Systems, Shadow Individuals

Let us now bracket worries about cognitive architecture; whatever the appropriate demands are, they have been met; in other words, assume now that certain groups do satisfy the criterion for being the relevant kind of cognitive individual. What else is

required in order that there be group cognition? The componential and determinative relations that hold between group cognitive architecture and individual human architectures give rise to one significant complication. In the typical case, a collection of individual human cognitive architectures construct, by their very cognitive states, the group and maintain it. There's an oddness to this relation, and so as to convey the oddness, I lead the reader through some twists and turns in the group-cognition debate.

Begin with a basic concern about causal-explanatory superfluity, driven largely by considerations of parsimony. Elsewhere (Rupert 2005, forthcoming *b*), I have pursued the following kind of argument against the existence of group cognitive states (cf. Wilson 2004, Ch. 12):

Premise 1. If it cannot be established that group cognitive states do distinctive causal-explanatory work – and there is no independent reason to think they exist – then we should tentatively conclude that they do not exist.

Premise 2. The antecedent of Premise 1's conditional is satisfied.

Conclusion. Thus, we should tentatively conclude that group cognitive states don't exist.

The form of parsimony that drives this argument is sometimes referred to as quantitative parsimony (Nolan 1997).⁵ This take on Occam's razor recommends that, once a given kind, force, or entity has been introduced, one not commit to more instances or individuals of that kind than is necessary for causal-explanatory purposes. When it comes

⁵ Thanks to Dave Statham for pressing me to clarify this (and for putting me on to Nolan's paper).

to courts, corporations, teams, unions, and the like, it would seem that the positing of group-level cognitive states violates just this principle of parsimony. For any action that, for example, the U.S. government takes, we can account fully for it (and for its intelligent nature, in particular) by adverting to the cognitive processes of the individuals involved; even if the individuals don't personally endorse the decision or the action, they commit to following certain decision-making procedures and carrying out certain attendant actions; they individually choose to live up to their commitments, and, voila, this provides all the cognitive materials needed to account for the outcome. There's no cognition left for the group to contribute to the production of the intelligent behavior in question.

Huebner (2008) objects that such an argument proves too much (also see Goldstone and Gureckis 2009), arguing that, if it effectively eliminates our commitment to group states, it also eliminates our commitment to personal-level states as well. After all, for any personal-level state one has – recognizing one's mother, for instance – there is a host of subpersonal representational states that subserves, and in some way determines, the presence of the personal-level states, perhaps such subpersonal-level states as appear in early visual processing, when cortex is computing the contour of a person's cheeks from the pattern of shading. If the human cognitive contribution to the production of group behavior renders superfluous group-level cognitive states, then so should the subpersonal contribution to the production of intelligent behavior render superfluous such personal-level states as recognizing one's mother. And, surely that's an indication that something's gone wrong with the parsimony-based argument.

Huebner's objection misses the mark. Notice that the parsimony-based argument doesn't rely on a distinction in kind between cognitive states of groups and those of

individuals; this is an intra-level question, not an inter-level one, which is obscured by ambiguity in the use of 'level'. Sometimes I use it in the sense of Oppenheim and Putnam (1958), to indicate something like the size of the things in question and something about the composition relations that hold between things of different sizes. A small portion of cortex might be of a lower level than a whole human organism, because the former occupies less space, to put it crudely, and similar remarks apply to the relation between an individual human organism and, say, a typical corporation or team. Looked at in this way, subpersonal, personal, and group states are all at distinct levels. The more relevant meaning of 'level' means something more like 'distinctive domain of properties', in which case there is in play only the single level of the psychological or the cognitive. Looked at from this perspective, a new set of properties introduced – intentional, representational, cognitive, psychological, whatever – and quantitative parsimony urges that we spread them around (with regard to one's commitments to them, of course, not the property instances themselves) in as stingy a manner as possible, with no regard to the areas of space time to which they're committed (other things being equal, a single thought instantiated in a large region of space-time is no less respectable than a single thought instantiated in a small region). So, it would be inappropriate for Huebner to cast his objection, as he sometimes seems to, as a *reductio* based on an analogy between inter-level relations. It's not as if I want to get rid of a level, and I do so by applying inter-level logic that just as clearly grounds elimination in other inter-level cases. No, my point is that we should appeal only to as many instantiations of cognitive properties as needed for causal-explanatory purposes.

The real question, then, has to do with the causal-explanatory need for group-level states – and personal-level and subpersonal states as well. We simply must ask which, if any, of these we need. We hold all to the same standard, and all up for grabs. With regard to the personal-level, as it's normally understood, I incline toward skepticism (Noë, 2004). To the extent that I can see a use in cognitive science for something like a person or self, it is the cognitive system. The ideas that a distinct personal-level must exist to guide cognitive scientific inquiry (Dennett 1987) or to maintain a certain pretheoretical picture of ourselves (that our consciousness is in direct relation to the objects of perception – McDowell 1994) strike me as dubious, perhaps even pernicious, grounds for preserving a commitment to a distinctively personal level. What is it that we need to say about the production of intelligent behavior that we can't sensibly say by adverting to the components of the cognitive system and their architectural features? I find nothing. True, folk psychology often works, as Dennett emphasizes, and we are inclined to think that it is *we* who see objects, not parts of our cortex, a fact that impresses McDowell. Nevertheless, folk psychology will account for less variance in the behavioral data than we get by invoking what Dennett might treat as mere subpersonal states; and here I would include the activities of language-handling mechanisms, which do much of our reasoning and control much of our reports about consciousness (as Dennett himself accepts). Dennett sees a distinctive value to generalizing over lots of differently constructed believers; but one who accepts the determinacy of externalist content for so-called subpersonal representations can capture many of these patterns simply by doing intentional psychology (cf. Fodor 1994, 1998). So far as I can tell, nothing is added to the project by saying that folk psychological states are states of the *entire person* and that

they, *by their very nature*, have certain relations to each other and to the environment; one can achieve all of the same causal-explanatory goals by appealing to metaphysically contingent regularities grounded in something like causal laws. And although I readily acknowledge the sorts of intuitions that move McDowell, reports of such intuitions provide *explananda* to cognitive science (at least to those cognitive scientists who care to study the psychology of philosophy) (Rupert forthcoming *a, c*). McDowell emphasizes the way in which whole creatures are suited to their environments, but I see no reason to think this cannot be understood as the suited-ness of cognitive systems – mere collections of co-contributing mechanism of certain sorts – to their environments (after all, why shouldn't selfish genes be “happy” to replicate via the construction of *these* systems as opposed to something more metaphysically robust?).

What, however, does all of this tell us about the utility of personal-level theorizing as it compares to the utility of group-level theorizing (I continue to talk in terms of levels, but bear in mind my earlier warning about such talk). What if, in the end, we accept none of the deep divisions between personal and subpersonal levels, of the sort that Dennett and McDowell embrace? What if, instead, we account for intelligent behavior in humans by appealing only to states of mechanisms that co-contribute in characteristic ways and govern *changes* in the patterns of behavior produced over time, partly by mutually governing each other's contributions – the sort of thing normally associated with development and learning? If this cognitive system is allowed to be a sort of self, then the proponents of group cognition might think they're off to the races. Surely some extant groups have cognitive systems related to ours by sameness in architecture or by having an architecture best represented by a model that bears a tweak-and-extend relation to the

best models of human architecture. And, if these cognitive systems survive parsimony-based skepticism, why not think group cognitive states have been vindicated?

We find ourselves in a strange position, then. The individuals in question – the group structures – seem to have no causal explanatory-work to do in the bringing about of behavior. What use are such shadow individuals, constituted as a by-product of the activities of, and relations between, other cognitive systems? Of course, we might ask the same of cognitive systems instantiated within individual organisms. I have argued that the distinction between cognitive systems – as relatively persisting architectures – appears as a constant feature across various kinds of modeling programs, but isn't that consistent with a kind of epiphenomenalism about the structured systems themselves? As indicated above, the persistence of cognitive systems helps to account for changes in the patterns of behavior; what we normally refer to as 'learning', 'memory', and 'development' are all patterns of change inside the cognitive system (changes that are determined partly by nonchanging features of the cognitive systems) that account for changes in the patterns of observable behavior (the child learns a language, for example). But, why, then shouldn't we appeal to changes in the group-level cognitive system to explain analogous longitudinal features of the behavioral output of the groups, finding just as much a causal-explanatory role for group cognitive systems as in individual human organisms'?

Shall we conclude, then, that groups are individual cognizers in just the same sense as individual organisms are? I'm inclined to resist such a conclusion, for the following reasons:

1. Consider the following disanalogy between group cognitive structures that may well meet formal, architectural requirements for being a cognitive system, on the one hand, and human cognitive architectures on the other. In the human case, there is a physically bounded entity that moves through the world (that is, which has spatio-temporal continuity as a whole and in its parts) and inside of which the persisting architecture (typically) resides, while in the case of many groups, this does not hold: corporations, stock markets, unions, and governments regularly change their membership in cognitively relevant ways. It's not simply a matter of losing neurons or taking in nutrients, as happens inside the individual human body; following Huebner's analogy, it's as if these groups are swapping out modules when human members come and go!

From the standpoint of the theory of cognitive systems, this difference appears to be important, for two reasons. First, we should want an independent criterion for the individuation of mechanisms (at least relative to cognitive scientific explanation – see Machamer, Darden, and Craver 2000). If cognitive systems are best characterized in terms of collections of co-contributing mechanisms, we should want some way to characterize those mechanisms that doesn't beg the question with regard to the question whether the same mechanism contributes to the production of a given form of behavior on different occasions; a physically oriented, lower-level (perhaps neural, in the human case) criterion of mechanism-individuation provides the best way to avoid such question-begging. (Consider, for instance, the difference between saying that language is “always in the air” as a cognitive resource and saying, in contrast, that a particular notebook with particular inscriptions is always available as a cognitive resource. We should want to be able to distinguish these cases and to register the significant difference in structural role

played by the two different kinds of resource. A physical or “lower-level” criterion of individuation allows for these distinctions to be drawn and differences to be explained – the notebook is a single “mechanism” in virtue of its being a spatio-temporally continuous entity; the various bits of language that enter an individual’s life are not.)

Second, and on a related note, one should wonder whether my diagnostic tool for identifying cognitive systems is too crude. A possible amendment might track the extent to which various mechanisms’ contributions have been tailored to each other’s contributions by the actual history of the interaction among *these* very mechanisms, physically individuated (Clark 1997) and how, for example, their patterns of co-contribution change over time (Klein 2010).

2. The problem of behavior returns as well. Many groups simply do not produce a wide range of forms of intelligent behavior. Many groups that make frequent appearances in the literature exhibit a high-degree of task specificity: the U.S. Supreme Court produces decisions, the stock market predicts companies’ future performance, a basketball team plays basketball, a group of sailors pilot, but in each case, the group doesn’t produce a whole lot else, as group.

Cognitive systems produce a wide range of forms of intelligent behavior; in fact, the characterization of any one form of a subject’s behavior as intelligent depends on the subject’s producing, or at least being capable of producing, a wide range of forms of the intelligent behavior. (To be more careful, there are sets of forms of behavior that individually become forms of intelligent behavior only when there is a significant number of such forms all at the subject’s disposal.) It’s the single subject’s being capable of

producing a wide range of forms of behavior that characterizes the very subject matter of cognitive science. (This complicates Robert Wilson's (2004) attempt to identify "minimal minds"; cf. Huebner, forthcoming.) It's one thing to be intelligent, but not to show it. It's plausible that some human cognitive systems have unusual biographies of this sort. In contrast, typical groups don't seem to have the right structure to exhibit a wide range of forms of intelligent behavior. (To the extent that they are flexible in this way, it has more to do with the flexible intelligence of the group members, who subvert, ignore, or redeploy standing group structures, than it does with anything about the group itself.)

3. Lastly, I focus – not decisively, but I hope fruitfully – on a philosophical puzzle to do with the hierarchies of natural kinds (or natural properties – I don't differentiate between the two; they are simply the kinds or properties of scientific interest that play a genuine causal-explanatory role, not merely an organizational role – see Rupert 2013). I propose to supplement the tweak-and-extend account of the unity of generic kinds, for principled reasons. Whether two states instantiate the same generic natural kind should also be constrained from below, as one might say, by the sharing of more determinate kinds (or all instantiating a significant number of kinds of those kinds. Individual humans instantiate *cognition* and instantiate a broad range of narrower kinds – perception, memory, language processing, etc. If Martians or groups of humans instantiate *cognition* of any kind (the same kind of cognition as individual humans instantiate or any other kind of cognition), then they must instantiate some substantial number of these more determinate kinds, and the model of shared overarching kinds applies here as well: Martian perception doesn't have to be the same as human perception, but in order that it

be a different kind of perception, the relevant behavior should be best accounted for by the application of models that bear a tweak-and-extend relation to the models that account for the relevant behavior in humans.

This addition of this requirement is not *ad hoc*. Think about the generic kind *animal*. There are substantially different ways of being an animal – being a mammal, being a fish, being a bird – but the category would seem useless if its various instances were not to share even variations on the less generic kinds such as *reproduction* and *metabolism*. This sort of judgment motivates the proposed constraint from below.

In the case of extant groups, it is very implausible that they instantiate a variety of the same kinds of kinds of cognitive states. That is, I suspect when the tweak-and-extend test is applied to group perception, group motor control, group language processing, etc., these will not be related as different kinds of the same overarching natural kinds – *perception, motor control, etc.* – that humans instantiate kinds of. In other words, we are likely to arrive at a very peculiar result, in which some extant groups would, by one measure, cognize, but without perceiving, remembering, exerting motor control, parsing language, and so on. I contend that the right response in such a fantastic case is simply to resist. A central part of the work done by the natural kind *cognition* is to unite various kinds of cognitive process: perception, memory, perception, reasoning, action, the giving and receiving of reasons, etc., and to organize and structure the interplay of these processes.

I should say, though, that I have only a modest amount of confidence that this final consideration by itself precludes typical groups from cognizing. Rather, it might be that I've merely identified weighty, substantial differences between the sort of cognizing that

humans do and the sort that groups do. This itself would be an intriguing result, however, and deflate many of the claims to do with group cognition. The idea that groups might cognize in just the way individual humans do or in a way very similar to the way in which individual humans do seems to motivate much of the interest topic of group minds and group cognition. If groups cognize only in the sense that they share some highly abstract structural properties with the human cognitive system (e.g., a pattern of co-contribution of mechanisms), but don't perceive, reason, remembers, etc., then the hypothesis of group minds or of group cognition lacks much of its intended appeal.

V. Afterword: Only Group Minds?

Questions to do with group cognition and group minds are partly conceptual, but they are also largely empirical, and we should be open to a wide range of outcomes. Here's an especially provocative one: that eliminativism about folk psychology (or some significant part of it) is true of individual-level mental states, but not at the group-level. After all, cognitive science might establish that intelligent behavior in individual humans is not produced by states that even roughly fit the folk descriptions of belief and desire (and their ilk) while at the same time showing (along with economics, sociology, etc.) that group-level behavior is produced by states that fit the functional profiles associated with our mental states as described by our everyday folk psychology.

Huebner (forthcoming) emphasizes the ways in which group members might intentionally construct social structures so as to make them mirror the kind of architectural structure one finds in humans. Imagine, though, that such purposeful construction succeeds in creating a group with the properties hoped for by those who organized it, but that the model for the group processing so constructed was an erroneous

theory of the human mind, endorsed by the individuals who structured the group. In that case, it might turn out that the group has mental states (its intelligent behavior is produced by states or processes that satisfy well enough the descriptions associated with ‘belief’, ‘desire’, etc.) even though individual humans don’t have any such states (because their intelligent behavior is not produced by states that satisfying such descriptions). Perhaps given the ever-increasing access to computing power, individual humans could structure a group so that it couches all of its “thoughts” in explicit symbol strings and has the computing power to check all of them for consistency. Assume for the moment that the transparency of logical relations among thoughts is central to our conception of beliefs and agency and also that, as a matter of fact, humans are in no position to test for such consistency in their own individual sets of beliefs (Cherniak 1986). This might deliver the peculiar result bruited above.

Many philosophers and cognitive scientists have emphasized the social nature of cognition (Hutchins 1995), which suggests a more fundamental connection between the idea of a mind and that of a group. Perhaps our very concept of mentality takes its character from our interaction with groups, and what generates the erroneous description of ourselves arises in the first instance from our acquisition of a certain concept from the observation of group dynamics.

These closing remarks have the flavor of musings, admittedly. I’m inclined to agree with Huebner that it is incredibly difficult to construct a group that satisfies the robust descriptions that the folk associate with human mentality, and one should also wonder how conceptions to do with consciousness might have been derived from the observation of groups. But, I also agree with Goldstone and Gureckis (2009) that rigorous

investigation of the possibility of group minds is in its early stages, and it is ill-advised to attempt to predict with confidence the outcome of the investigation of the relation between groups processes and those in such individual thinkers as are we.

References

Adams, F., and K. Aizawa (2001). The bounds of cognition. *Philosophical Psychology* 14, 43–64.

Anderson, M. L., (2010) Neural reuse: A fundamental organizational principle of the brain. *Behavioral and Brain Sciences* 33, 245–313.

Block, N. (1978). Troubles with functionalism. In C. Savage (Ed.), *Perception and Cognition. Issues in the Foundations of Psychology, Minnesota Studies in the Philosophy of Science*, vol. 9 (pp. 261–325). Minneapolis: University of Minnesota Press.

Block, N. (1980). Introduction: What is Functionalism? In N. Block (Ed.), *Readings in the Philosophy of Psychology*, volume one, pp. 171–84. Cambridge, MA: Harvard University Press.

Chalmers, D. (1996). *The conscious mind: In search of a fundamental theory*. Oxford: Oxford University Press.

Cherniak, C. (1986). *Minimal rationality*. Cambridge, MA: MIT Press.

Churchland, P. M. (1981) Eliminative materialism and the propositional attitudes. *Journal of Philosophy* 78, 2, 67–90.

Clark, A. (1997). *Being there: Putting brain, body, and world together again*. Cambridge, MA: MIT Press.

Clark, A. (2008). *Supersizing the mind: Embodiment, action, and cognitive extension*. New York: Oxford University Press.

Clark, A. (2010). Memento's revenge: The extended mind, extended. In Menary (2010), 43–66.

Clark, A. (2011). Finding the mind. *Philosophical Studies* 152, 447–461.

Clark, A., and D. Chalmers (1998). The extended mind. *Analysis* 58, 7–19.

Dennett, D. (1987). Three kinds of intentional psychology. In *The intentional stance* (pp.

43–68), Cambridge, MA: MIT Press. Reprinted from R. Healy (Ed.), *Reduction, time and reality* (Cambridge: Cambridge University Press, 1981).

Dennett, D. (1988). Quining qualia. In A. Marcel and E. Bisiach (Eds.), *Consciousness in Modern Science* (pp. 42–77). Oxford: Oxford University Press.

Dennett, D. (1991). *Consciousness explained*. Boston, MA: Little, Brown and Company.

Fodor, J. (1983). *The modularity of mind: An essay on faculty psychology*. Cambridge, MA: MIT Press.

Fodor, J. (1994). *The elm and the expert: Mentalese and its semantics*. Cambridge, MA: MIT Press.

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

Giere, R. (2006). The role of agency in distributed cognitive systems. *Philosophy of Science* 73: 710–719.

Goldstone, R., and T. Gureckis (2009). Collective behavior. *Topics in Cognitive Science* 1: 412–438

Gopnik, A., and Wellman, H. M. (1992). Why the child's theory of mind really is a theory. *Mind & Language* 7: 145–171.

Grush, R. (1997). The Architecture of Representation. *Philosophical Psychology* 10: 5–23.

Huebner, B. (2008). Do you see what we see? An investigation of an argument against collective representation. *Philosophical Psychology* 21: 91–112.

Huebner, B. forthcoming. *Macro cognition: A theory of distributed minds and collective intentionality*.

Hutchins, E. (1995). *Cognition in the Wild*. Cambridge, MA: MIT Press.

Klein, C. (2010). Critical Notice: *Cognitive Systems and the Extended Mind* by Robert Rupert. *The Journal of Mind and Behavior* 31, 3&4, 253-264.

Machamer, P., L. Darden, and C. Craver. (2000) Thinking about mechanisms. *Philosophy of Science*, 67, 1, 1-25.

McDowell, J. (1994). The content of perceptual experience. *Philosophical Quarterly* 44, 175: 190-205.

Menary, R. (Ed.) (2010). *The extended mind*. Cambridge, MA: MIT Press.

- Noë, A. (2004). *Action in perception*. Cambridge, MA: MIT Press.
- Nolan, D. (1997). Quantitative parsimony. *British Journal for the Philosophy of Science* 48 (3): 329-343
- Oppenheim, P. and H. Putnam (1958). The unity of science as a working hypothesis. In H. Feigl, M. Scriven, and G. Maxwell (Eds.), *Minnesota Studies in the Philosophy of Science*, vol. 2, Minneapolis: Minnesota University Press.
- Quine, W. (1969). Epistemology naturalized. In *Ontological Relativity and Other Essays* (New York: Columbia University Press).
- Quine, W., and Ullian, J. (1970) *The web of belief*. New York: Random House.
- Ross, D., and J. Ladyman (2010). The alleged coupling-constitution fallacy and the mature sciences. In Menary (2010), 155–166.
- Rowlands, M. (2009). Extended cognition and the mark of the cognitive. *Philosophical Psychology* 22, 1, 1–19.
- Rumelhart, D. E., G. E. Hinton, and J. L. McClelland (1986). A general framework for parallel distributed processing. In Rumelhart, D. E., McClelland, J. L., and the PDP Research Group (Eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vol. 1, *Foundations* (pp. 45–76). Cambridge, MA: MIT Press.
- Rupert, R. (2001). Coining terms in the language of thought: Innateness, emergence, and the lot of Cummins’s argument against the causal theory of mental content. *Journal of Philosophy* 98, 499–530.
- Rupert, R. (2004). Challenges to the hypothesis of extended cognition. *Journal of Philosophy* 101, 389–428.
- Rupert, R. (2005). Minding one’s cognitive systems: When does a group of minds constitute a single cognitive unit? *Episteme: A Journal of Social Epistemology* 1, 177–88.
- Rupert, R. (2009). *Cognitive systems and the extended mind*. Oxford: Oxford University Press.
- Rupert, R. (2010). Extended cognition and the priority of cognitive systems. *Cognitive Systems Research* 11: 343–56.
- Rupert, R. (2013). Memory, natural kinds, and cognitive extension; or, Martians don’t remember, and cognitive science is not about cognition. *Review of Philosophy and Psychology*, on-line first, DOI 10.1007/s13164-012-0129-9 .

Rupert, R. forthcoming *a*. Embodiment, consciousness, and the massively representational mind. To appear in *Philosophical Topics*.

Rupert, R. forthcoming *b*. Against group cognitive states. To appear in S. Chant, F. Hindriks, and G. Preyer (eds.), *From Individual to Collective Intentionality* (Oxford: Oxford University Press).

Rupert, R. forthcoming *c*. “On the sufficiency of objective representation.” To appear in U. Kriegel (Ed.), *Current Controversies in Philosophy of Mind* (New York: Routledge)

Sprevak, M. (2009). Extended cognition and functionalism. *Journal of Philosophy* 106, 503–27.

Sutton, J. (2010). Exograms and interdisciplinarity: History, the extended mind, and the civilizing process. In Menary (2010), 189–225.

Tollefsen, D. P. (2006). From extended mind to collective mind. *Cognitive Systems Research* 7 (2006): 140–50.

Wheeler, M. (2011). In search of clarity about parity. *Philosophical Studies* 152, 417–425.

Wilson, M. (2002). Six Views of Embodied Cognition. *Psychonomic Bulletin and Review* 9: 625–36.

Wilson, R. A. (2004). *Boundaries of the Mind: The Individual in the Fragile Sciences*. Cambridge: Cambridge University Press.