Group Minds and Natural Kinds*

Robert D. Rupert, U. of Colorado, Boulder

Draft, August 18, 2019

I. Introduction

This essay addresses the question whether groups of individual humans – for instance, courts, corporations, or labor unions – have mental states of their own, over and above those of their members. My methodological orientation is naturalistic, and thus I treat the matter as a question in philosophy of science, primarily in the philosophy of cognitive science.[1] Accordingly, I focus on such questions as "In virtue of what is a state or process cognitive?" and "What kind of evidence would support the claim that groups have cognitive states or engage in cognitive processing?" With regard to the latter question, I take particular interest in cases in which group processing differs significantly from processing in individual humans, asking what sameness of kind amounts to – or, rather,

---

* This paper descends from the previously circulated manuscript "Individuals as Group Minds, Group Minds as Individuals." Significant portions of this material were presented at Macquarie U., U. of Auckland, U. of Edinburgh, Ruhr U. Bochum, *Collective Intentionality IX* at Indiana U., *Trends in Interdisciplinary Sciences III*, Johns Hopkins U., and Humboldt U., and to New York U.'s Philosophy of Mind Discussion Group. Many thanks to numerous audience members and group participants for their helpful feedback. Special thanks in this regard to Georg Theiner.

[1] A large literature on group minds has been driven primarily by reflections on ordinary ways of ascribing mental states. For example, in describing her methodological assumptions, Margaret Gilbert says, "Suffice it to say that my own tendency is to go with everyday discourse, as I understand it. If such-and-such a phenomenon is referred to, seriously and literally, as belief, it is hard to argue that it is not, after all, belief as this is ordinarily understood" (2003, 103). I set this literature aside. I would like to know whether groups really are in cognitive states of their own, and to my mind, that requires making our best sense of scientific work on cognition, largely disregarding what the folk tend to think about matters mental and cognitive (acknowledging, however, that folk patterns of usage may, in the end, help to determine, not so much where nature's joints lie, but how to label the categories thereby carved out – see Stich 1996, chpt. 1). For similar reasons, I set aside empirical investigations of folk tendencies to ascribe psychological states to groups (Knobe and Prinz 2008, Sytsma and Machery 2009, 2010, Arico 2010, Huebner et al. 2010).

Note, too, that a complete naturalistic picture must take into account results from the social sciences, for the question at hand resolves into two main components: "What is cognition?" and "Do extant groups have cognitive states or engage in cognitive processing?" I here focus on the first component question, and thus my emphasis on philosophy of cognitive science rather than on philosophy of social science.

what sort of evidence would support the claim to sameness of kind – across such differences.

Certain trends in the cognitive and social sciences frame the discussion. Over the past three decades or so, it has become increasingly clear that distributed collections of processes produce a substantial portion of intelligent human behavior (Dennett 1991, Brooks 1999, Gendler 2008, Evans and Frankish 2009, Clark 1997). During roughly the same period, an increasing number of researchers working in the cognitive and social sciences, as well as in empirically oriented philosophy of mind, have argued that some extant groups of individual humans have mental or cognitive states of their own (Huebner 2013, Hutchins 1995, Goldstone and Gureckis 2009, Goldstone and Theiner 2017, Theiner and O'Connor 2010, Barnier et al. 2008, List and Pettit 2011, Schwitzgebel 2015). This appears to be no coincidence; recognition of the decentralized nature of human cognitive processing lends an increased air of plausibility to the thought that such distributed entities as courts, corporations, and citizenries might be in genuine cognitive states. In fact, some authors discuss the two questions explicitly in tandem (Hutchins 1995, Wilson 2004, Tollefsen 2006, Theiner, Allen, and Goldstone 2010).

The question naturally arises, then, whether these two lines of research – on distributed cognition in individual humans, on the one hand, and on group processes, on the other – investigate the same natural kinds, states, properties, or processes,[2] for example, *belief*, *memory*, and *cognition*. If groups do appear to have memories of their own, are those states of the same natural kind as the individual human states that we refer to as 'memories'? What does it amount to for there to be both an overarching kind

---

[2] For present purposes, I use these terms interchangeably, all providing a way of referring to joints in nature that the sciences attempt to identify.

memory as well as particular species of it? If there is group-level cognition, is it the same *kind* of cognition as individual humans engage in? And what, one might wonder, would it amount to for there to be different species of the genus *cognition*? How deep do sameness-of-kind relations run across the individual and group cases, and how should one should go about trying to determine how deep they run?

Here, in preview, is what follows. In Section II, I consider two arguments for realism about group states. I find both lines of reasoning inconclusive, largely because they proceed absent an account of sameness of natural kind and thus leave the central question unanswered. In Section III, I articulate a theory of the sort of evidence that could support claims to the sameness (or difference) of natural kinds (including, most importantly, thin, generic, or superordinate kinds) across cases. Elsewhere I've called this the 'tweak-and-extend' account (Rupert 2013). In order to apply the tweak-and-extend test, one must have in hand models of the data that are meant to be explained by the hypothesized instances of the kind in question. Therefore, in Section IV, I present what I take to be the best going model of the generic kind *cognition*, at least as it appears in the human case. If this view of cognition is correct, it removes a potential barrier to genuine group-level cognition; for, the view allows an individual with cognitive states to be decentralized. Section V pursues this thought, drawing on the material presented in Sections III and IV to develop a more definite version of the proposal that groups have cognitive states that are of a different kind of cognition from individual human cognition. Section VI canvasses objections and ultimately sounds a pessimistic note.

II. Realism, Reduction, and Functional Similarity

A view is realist with regard to group cognition if and only if, according to that view, (a) in a significant number of cases, an actual groups of humans, as a single entity, has cognitive states of its own and (b) such states are of the same relevant kind, qua cognitive states, as the states of individual humans. The kind of realism I have in mind does not rest on equivocation, metaphor, as-if attribution, or the deployment of homographs or "merely nominal" kind terms. The radical cachet of the view stems from its bold realism about many actual groups: when the sentence "England believes that withdrawing from the European Union is in its best interest" is truly asserted, that sentence is made true by the state of a single entity, according to the realist, viz. the United Kingdom's having a single belief of same natural kind as beliefs held by individual humans (whether for or against Brexit, or concerning any other matter).

In the present section, I consider two arguments for realism. The first has the flavor of an anti-reductionist argument (though see section VI for reservations about this characterization of it). The second focuses on the similarity of cognitive processing in groups to cognitive processing in the individual human. Neither argument clearly carries the day, or so I contend. Discussion of the arguments' weaknesses throws into strong relief the adjudicative potential of an account of the relation *being-of-the-same-natural-kind-as*.

1. Argument from naturalistic principles of ontological commitment

The first argument appears in a recent paper by Christian List (2018), which focuses primarily on the possibility of group phenomenal consciousness, a topic I here set aside. As a way to frame the discussion, List summarizes what he takes to be the strongest, extant argument for realism about group cognition,[3] an argument that he has helped to develop in a series of prior publications, perhaps most notably in a book co-authored with Philip Pettit (List and Pettit 2011):

> Premise 1: Our best social-scientific theories of certain social phenomena—for instance, our best theories of the behaviour of firms in the market place—attribute belief-desire agency of the functionalist kind to (some of) the collectives involved, often by representing them as agents in the decision- and game-theoretic sense.
>
> Intermediate conclusion: According to a naturalistic definition of ontological commitment, those theories are then ontologically committed to group agents.
>
> Premise 2: We should, at least defeasibly, take the ontological commitments of our best scientific theories in any given domain at face value.
>
> Conclusion: We should, at least defeasibly, take our best social-scientific theories' commitment to group agents at face value. (List 2018, 298)

And, to be clear, taking these commitments at face value commits one, at least provisionally, to the existence of genuine group agents with their own beliefs and desires (or judgements and preferences, as it is often put in List and Pettit [2011]).[4]

---

[3] List is a realist about group cognition, though not about group phenomenal consciousness.

[4] This represents only one of the two realist arguments developed by List and Pettit – their argument from positive social science (2011, 10*ff*.). List and Pettit also build a normative case for realism, which rests on such intuitions as that every good or bad action requires the apportioning of an appropriately corresponding amount of praise or blame; as List and Pettit see things, in some cases, the blame or praise properly distributed among individual members of a group does not sum to a sufficient amount given the nature of the act in question, and thus a group agent exists as something to be held responsible and given the remainder of the praise or blame (2011, 166). Such normative intuitions seem clearly out of place in the present context, the theme of which is, roughly, "what cognitive science tells us about realism," for such intuitions have no probative force in cognitive science; rather, they exist only as phenomena, that is, as data the production of which is to be modelled using standard cognitive scientific tools; and there is no reason to think cognitive science's best models of the production of such intuitions will include normative facts (in the philosopher's sense) as elements.

One might wonder, however, whether normativity plays a more direct role in cognitive science, by being presupposed, for instance, by Bayesian-based research programs (see, e.g., Griffiths, Chater, Kemp, Perfors, and Tenenbaum 2010). Such research seems driven by claims about how problems *ought* to be solved, viz. in a Bayes-optimal fashion. Note, however, that, in this research program, normativity does not play a role analogous to the role it plays in List and Pettit's reasoning. On List and Pettit's view, intuitions

Given List's reference to social-scientific theories and his appeal to a naturalistic principle of ontological commitment, the reader can safely assume that he has in mind natural, or scientifically respectable, kinds. If *Premise 1* is true and the successes at issue are genuine, then the kinds in question are, in fact, scientifically respectable. (In some cases, our best models are not terribly successful; but in the present context, charity demands that we understand something's being the best theory as entailing that it is at least moderately successful – successful enough to warrant ontological commitment.)

Of central concern in the present context is the question whether beliefs and desires as characterized by the social-scientific theories in question are of the same scientifically respectable state-types as the state-types that humans are typically in when they have what are commonly called 'beliefs' and 'desires'. It is hardly guaranteed that all applications of 'belief' and 'desire' refer to the same natural kind, any more than all

---

or judgments about normative matters – how much holding of responsibility is required relative to a given action – play the role of fixed premises. But, in cognitive science, Bayesian thinking plays a much more pragmatic role, providing inspiration for the development of models that are then evaluated by independent standards. Bayesian cognitive modelers attempt to match human performance by guiding the search for mechanisms that perform Bayesian inference "in a variety of implicit and approximate ways" (Griffiths et al. 2010, 362). More generally speaking, Griffiths et al. hope for a "synthesis with more bottom-up, mechanistically constrained approaches to modeling the mind" (*ibid.*, 362). On this view, "Probabilistic models are a tool for exploring different sets of assumptions about representations and inductive biases, making it possible for data to lead us to an account of human cognition" (*ibid.*, 363). Bayesian principles do not represent *a priori* or commonsense truths about human cognition, truths that constrain the selection of models in cognitive science (leading cognitive scientists to reject otherwise superior models of the data because they do not portray human cognition as optimally Bayesian). Rather, the use of Bayesian principles helps researchers to formulate possible models of the human cognitive process and understand why a process, so modeled, might be useful to the organism, even if actual human cognition is not Bayesian. For Griffiths et al., the Bayesian top-down approach is an empirical bet about methodology. They offer empirical arguments for their approach, for example, that Bayesian cognitive modeling is more likely to explore the space of possibilities effectively, and less likely to get bogged down in dead-ends, than is a mechanisms-first approach (*ibid.*, 358). Implicit in this style of argument is a commitment to contingency: if the competing, mechanisms-first approach produces models that account well for human behavior, while the top-down Bayesian approach flounders, Griffiths et al. will have lost their empirical bet and will abandon the Bayesian program. The attitude Griffiths et al. take toward the normativity presupposed by Bayesian modeling, treating it as a kind of tool that might be fruitful in the context of discovery or as a way of seeing the advantages of certain problem-solving strategies, seems well advised and to apply equally to the general question of how much normativity is presupposed by cognitive-scientific methodology. Thanks to Antonia Peacocke, David Chalmers, and Steven Gross for pressing me on normativity-related issues.

applications of 'jade' do (Putnam 1975). Consider that typical groups and typical individual humans differ enormously in ways that are, at least *prima facie*, psychologically relevant. Take, for example, matters to do with domestic life and interpersonal interaction. Groups do not post dating profiles on Web sites and make arrangements to join each other for coffee or drinks. Groups do not put their offspring down for naps, making sure first to change their baby-groups' diapers. Individual humans have characteristic developmental patterns and experiences that groups do not have, for example, the experience of learning a song about their first-language's alphabet or of coming to understand aspects of group dynamics by participating in organized sporting activities. And so on, throughout a wide range of humans' everyday affairs. It would be no surprise, then, to discover that the cognitive states operative in the production of a distinctively human life were of a different kind from the group-level states that produce, for example, court decisions, even if the latter states are sometimes called 'beliefs' and 'desires' and even if they play an important causal-explanatory role in the social sciences, and thus are not at all uninteresting.[5]

How shall the issue be decided? A convincing account of sameness and difference of natural kinds – or of what sort of evidence we might have for sameness or difference of two members of a natural kind – would provide one promising path forward. This is the subject of Section III, below.

[5] The point here is that group states might be of enormous scientific and philosophical interest even if groups have no cognitive or psychological states – or, to be more cautious, have no states of the same natural kind as the human states picked out when we use cognitive, psychological, or mentalistic terms. I include this qualification because my immediate concern is not to identify *real* cognitive, mental, or psychological states, where being the genuine article is thought to carry some sort of privilege or normative status, only in the realist's claim that the states of interest are of the same kind across the human and group cases.

To be fair, though, List might take the question about sameness of kind to have been answered. After all, List and Pettit go to some lengths to state and explain their account of what it is to be an agent: an entity is an agent if and only if it has representational states, motivational states, and the ability to combine these in appropriate ways to act on its environment (List and Pettit 2011, 20). This says, essentially, that something is an agent if and only if it has a belief-desire psychology, behaving in ways that obey, and can be predicted and explained by, applications of the practical syllogism.

This, however, is too schematic to be satisfying. It trades on the reader's familiarity with folk psychology, that is, with everyday explanatory practices of value in our interaction with conspecifics; but it does not address the difficult questions about genuine sameness of natural kind. In particular, it does not establish that (a) folk psychological kinds are natural kinds (that is, whether representational and motivational states, *as they're conceived of by the folk*, appear in any scientific theories successful enough to warrant ontological commitment)[6] or (b) that the states adverted to in List's Premise 1 are of the same kind as whatever sort of folk states are vindicated by cognitive science (or are of the same kind as whatever states cognitive science puts in place of folk psychological state-types). Cognitive science has produced accounts of motivational and representational states, but typically these characterize states, processes, and mechanisms not likely to appear in extant groups, and thus the success of such accounts hardly shows that the thin property *being a motivational state* plays a robust causal-explanatory role.

---

[6] I duly acknowledge (while also remaining skeptical about) Fodor's reasons for thinking that computational psychology vindicates an ontologically robust version of folk psychological theorizing (Fodor 1987, ch. 1; *cf*. Churchland 1981).

Let me come at this concern from a somewhat different angle. List and Pettit set their discussion of state-types in a theoretical context, a context indicated by List's Premise 1. They are functionalists about mental states (Putnam 1967, Lewis 1980, Shoemaker 1981). List and Pettit cannot be faulted for this; the functionalist approach to mental state-types holds the position of front-runner in philosophy of mind. But, this functionalist perspective also provides a perspicuous framework within which to pinpoint the lacuna in List's argument.

On the functionalist view, to be in a given type of mental state is to be in a state that plays a distinct causal-functional role in the subject's overall economy of states that mediate between input and output. To be in the sort of state-type that a human is in is to be in a state that holds the same place as the one held by the corresponding human state in a massively internally interconnected construct, involving hope, perception, fear, memory, dreaming, inference, etc. (or their cognitive-scientific descendants). How much detail should be built into this network is a matter of controversy (Block 1978), but this much is clear: List must identify a functionalist conception of belief and desire sufficiently fine-grained that the states in question play a significant causal-explanatory role (or appear as non-extraneous elements in our best, reasonably successful models of human behavior), but not so fine-grained that they clearly are not possessed by the typical groups of interest.[7]

Assume reasonably enough that the Ramsey sentence (Lewis 1970) of human psychology, as delivered by cognitive science (not folk psychology), contains at least a modest amount of detail. What, then, is the relation between it and social-scientific

[7] This line of argument parallels the dilemma presented in Rupert (2004, sections V–VIII) in the context of the debate about individual-level extended mind and cognition (Clark and Chalmers 1998).

models? It is highly doubtful that Ramsified social-scientific theory as a whole – a theory of political organizations or a theory of judicial organizations – has anything like the same structure of the Ramsey sentence of human psychology. In which case, given the holistic nature of functionalist state-individuation, humans and the groups in question share *none* of the relevant states. At the very least, we should demand a positive account of sameness of natural kind – beyond the bare comparison of our explanation of group behavior to folk-psychological belief-desire explanation – to support the realist's claim that the natural kinds in question are the same.[8]

What tends to obscure the problem at issue is the fragmentary nature of many influential models in the social sciences of interest to List, including typical game-theoretic or decision-theoretic explanatory structures, for example, those discussed in connection with the Prisoner's Dilemma (List and Pettit 2011, 107–108) or the mechanism for maintaining overall consistency in group-level decisions when confronting the Discursive Dilemma (*ibid.*, 45–46). Of course, many explanatory structures deployed in cognitive modeling are also mere fragments of psychological theory, but the holy grail of cognitive science is the construction of an architecture (e.g., SOAR, LEABRA, ACT-R, or GPS), within the context of which the various fragments will cohere as part of a single (ideally autonomous) agent – in which the language parser is functionally interwoven with the face-recognition mechanisms, and so on. Consider the claim that a corporation can act rationally in the market place, a claim that rests on a

[8] The device of Ramsification provides a formal tool for schematizing psychological theory. A Ramsey sentence does so by amalgamating or stitching together many successful fragmentary models (that is, models of specific data sets or models of families of related data sets). Overarching state-types – such as belief – might be made explicit in the relevant models and thus in the Ramsey sentence built from them; or such types may be only implicitly defined, say, by common structural patterns detectable in various parts of the Ramsey sentence at issue. In either case, the question remains open whether the types, so defined, are shared by groups and individual humans.

certain conception of what's optimized by the corporation's input-output function. Taken out of the context of the Ramsification of human psychology as a whole, this might seem to reflect the operation of the same kind of states as are at work when an individual human optimizes his performance with regard to, say, household finances. But, when set within the context of the Ramsification of individual psychology as a whole and the Ramsification of the theory of corporate structure as a whole, there would seem to be no merit to the claim of sameness of kind, given that the overlap between structures is likely to be very small and fragmentary. The particular group-level explanation in question relies on a collection of interacting states the behavior of which might match a small collection of states in the Ramsification of human psychology, and even then, only within a highly limited range. The matching of a small fragment across a highly limited range of cases is hardly grounds for asserting sameness of natural kind – at least not if one assumes, as List does, a functionalist conception of mental states. Anyone attempting to evaluate List's argument should want, at the very least, a more explicit account of the kind of evidence that would support his claim to sameness of natural kind across individuals and groups (that is, of the kind of evidence that would support the claim that the various uses of 'belief', etc. in question refer to the same natural kind across the cases of humans and groups) and an application of that standard to the matters at hand.[9]

---

[9] List and Pettit do not focus exclusively on agents' individual states, such as beliefs and desires. They frequently emphasize properties of the agent's psychology as a whole, such as consistency and coherence over time. They argue that in order to construct group-level institutions that serve our needs, we must carefully structure these institutions – by codifying the use of certain methods of aggregating individual humans' judgements and preferences – in ways that endow groups with such properties as consistency, coherence, and completeness (where to be complete is to be such that, for any of the relevant propositions, the group either endorses it or endorses its negation; List and Pettit 2011, 53), the idea partly being that such properties must hold of groups in order that belief-desire psychology be fruitfully applied to them.

List and Pettit do little, however, to support the claim that the properties in question in fact hold of individual humans. They appeal to intuitive conceptions of the role such properties play in human psychology and on claims about how we would like our institutions to function, without making a case that

2. Similarity-based arguments

Many authors have argued for realism by drawing direct comparisons to the case of individual humans. It is claimed that group-level processing involves the complex computational transformation of representations (Hutchins 1995), as is widely thought to be the basis of human cognition, or it comprises the operation of algorithms or mechanisms – such as lateral inhibition (Goldstone and Theiner 2017) – that play a significant role in models of human processing. It has seemed to many to follow from such similarities that groups have the same kinds of cognitive or mental states as humans (Huebner 2008, 109, 2013; Rovane 2014, 1663; Tollefsen 2006, 144; Theiner and O'Connor 2010, 89; Theiner et al. 2010, 382–383).

Such arguments would seem most convincing when they appeal to fine-grained similarities across humans and groups, similarities in the sorts of properties and processes of interest in the scientific study of human cognition. If we can identify similarities with respect to the very processes, structures, and kinds that do causal-explanatory work in the cognitive-scientific modeling of data involving intelligent human performance, we can be confident that we have found shared natural kinds of the relevant sort.

This, however, is a tall order, unlikely to be filled. As an illustration, consider the influential proposal that the architecture of the human mind contains specialized modules. On Fodor's view (1983), the human architecture consists of (at least) (a) peripheral transducers, (b) domain-specific modules (e.g., sensory systems), (c) a central-processing

---

the global properties we might thereby create in groups are of the same natural kind as global properties of individual human cognitive systems. Although some vague conception of rationality is likely to apply to humans and to certain groups, the questions remain open just how thin the shared kind in question is and whether a kind that thin does any causal-explanatory work.

unit that reasons holistically, and (d) output systems. This is a schematic start. But, of course, cognitive scientists spend most of their time filling in details of such architectural schemata. For example, with regard to output systems, it has been claimed the human motor system includes emulator circuits; part of such a circuit receives a copy of an outgoing motor command (a so-called efferent copy) and runs a very fast simulation that can, for instance, predict whether the appendage being moved will land where it was meant to land, all in time for a compensatory motor command to be sent, if the appendage does not appear to be on its proper path (Grush 1997). With regard to the fleshing out of the details of human cognitive processing, that is the tip of the iceberg. As such detail mounts, the idea that group processing might have the same structure as human processing – at the same level of grain typically of interest to cognitive scientists – seems incredible. Detailed profiles of human cognitive functioning fill millions of pages, tome after tome, journal article after journal article, cataloguing the often-quirky, highly specific workings of, for example, human memory and human vision. If one focuses on the sorts of states, processes, and causal profiles of interest to working cognitive scientists, similarity reasoning appears to be a dead end.

At the same time, as such details pile up, one begins to chafe at the chauvinism of it all; perhaps the human architecture – with all of its quirky, kludgy forms of processing – suffices for cognition. But, is it necessary? Must a group really work in *just that way* in order to cognize? Why not say, instead, that some groups cognize, but that it is a different species of cognition from human cognition?[10] To make such a case convincing, however,

[10] In this respect, the present discussion dovetails a long-running debate about functionalism and chauvinism at the individual level (Block 1978). On the role this issue has played in debates about distributed cognition and the extended mind, see Rupert (2004, 2013), Clark (2008), Sprevak (2009), and Wheeler (2011).

requires an account of when two kinds are species of the same overarching, generic kind, *cognition*.

III. The Tweak-and-Extend Account of Shared Kind-hood

What, then, distinguishes a generic but still natural kind from a merely nominal kind? Minimally, various species of a generic natural kind bear a family resemblance to each other (Wheeler 2011), but not just any family resemblance will suffice; there must, I contend, be a causal-explanatory unity to various instances of a kind. Claims to sameness of kind are most convincingly legitimated by a certain pattern in models of the phenomena of interest: our best models of the way in which various species produce instances of the relevant *explananda* must have significantly overlapping elements and relations among those elements. 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), and then related phenomena are modeled by the "tweaking" of the initial model – terms and model elements are added, parameter values adjusted, etc. – and the application of the family of models in question is thereby extended and perhaps sharpened (that is, made more accurate as models of systems to which their pre-tweaked cousins had been applied). If a phenomenon we might have thought to be of a piece with some 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; we should conclude that, pending further developments, there is no generic kind that subsumes all of the relevant cases.

How might the tweak-and-extend approach play out in cognitive-scientific practice? Consider, for example, Rumelhart, Hinton, and McClelland's (1986) schematic presentation of the idea of a connectionist model; it contains, among other things, units with individual levels of activation, arranged in layers, and that can pass activation forward to units in the adjacent layer, the activation levels of which are determined by a function that takes incoming activation (or inhibition) as input, and so on. The authors lay out these basic elements and, with respect to each of them, describe differential effects of, for example, different possible parameter settings or choice of activation functions. Different ways of filling in the schema produce different kinds of behavior in the resulting networks, introducing the possibility that one can construct an expanding family of interrelated models that accounts for a broader and broader range of forms of intelligent behavior (*ibid.*, 46*ff*).

What it is to be a connectionist model, then, is to be a model that instantiates Rumelhart et al.'s recipe or to models that themselves bear a tweak-and-extend relation to instantiations of Rumelhart et al.'s schema. And our best evidence that two kinds of process are both instances of *connectionist processing* – that is, that they share the natural kind in question – is that our best models of the two processes are both connectionist models. This approach rules out certain possibilities: one cannot without excessive contortions begin with an *ur*-PDP model (a straightforward instantiation of Rumelhart et al.'s schema) and transform it by tweaks into a look-up table.

I propose the tweak-and-extend approach as a general account of evidence relevant to judgements about the sameness of kinds, one that makes sense of work across the sciences. Take, for example, *being a harmonic oscillator*. One begins with a simple equation describing the behavior of an idealized pendulum; in an attempt to model a wider ranges of systems, one adds a term for friction, and then a forcing function, and so on. That this family of "tweaked-and-extended" models has fruitful application to a variety of physical systems is strong evidence that those systems instantiate a shared kind – *harmonic oscillator*.[11]

A full-dress presentation and defense of the tweak-and-extend approach would exceed limitations of space. Here I content myself with the discussion of one potential objection, to do with the boundary between models that are related by tweak-and-extend, on the one hand, and those that are not so related, on the other. How, one might wonder, does the view handle a long series of tweaks? Might it be easy enough to construct, in the incremental fashion of a sorites argument, a family of models interrelated by tweaks, two of the members of which, considered pairwise, seem quite unrelated to each other?[12] Perhaps the simplest reasonable approach to such cases would be defer to the intuitions of experts and practitioners, to tell us when a series of tweaks produces a result "beyond recognition."[13]

Ideally, though, we should want a more substantive way to separate apparent similarity from similarity that genuinely indicates sameness of kind. This might involve identification of aspects of the data variation in which is accounted for by two models' shared components (that is, whatever remains common to them after the tweaking required to construct one model from the other). The method might take something like the following form: take two candidate models; identify their shared elements or structure; vary systems that those models have successfully targeted in the past along the

[11] For a further example, see Chemero's discussion of Haken-Kelso-Bunz model and its various extensions and refinements (2009, 85–96).
[12] Thanks to Jackson Kernion for pushing me on this point.
[13] Although one might wonder about the extent to which such judgements will vary with the particular context, with, for example, the purpose the expert has in mind when asked the relevant question (*cf*. Block 1997).

dimension that the models' shared components were mapped onto in the cases of previous successful applications; see whether corresponding variations in the shared components of the models account for variance in other measurable quantities of the target systems (variance that accompanies variation in the part of the target systems that corresponds to the models' shared component). The idea is to uncover evidence that what is shared by two candidate models plays a causal-explanatory role or whether it is, in some way, incidental. If it does play a causal-explanatory role, then we place the candidate models in the same family, even if, *prima facie*, the extent of the tweaking and extending involved renders the models distant cousins.

Note that even if a shared element does causal-explanatory work, the question remains how best to individuate the natural kind for the presence of which evidence has been provided. The relevant aspects of the target system (the aspects that correspond to the shared component in question as well as those variance in which is accounted for by changes in the value of the shared component) should guide our characterization of the kind. If, for example, we are to treat the models' shared component as indicative of the presence of the generic kind *cognition*, then the data in question had better be relevant to what we take to be cognitive phenomena, not to, for instance, overall size or mass of the target systems.

The preceding is, by needs, sketchy and incomplete. Bear in mind, however, that my purpose is not to present a theory of what it is to be a generic kind, but to articulate a more structured framework within which to attempt to construct a naturalistic argument for realism.

IV. What Is Cognition?

Whether groups have cognitive states – that is, are in states of the same natural kind as human cognitive states – depends partly, of course, on the empirical facts about cognition itself. Cognition is a scientific kind, hypothesized by the relevant sciences to explain (what we take to be) a particular domain of phenomena concerning, in the first instance, the behavior of individual humans: conversation in real time, patterns of similarity in the treatment of objects (reidentification), the production of works of art, the formulation and testing of scientific theories, the playing of chess, performance on reading comprehension exams, etc.[14]

What, then, is cognition? In other work (Rupert 2004, 2009, 2010), I have argued that virtually all successful forms of cognitive modeling – computational, brute biological, robotics-based, connectionist, and dynamicist – distinguish between, on the one hand, the relatively integrated, relatively persisting architecture, and, on the other, more transient causal contributors that, together with aspects of the architecture, produce intelligent behavior (*cf.* Wilson's [2002] distinction between obligate and facultative systems), and, moreover, that this provides our best clue to the distinction between genuinely cognitive causes to the production of intelligent behavior and other contributing causes. Think of this as an inference to the best (available) explanation, twice over. First, that some distinctive and central aspect of cognitive processing is typically instantiated within the boundary of individual organisms explains why various modeling approaches that focus on the individual have been as successful as they have

[14] This approach might seem to beg the question in the current context, but it represents standard methodology in the sciences. If the process of investigation works properly, then, in the normal course of scientific research, which of the various phenomena turn out to be of a piece – that is, which are in fact produced by processes of the same kind – will emerge, regardless of the initial inclinations of researchers to group phenomena in one way rather than in another.

been. Second, that the persisting architecture is the distinctively cognitive thing in question best explains why it runs through the various forms of successful individual-level modeling.

Can we say anything more precise about the integrated nature of the system appearing inside the skin, anything that sheds light on its role as a cognitive system, that is, a system that flexibly produces a wide range of forms of intelligent behavior? In earlier work on the topic (Rupert 2009, 2010, 2011, 2013), I proposed that a cognitive system consists of a collection of mechanisms that co-contribute in overlapping subsets to the production of a wide range of forms of intelligent behavior as well as a mathematical measure meant to cash out the requirement "in overlapping subsets" (Rupert 2009, 2010). This mathematical measure is, in essence, a theory of integration. Though motivated by an attempt to characterize the internal cognitive system, the measure is location-neutral; it distinguishes between two kinds of causal contributor, wherever they appear, to the production of intelligent behavior.

Here, then, is what I now call the 'conditional probability of co-contribution' account of cognitive integration, or CPC, now refined so as to clarify its structure. Bear in mind that, although the description to follow has a procedural flavor – as if it were a recipe for carrying out a construction – it is meant to characterize the property of cognitive integration itself:

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 such 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 every member of the complement of that set's contributing causally.

3. Rank order all such conditional probabilities.

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

5. For each mechanism appearing on the list of sets with higher conditional probabilities (that is, the sets above the gap referred to at Step 4.), count the number of times that mechanism appears and rank order individual mechanisms accordingly (that is, according to their number of appearances above the gap on the list produced by Step 4.).

6. A statistically significant gap separates those mechanisms that appear frequently on this second list from those that do not.

7. The integrated cognitive system comprises all and only those mechanisms appearing above that gap on the second list.

Presented in this formal way, CPC's implications may remain obscure. Consider an example, then. The typical subject is quite good at avoiding obstacles as she moves about, and an visual edge-detection mechanism has almost certainly causally contributed to such behavior. A mechanism that computes distance from retinal disparity will likewise have contributed to obstacle avoidance in the typical subject, as will have a mechanism that calculates shape from detected shading (Marr 1982). With regard to the avoidance of obstacles, many further mechanisms have contributed, for instance, various motor control mechanisms. To keep matters relatively simple, let us add only one such motor-control mechanism to the mixture of mechanisms under consideration. The resulting set of four mechanisms allows the possibility of six two-membered sets, four three-membered sets, and one four-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; this yields a total of twelve entries on the rank-ordered list constructed at CPC's Step 3. For each of the four three-membered sets, 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; this yields a total of twenty-four additional entries on the rank-ordered list constructed at CPC's Step 3. For the four-membered set, there are fourteen relevant conditional probabilities (which thus represent fourteen further entries to the rank-ordered list in question). For any one of the four, we must include the probability of its contributing conditional on the contribution of the remaining three, and vice versa, which yields eight entries. The remaining proper subsets of the four-membered set are pairs, as are the complements in all such cases. For any such pair, and there is a conditional probability of its contributing given that its complement is contributing. That yields six entries, which together with the eight from our lopsided divisions of the four-membered set, equals a total of fourteen entries contributed by the four-membered set. Relative to only this one kind of behavior and only these four elements, we already have fifty entries on the rank-ordered list associated with CPC's Step 3. Now go through this procedure – in principle! – for every grouping of all causally contributing mechanisms relative to each form of intelligent behavior that has been exhibited by the subject in question (so long as the subject has exhibited a reasonably wide range of forms of intelligent behavior – if not, all bets are off, for this richness of repertoire is one of the central features of the *explananda* of cognitive science).

With regard to the example at hand, each of the four mechanisms will presumably appear in many subsets with high conditional probabilities (in the sense that the probability of a proper subset of a set's contributing will be high given that the complement of the set is contributing). This is a feature of the mechanisms and the form of behavior chosen. For instance, one might reasonably think that the probability of the edge-detection mechanism's contributing given that the shape-from-shading mechanism is contributing is close to one; it would seem that every time the shape-from-shading mechanism contributes to the avoidance of obstacles, the edge-detection mechanism also contributes, at least for the typical subject, partly because, as we might say informally, they are both fundamental mechanisms of visual processing. Similarly for P(edge detection|shape-from-shading & distance from retinal disparity) and for P(distance from retinal disparity & edge detection|shape-from-shading). Notice, however, that sets including only the three visual mechanisms may well deliver higher conditional probabilities than sets that mix the motor-control mechanism with the visual mechanisms, particularly where the motor-control mechanism is being conditioned upon. It seems highly probable that if the visual mechanisms are guiding obstacle avoidance, then the motor-control mechanism is. But, perhaps the motor control mechanism also contributes to obstacle avoidance in cases in which, for example, one successfully navigates a familiar room in the dark, with little visual guidance. Thus, P(shape-from-shading|motor control) may be significantly lower than the conditional probabilities just considered. This will likely not be the case when the motor-control mechanism is being conditioned upon alongside a visual mechanism. For example, P(shape-from-shading|motor control & edge detection) is not likely to be any lower than conditional probabilities involving only our three visual mechanisms; for, if the motor-control mechanism in question is contributing along with the edge detection mechanism to obstacle avoidance, then we're almost certainly talking about visually guided obstacle avoidance, in which case shape-from-shading is almost certain to be contributing as well. As a result, consideration of our four mechanisms in connection with obstacle avoidance would presumably yield many subsets with high conditional probabilities (those that appear above the cut-off point at CPC's Step 4), even if the motor-control mechanism shows up in fewer than do the other three.

CPC was initially formulated in an attempt to adjudicate claims about extended cognition, in particular, the claim that contemporary cognitive science has revealed human cognition to be extended in a deep and theoretically important way (Clark and Chalmers 1998). If cognition must occur within the cognitive system, as delineated by CPC, then it would seem that for most individual human subjects at most times, cognitive processing occurs within the boundaries of the subject's body; for, generally speaking, the preceding characterization of the cognitive system cuts against the inclusion of special-purpose tools and one-offs, which tends to be the status of causal contributors beyond the boundary of the body. (A special purpose tool will likely appear in many sets with high conditional probabilities relative to a single form of intelligent behavior, but will not appear in such sets relative to other forms of intelligent behavior, putting that special-purpose mechanism at a significant disadvantage at Step 5 relative to mechanisms that contribute to a variety of forms of intelligent behavior.) The location of individual human cognition is largely an empirical matter, though. The systems-based proposal CPC leaves open the possibility that a tool – perhaps an iPhone (Chalmers 2008) – that consistently contributes to the production of a variety of forms of intelligent behavior across a variety of contexts, alongside a shifting set of co-collaborators that themselves have similar standing, is part of a human's cognitive system.

But why think CPC is correct? Flexibility is the heart of cognition and intelligence – flexibility in learning, in the acquisition of concepts and skills, in problem-solving, and in the deployment of a variety of resources in the pursuit of and revision of goals in an oft-changing environment. It is this flexibility that attracts attention to certain forms human behavior and performance, and motivates the development of a distinctive science (cognitive science) to study them, in contrast to tropes and other stereotyped forms of behavior. It is the lack of such flexibility that drives continuing complaints about extant forms of artificial intelligence. "It's not intelligence at all," one is tempted to say about such systems, "It wouldn't have any idea what to do if an unexpected situation were to arise! It does only that one thing!" – whether that one thing is playing chess, answering quiz-show questions, or controlling an automobile.

CPC is grounded in the idea that flexibility is achieved in humans only by the presence of many units poised to work together in various combinations. There's plentiful evidence that this sort of thing happens in the human brain (Anderson 2010, 2014; Cole et al. 2013; Botvinick and Cohen 2014). On some accounts of this sort of process, subnetworks with overlapping members wrest control from each other via competitive processing. When two functional subnetworks have overlapping members, it may take only a bit of differential stimulus to shift the agent's activity from the performance of one task to the performance of a different one. On this approach, a shift in task doesn't require an entirely new network to take control from a previously dominant one; more subtle shifts in the co-activation of elements, some of which are already active, can more smoothly effect such a transition. The systems-based view CPC emphasizes what seems likely to be a central trait of such a system – that any given mechanism is capable of cooperating with various other subsets of mechanisms to complete a variety of tasks.

V. Group Cognitive States?

We have in hand, then, an abstract characterization of the cognitive system that delivers a minimal subject of cognitive states. Given the thin-ness of that subject, proponents of group cognition might take heart. It appears that the individual human's cognitive system – regardless of whether it is wholly internal – is little more than a jointly packaged collection of "demons," all doing their individual bits to produce flexible behavior, absent a Cartesian theatre and with little in the way of a central controller (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 behavior! Thus, some, perhaps many, groups would appear to meet the systems-based necessary condition for the possession of cognitive states.

Given the extensive differences in detail across the group and individual case, however, we should want to know what exactly to look for, if we are to find cognition at the group level. The material from Sections III and IV suggests the following picture. Imagine that social scientists have completed careful studies of organizational structure, group dynamics, etc. In fact, sufficient resources have been devoted to such study as to provide for experimental interventions. And, imagine that our best models of the data collected draw a consistent distinction, at the level of the group architecture or organization, between relatively persisting resources and the passing parade of other resources that causally contribute to the production of the phenomena of interest. But, imagine that when one models the co-contribution relations between the components of the persisting package, the relation that emerges is not best captured by conditional probabilities of co-contributions in the manner of the construction of Section IV. The probabilistic measure the fitting of which best captures the relations between what would appear to be the obligate parts of the group system is instead grounded in the relation *changing the likelihood of co-contribution of two others* (*cf.* Klein 2010): take any three causally contributing mechanisms and ask to what extent the contribution of any one of the three alters the probability of the contribution of the other two; then cluster the resulting values in a fashion analogous to the manner in which conditional probabilities of co-contribution are clustered above.

Such a model would bear a tweak-and-extend relation to Section IV's model of a cognitive system. For example, "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..." becomes "1. For a subject at a time, form each triplet of the mechanisms that have distinctively causally contributed to the production of any form of intelligent behavior...", which clearly constitutes a tweak. Common to the models is some probabilistic relation of co-contribution that binds some of the causally contributing mechanisms together, into a single system, in contrast to the status of other contributing mechanisms. This provides a concrete sense of what it would be for a group to engage in a different kind of cognition from an individual human, but cognition nevertheless.

VI. Complications

Some readers may wonder whether I have made matters too easy for the proponent of group cognition. In virtually any case in which a persisting system contributes to the production of some phenomenon of interest (alongside various other causal contributors), it's almost certain that *some* probabilistic measure or other partitions the causes into two groups, where the members of one group are components of the persisting system in question and the members of other group are the remaining causes. And, it would seem that an expression of just about any such probabilistic measure could be constructed by tweaking CPC. But, what makes a model substantive and interesting, one might think, are the interrelations between elements specified by the model, not merely by such a thin fact as that the causally contributing elements of the model resolve into two groups.[15] A connectionist model must include processing units that have activation values and that can pass activation to other units to which they are directly connected. This places a significant constraint on the "same-generic-kind-as" relation in the case of *being a connectionist network*. But, a model that merely applies probabilistic measures to cluster certain of the contributing causes together makes very few demands on a system – too few, one might think to be taken seriously as an account of *cognition*, as a causal-explanatory kind.[16] The remainder of this section explores this potential problem and two others, concluding with a pessimistic evaluation of the prospects for a realism about group cognitive states.

1. *The problem of missing conditions.* Something is clearly missing from section IV's characterization of the cognitive system, at least if it is intended to capture the kind *being cognitive*. Think of the problem in this way: my characterization requires, of the mechanisms in question, that they causally contribute to the production of intelligent behavior; nevertheless, the systemic integrity in question seems to have nothing

[15] Compare the search for a generic conception of cognition to the search for a generic grounding relation. In particular, consider Jonathan Schaffer's response to Jessica Wilson's criticisms of the claim that there is an interesting generic conception of grounding (Wilson 2016). Schaffer contends that what is common to various specific dependence relations (which, on Wilson's view, are the objects of genuine interest) is that they can all be represented using structural equation models (Schaffer 2016), and that this yields a useful general understanding of the grounding relation. One might reasonably wonder, though, whether the thin property *subject to being represented by a structural equation model* does significant explanatory work, absent the filling in of the dynamics relative to a particular relation, domain, or class of systems. It is unclear, for instance, what interesting range of counterfactual inferences is licensed just by a relation's instantiating the property *subject to being represented by a structural equation model*.

[16] Consider a corresponding point about Ramsey sentences: one might worry that a Ramsey sentence can be used to represent a theory of any domain that contains interrelated properties, without doing violence to the theory. It would thus be ill-advised for a functionalist in philosophy of mind to assert that any system that can be modeled using a Ramsey sentence has a mind! (And the flexibility of the Ramsey-sentence formalism should come as no surprise. Consider Lewis's use of it as a general tool for characterizing theoretical terms – Lewis 1970.)

specifically to do with cognition. The measure of probabilistic clustering of contributing mechanisms could be used to capture, for example, biological or chemical systems. It is partly on account of this worry that I'm inclined to treat CPC as an articulation of only a necessary condition for a state's being cognitive.[17]

What must be added? One obvious suggestion is that the system trade in representations (Rupert 2005), perhaps representations that are governed by a specific kind of algorithm or process (Rupert 2018*b*). Or, one might plausibly require that for something to be a *cognitive* system, it must have cognitive states of specific types – that it must, for example, have some genuine memories, beliefs, or perceptual states. But, whatever else is demanded of a family of models that captures cognition, we should worry that the inclusion of such further conditions raises the bar to a height that typical groups do not meet.

Here's another way to come at this concern. To a great extent, an ecumenical attitude toward the variety of forms of successful cognitive-scientific modeling motivated the formulation of Section IV's systems-based criterion. Part of the idea was to show, without begging the question against, say, anti-representationalists, that a plausible necessary condition on cognition – one that everyone should agree to – cuts ice in the debate over extended cognition. At other times the motive was pluralist: perhaps all of the even moderately successful forms of modeling in cognitive science have got hold of something important, and thus what is common to the programs illuminates the nature of

[17] Note that treatment of it as a necessary condition serves certain dialectical purposes sufficiently well. If, for example, *(a)* anywhere mental states appear, there must also be cognition, *(b)* a state's obtaining within a cognitive system of the sort characterized in Section IV is a necessary condition on a state's being cognitive, and *(c)* extant groups do not have cognitive systems, an interesting conclusion follows: extant groups do not have mental states.

cognition. The value of such open-minded-ness duly acknowledged, it is entirely possible that a certain family of models of human cognition will, as cognitive science matures, win out decisively; in which case, ecumenical and pluralistic motives will appear excessively circumspect. Instead, cognitive science will have identified cognition, in the human case, with some fairly specific kind of architecture and processes, leaving little reason to think the study of our paradigmatic cases of cognition will yield a thin, overarching property *being a cognitive system* that groups might instantiate – as opposed to their being merely persisting systems, which many groups clearly are.

2. *The problem of flexible, intelligent behavior*. Many groups simply do not produce a wide range of forms of intelligent behavior. Those that make frequent appearances in the literature on group cognition 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 does not do a whole lot else, as a coherent group.

In contrast, properly cognitive systems produce a wide range of forms of intelligent behavior; the characterization of any one form of a subject's behavior as intelligent depends on the subject's producing a wide range of other forms of intelligent behavior. The single subject's ability to produce a wide range of forms of behavior in a flexible manner constitutes one of cognitive science's central *explananda*. As noted above, a longstanding complaint about artificial intelligence turns on just this point: the achievements of Deep Blue, MYCIN, self-driving cars, Alpha Go, and many other specialized and expert systems are too narrowly targeted to manifest genuine intelligence.

Intelligence requires the ability to solve all sorts of problems flexibly – to prioritize tasks, to switch between tasks in appropriate ways at appropriate times, to switch between strategies for the accomplishing a given task as the situation demands, and so on (to tell a joke or sing a song when it's called for, to study for the Medical College Admission Test, when that's called for, etc.).

Typical groups do not seem to have the right structure to exhibit a wide range of forms of intelligent behavior. And, to the extent that groups exhibit some flexibility, a certain asymmetry in the production of such behavior seems telling. It is virtually always the case that where a group exhibits flexibility of degree *y*, each contributing individual in the group who helps to bring about the behavior in question is capable of cognitive flexibility of degree *x*, where *x* is significantly greater than *y*. This would seem to be the opposite of the situation in the individual human's case. Individual humans may well have specialized subsystems responsible for performing much of the work in their purview. Nevertheless, the individual subject as a whole has a much more flexible capacity than any individual module or specialized subsystem. In contrast, group actions are virtually always carried out, or enacted, by individual members (or representatives) of the group, which members are individually significantly more flexible than the group. Perhaps this fact alone does not speak against realism about group states. But, this asymmetry – the fact that the individuals are more flexible than the group – should give us pause; in such a case, it would appear that genuinely intelligent beings are cooperating to organize a group that is, as in the case of A.I, only a relatively inflexible replica of a specialized skill.[18]

[18] Thanks to Gabriel Rabin for pushing me to clarify my view about these matters.

3. *The problem of constraints on model-selection*. We derive substantive conclusions from tweak-and-extend relations only when the models in question are our best models of the relevant phenomena. Thus, canons of model-selection – and methodological considerations, more generally speaking – bear on the debate about group cognition. These include such methodological desiderata as simplicity, conservatism, and unity of science, as well as the guiding principles of statistical modeling.

The contribution of methodological principles threatens to cause problems for realists. If, for example, a constraint on a state's being cognitive is that it be representational, then concerns about, for example, quantitative parsimony (Nolan 1997) rear their head. Consider a case in which all of the members of an appellate court vote to uphold a criminal conviction, although not one of them believes the party in question is guilty. Does that show that the court is in a representational state – perhaps a belief that the accused is guilty – that differs from, and exists in addition to, the cognitive states of the individual court members? Considerations of quantitative parsimony suggest not. Positing an additional representational state – the court's cognitive state – is gratuitous and thus would not find its way into the model of the phenomenon in question (that is, the court's issuing of a decision). For, presumably, each individual member of the court has his or her reasons for voting to uphold the conviction even though she or he does not believe the accused is guilty. Such reasons might range from an individual justice's principled commitment to precedent to a belief that there will be a riot if the accused is set free to a belief in a premise-based approach to judicial decisions. Once, however, such reasons have been included in one's model of the court's issuing of the decision, all of the representational resources needed to explain the event are at hand (Rupert 2005, 2011, 2014; Wilson 2004). Adding more of the same, that is, more representational resources, is gratuitous. *Ceteris paribus*, then, canons of model selection dictate the choice of a model in which representational states do not appear as states of the group as a whole. Therefore, our best overall model of group-involving processes that produce intelligent behavior (if such behavior exists) will not attribute representational states to groups, *contra* List's Premise 1.

This moral might be expressed not so much in terms of parsimony but as a point about reduction and the unity of science. It is a canon of model selection that, other things being equal, we should prefer models of phenomena that dovetail with successful models in adjacent domains. Or, to water this down a bit, it is a canon of model selection that *ceteris paribus*, when attempting to draw ontological conclusions from successful modeling, favor collections of models that present a unified picture of nature over those that do not. Consider, on the one hand, a model according to which social and political institutions have cognitive states, because they exhibit rationality over time, and, on the other hand, a model according to which only individual human cognitive processes explain measurable phenomena related to social and political institutions (ignoring all of the other components of both kinds of model, such as sheets of paper, gasoline pumps, etc.). The latter view is more consistent with an emphasis on the unity of science. Or, one might combine these two concerns in the form of a dilemma: either the proponent of group cognitive states opts for freestanding and mysterious group processes, in which case she offends against the unity of science, or she makes a genuine effort to integrate models in a single package and thereby offends against quantitative parsimony, if she insists that groups have cognitive states of their own.[19]

One should be careful not to be misled by talk of reduction, however. Return to List's realist argument presented in Section II, which vaguely suggests an anti-reductionist line of thought – the idea being that if the social sciences make productive use of attributions of group beliefs and desires, a card-carrying naturalist should not resist (unless the states in question can be reduced). This vague impression is reinforced by various remarks in List and Pettit (2011) about, for example, interlevel relations (*ibid.*, 4), multi-level causal competition (*ibid.*, 161), the realization of group attitudes by individual attitudes (*ibid.*, 77), supervenience (as a relation between levels – *ibid.*, 65), and the difficulties of reducing group attitudes to individual attitudes (*ibid.*, 76–77, 194). But, this framing of the issue strikes me as confused or at least misleading. Generally speaking, the question of reduction is an *inter*level question (as are questions about supervenience and realization), arising in the case of different scientific domains with different sets of properties, and this is often put in terms of relations between different levels. Yet, that is decidedly not the sort of case at hand. The group and individual cognitive states are supposed to be of the same kind, parts of the same domain: the mental, psychological, or cognitive. This is what's supposed to be striking about the realism in question. After all, it would be no surprise to find out merely that some states or other at the group level play

[19] It might be objected that the invoking of canons of theory selection is heavy handed, for the status of such canons is contested, and their historical contribution to scientific reasoning complex and opaque. Rather, it might be thought, the discussion is best limited to consideration of first-order reasons for or against proffered explanations of specific phenomena, data, or effects. Point taken. But, so far as I can tell, the concerns expressed here can be translated into first-order arguments on a case-by-case basis: we know why the justices voted the way they did, and thus we know why the decision is what it is, without any reference to a group cognitive state. And this way of viewing the matter dispenses with any need to appeal to an overarching methodological principle. Thanks to Peter Achinstein for pressing me on this point.

a causal-explanatory role (*cf.* "the weight of the group caused the terrace to collapse"). List and Pettit seem to lose sight of this when making what they take to be a realism-vindicating comparison between (a) the relation between group cognitive states and individual cognitive states and (b) the relation between individual cognitive states and "the configuration and functioning of biological subsystems" (*ibid.*, 7); they argue that, if it is legitimate to resist reduction in case (b), then case (a) should be treated likewise (*ibid.*, 78; *cf.* Goldstone and Gureckis 2009, Huebner 2008). This argument by analogy fails, however. Case (b) concerns two genuinely different sets of properties (cognitive properties, on the one hand, and biological properties, on the other), aptly raising questions of reduction and anti-reduction. But, case (a) concerns the very same kinds of properties and states, on both sides of its comparison, not one set of properties or states of one kind realizing or supervening on properties or states of another.[20] Thus, one cannot simply take on board anti-reductionist resources used in case (b) to defuse parsimony-based arguments aimed at case (a).[21]

Here's another way to see this: in case (b), reduction, in the intertheoretic sense, is blocked by anti-reductionist arguments (Fodor 1974, for instance), and we thereby ontologically certify a domain of properties – psychological, mental, or cognitive ones.

[20] List and Pettit sometimes seem to acknowledge this fact (2011, 1, 78) but do not seem to appreciate the extent to which it undermines their defense of group agents.

[21] In response, one might advert to the widespread acceptance of both personal and sub-personal levels (rather than the relation between the psychological and the neural levels) in cognitive theorizing (Huebner 2008). If, disregarding the pull of parsimony, we refuse to eliminate the personal level in favor of the subpersonal one, we should not allow the individual level to crowd out the group level, the thinking goes. To my mind, however, the personal-subpersonal distinction has no substantive role to play in cognitive science (see Rupert 2018*a* for arguments to this effect); in which case, the realist can take no consolation here. Moreover, most philosophers who accept the existence of a personal level take the states appearing there to be of a different kind from those that appear at the subpersonal level. Thus, the supposed distinction between the personal and subpersonal levels does not provide an example of a case in which the same psychological, mental, or cognitive properties appear at different levels.

Case (a), however, involves the situation in which we have already let cognitive or mental kind into our ontology, and the only question is how far to spread them around. Case (b) is an interlevel matter; case (a) an *intra*level matter. That is why List and Pettit's argument by analogy is inapt, and in case (a), the case at hand, we should focus on quantitative parsimony (Nolan 1997), the question of how many instances of a given, already countenanced kind of state or property we should include in our scientific models.[22]

Consider a final response to concerns about parsimony and reduction. List and Pettit claim, "The agency of the group relates in such a complex way to the agency of individuals that we have little chance of tracking the dispositions of the group agent and of interacting with it as an agent to contest or interrogate, persuade or coerce, if we conceptualize its doings at the individual level" (2011, 76). I find this remark puzzling. Throughout their book, List in Pettit develop precise analyses of how best to construct groups so that the groups will exhibit rationality (as best as can be done given various impossibility results). Moreover, these recipes tell us exactly who to talk to in an organization, because they tell us, relatively transparently, who will be voting on (or otherwise having a determinative say about) what. For example, List and Pettit

[22] In a discussion of some of my previous work on group cognition (Rupert 2014), Himmelreich (2015) complains that in making a quantitative-parsimony-based argument against realism, I do not avail myself of various conceptual tools emerging from debates about causal exclusion. But, given comments in the main text, it might seem clear why the proposed path is less promising than Himmelreich takes it to be. The conceptual tools developed in debates about causal exclusion are generally aimed at interlevel relations; in Kim's standard diagram (e.g., in Kim 1998), the vertical relations at issue are supervenience relations, which are normally taken to hold between distinct families of properties. But, the relation between group states and individual states are not meant by realists to concern relations between distinct sets of properties. After all, if the sets of properties are distinct, why do realists use the same terms – e.g., 'belief' and 'desire' – to talk about group states as are used to talk about individual states? Note, too, that when List and Pettit pursue the tack proposed by Himmelreich, they approach the problem specifically as a problem about multi-level causation (2011, 161), which doesn't suit the case of cognitive states competing with cognitive states; that wouldn't be a matter of multi-level causation, for the reasons given above.

demonstrate formally that, under certain conditions, an effective procedure for maintaining group rationality is to divide logically interrelated questions on an agenda in the following way: identify premises, assign a subgroup of members with relevant specialized knowledge to each of the premises, take a majority-rules vote among each of the subgroups on that group's assigned premise, and then reach a conclusion based only on the outcomes of those votes (that is, follow a premise-based procedure) (List and Pettit 2011, 95). In that case, it seems apparent to whom one should speak if one would like to persuade the group to reach one's favored conclusion or to coerce a certain outcome: the members of the specialized subgroup that one thinks one has the best chance of winning over. (There exists an entire industry in the United States – the lobbying industry – built on this kind of observation.) What List and Pettit take to be group attitudes supervene on individual attitudes and, as they emphasize, sometimes in a not entirely straightforward way. Fair enough. But, the lion's share of their book demonstrates exactly why and for what principled reasons the pattern of supervenience should be less-than-entirely-straightforward. And, by explaining the less-than-entirely-straightforward relations to individual decisions that fix supervening facts about the group, List and Pettit themselves demonstrate that a less-than-entirely-straightforward supervenience relation does not entail the sort of opacity that might motivate the positing of autonomous group states. They provide recipes for the construction of group states by individuals and thereby provide the guidebook for conceptualizing those states as amalgamations of individual decisions (*cf.* Rupert 2005, 179; Roth 2014)[23] – not as genuine group states – which tells

[23] Rupert (2005) and Roth (2014) emphasize the role of higher-order individual states in reduction- or otherwise parsimony-based criticisms of realism, such higher-order states as an individual's endorsing a rule of aggregation or supporting a canonical procedure for constructing a group's officially stated conclusion. An example of such a canonical procedure would be the following of precedent in the legal

us which individuals should be influenced on which questions at what stage in a given process in order to bring about the results we would like.

Conclusion. In Sections II-V, I motivated an interest in the theory of natural kinds as a tool for evaluating realism about group cognition; offered a theory of the sort of evidence germane to the determination of sameness of natural kind; and applied the theory to the case of the overarching kind *cognition*, in the hope of offering the realist a way forward. In the final section, prohibitive difficulties mounted. I conclude, then, that the prospects for the discovery of genuinely cognitive states in groups appear dim, though much depends on empirical work yet to be done.

In particular, although this closing section has been generally pessimistic in its evaluation of the prospects for realism, it is worth mentioning an empirical avenue worth pursuing. It might be that, although CPC does not seem to capture a robust or interesting property – beyond something's being a coherent system – the kind of structure characterized is, in fact, just the kind of structure needed to support flexible behavior (timely switching between tasks, etc.). Perhaps, nomologically speaking, there is no other way to construct a system that exhibits flexibility of the right sort except to endow it with a broad collection of mechanisms that contribute in overlapping subsets to the production of a wide range of forms of behavior; perhaps it is this structure in particular that facilitates flexibility. In which case, such structure would have more to do with the nature

---

arena. It's one thing to believe that the legally correct decision, based only on law as it's written, is P. It's another thing to believe that the correct judgement, based only on law as it's written, is not-P, but that the best thing for the court to do on balance, given P-related precedents, is to decide that P. A justice with a higher-order commitment to the role of precedent may well find herself in the latter situation. In defending realism against reductive or otherwise parsimony-based criticisms, List and Pettit, as well as Gilbert, argue as if the critic has only first-order states of the individual (such as the belief that the defendant is guilty) to work with. That, however, is to argue against a straw individualism.

of cognition than sometimes suggested above and the search for such structure in groups may bear cognition-related fruit after all.

References

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

Anderson, Michael L. (2014). *After Phrenology: Neural Reuse and the Interactive Brain*. Cambridge, MA: MIT Press.

Arico, A. (2010). Folk psychology, consciousness, and context effects. *Review of Philosophy and Psychology* 1: 371–393.

Barnier, Amanda J., John Sutton, Celia B. Harris, and Robert A. Wilson. (2008). A conceptual and empirical framework for the social distribution of cognition: The case of memory. *Cognitive Systems Research* 9 (1): 33–51.

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. (1997). Anti-reductionism slaps back. In J. Tomberlin (Ed.), *Mind, Causation, World: Philosophical Perspectives* 11, pp. 107–133. Boston: Blackwell.

Botvinick, Matthew M., and Jonathan D. Cohen. (2014). The computational and neural basis of cognitive control: Charted territory and new frontiers. *Cognitive Science* 38: 1249–1285.

Brooks, Rodney A. (1999). *Cambrian intelligence: The early history of the new AI.* Cambridge, MA: MIT Press.

Chalmers, D. (2008). Foreword to Andy Clark, *Supersizing the mind* (see Clark 2008).

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. (2011). Finding the mind. *Philosophical Studies* 152: 447–461.

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

Cole, Michael W., Jeremy R. Reynolds, Jonathan D. Power, Greg Repovs, Alan Anticevic, and Todd S. Braver. (2013). Multi-task connectivity reveals flexible hubs for adaptive task control. *Nature Neuroscience* 16(9): 1348–1355.

De Brigard, F. (2017). Cognitive systems and the changing brain. *Philosophical Explorations* 20 (2): 224–241.

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

Evans, J. St. B. T. and K. Frankish (eds.). (2009). *In two minds: Dual processes and beyond*. Oxford: Oxford University Press.

Fodor, J. A. (1974). Special sciences (Or: The disunity of science as a working hypothesis). *Synthese* 28 (2): 97–115.

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

Fodor, J. (1987). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind.* Cambridge, MA: MIT Press.

Gendler, T. (2008). Alief and belief. *Journal of Philosophy* 105 (10): 634–663.

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

Gilbert, M. (2004). Collective epistemology. *Episteme: A Journal of Social Epistemology* 1: 95–107.

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

Goldstone, R., and G. Theiner. (2017). The multiple, interacting levels of cognitive systems (MILCS) perspective on group cognition. *Philosophical Psychology* 30 (3): 338–372.

Griffiths, T. L., N. Chater, C. Kemp, A. Perfors, and J. B. Tenenbaum. (2010). Probabilistic models of cognition: Exploring representations and inductive biases. *Trends in Cognitive Sciences* 14: 357–364.

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

Himmelreich, J. (2015). Review of *From Individual to Collective Intentionality: New*

Essays, in *Economics and Philosophy* 31 (3): 479–486.

Huebner, B. (2008). Do you see what we see? An investigation of an argument against collective representation. *Philosophical Psychology* 21 (1): 91–112.

Huebner, B. (2013). *Macrocognition: A theory of distributed minds and collective intentionality*. Oxford: Oxford University Press.

Huebner, B., M. Bruno and H. Sarkissian. (2010). What does the nation of China think about phenomenal states? *Review of Philosophy and Psychology* 1: 225–243.

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

Kim, J. (1998). *Mind in a physical world: An essay on the mind-body problem and mental causation*. Cambridge: 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.

Knobe, J. and J. Prinz. (2008). Intuitions about consciousness: Experimental studies. *Phenomenology and the Cognitive Sciences* 7: 67–83.

Lewis, D. (1970). How to define theoretical terms. *Journal of Philosophy* 67: 427–46.

Lewis, D. (1980). Mad pain and Martian pain. In N. Block (ed.), *Readings in the Philosophy of Psychology, Volume one*, pp. 216–222. Cambridge: Harvard University Press.

List, C. (2018). What is it like to be a group agent? *Noûs* 52 (2): 295–319.

List, C., and P. Pettit. (2011). *Group agency: The possibility, design, and status of corporate agents.* Oxford: Oxford University Press.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. Cambridge: MIT Press.

Nolan, D. (1997). Quantitative parsimony. *British Journal for the Philosophy of Science* 48 (3): 329-343

Putnam, H. (1967) Psychological predicates. In W. H. Capitan and D. D. Merrill (eds.), *Art, Mind, and Religion*, pp. 37–48. Pittsburgh: University of Pittsburgh Press.

Putnam, H. (1975). The meaning of ‘meaning’. In H. Putnam (ed.), *Mind, Language, and Reality: Philosophical Papers*, vol. 2, pp. 215–271. Cambridge: Cambridge University Press.

Roth, A.S. (2014). Indispensability, the discursive dilemma, and groups with minds of their own. In S. R. Chant, F. Hindriks, and G. Preyer (eds.) *From Individual to Collective Intentionality: New Essays*, pp. 137–162. Oxford: Oxford University Press.

Rovane, C. (2014). Group agency and individualism. *Erkenntnis* 79: 1663–1684.

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. D. (2004). Challenges to the hypothesis of extended cognition. *Journal of Philosophy* 101, 389–428.

Rupert, R. D. (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. D. (2009). *Cognitive systems and the extended mind*. Oxford: Oxford University Press.

Rupert, R. D. (2010). Extended cognition and the priority of cognitive systems. *Cognitive Systems Research* 11: 343–56.

Rupert, R. D. (2011). Empirical arguments for group minds: A critical appraisal. *Philosophy Compass* 6 (9): 630–639

Rupert, R. D. (2013). Memory, natural kinds, and cognitive extension; or, Martians don't remember, and cognitive science is not about cognition. *Review of Philosophy and Psychology* 4 (1): 25–47

Rupert, R. D. (2014). Against group cognitive states. In S. Chant, F. Hindriks, and G. Preyer (eds.), *From individual to collective* intentionality, pp. 97–111 Oxford: Oxford University Press.

Rupert, R. D. (2018*a*). The self in the age of cognitive science. *Philosophic Exchange* 47. https://digitalcommons.brockport.edu/phil_ex/vol47/iss1/2 .

Rupert, R. D. (2018*b*). Representation and mental representation. *Philosophical Explorations*, 21, 2: 204–225.

Schaffer, J. (2016). Ground rules: Lessons from Wilson. In K. Aizawa and C. Gillett (eds.), *Scientific composition and metaphysical ground*, pp. 143–169. Springer.

Schwitzgebel, E. (2015). If materialism is true, the United States is probably conscious. *Philosophical Studies* 172: 1697–1721.

Shoemaker, S. (1981) Some varieties of functionalism. *Philosophical Topics* 12: 93–118.

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

Stich, S. (1996). *Deconstructing the Mind*. Oxford: Oxford University Press.

Sytsma, J. and E. Machery. (2009). How to study folk intuitions about phenomenal consciousness. *Philosophical Psychology* 22: 21–35.

Sytsma, J. and E. Machery. (2010) Two conceptions of subjective experience. *Philosophical Studies* 151: 299–327.

Theiner, G. and T. O'Connor. (2010). The emergence of group cognition. In A. Corradini and T. O'Connor (Eds.). *Emergence in science and philosophy*, pp. 78–117. New York: Routledge.

Theiner, G., C. Allen, and R. L. Goldstone. (2010). Recognizing group cognition. *Cognitive Systems Research* 11: 378–395.

Tollefsen, D. (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, J. (2016). The unity and priority arguments for grounding. In Kenneth Aizawa and Carl Gillett (Eds.), *Scientific Composition and Metaphysical Ground*, pp. 171–204. London: Palgrave Macmillan.

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.