Distributing cognition:
A defense of collective mentality

Bryce Huebner

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Philosophy.

Chapel Hill
2008

Approved by:

Jesse J. Prinz

William G. Lycan

Joshua Knobe

Ram Neta

Alison Wylie

ABSTRACT

BRYCE HUEBNER: Distributing cognition
(Under the direction of Jesse J. Prinz)

While ordinary language allows for the attribution of mental states to collectivities, there is broad agreement among philosophers and cognitive scientists that such attributions should not be taken literally because they are at best *explanatorily superfluous* and at worst *wildly implausible*. I argue that the widely shared philosophical assumption that mentality is exclusively a property of individuals is mistaken. One prominent objection to the idea that collectives could be in genuinely mental states is that they lack self-consciousness and the capacity for qualitative consciousness. I argue that neither self-consciousness nor qualitative consciousness is necessary for mentality. But I also show that both collective self-consciousness and qualitative consciousness *are possible*. Another objection states that collectives cannot possess representations above and beyond the representations in the minds of the individuals that compose them. I counter that representations in individual minds often depend on representations in lower-level subsystems and I argue that collective representations can arise in a similar way. I conclude by demonstrating that collective cognition is not a mere possibility; there are cases of collective cognition in the actual world.

For Sheri and Robert

ACKNOWLEDGMENTS

I am extremely thankful to a number of people for comments on various parts of this thesis. Chief among these are: Jacek Brzozowski, who exhibited an uncanny ability to convince me that there was a metaphysical side to these philosophical questions; Joshua Knobe, who helped to fuel my continued excitement in this project by persistently, and sincerely exhibiting interest in where it was heading; Bill Lycan, who taught to write about traditional questions in the philosophy of mind with rigor and style; and Jesse Prinz who helped to push me further on central arguments than I might have otherwise gone. I am also thankful to Mike Bruno, Ben Fraser, Eric Mandelbaum, Dylan Sabo, Hagop Sarkissian, Dave Ripley, and Susanne Sreedhar for insightful and constructive conversations in working through the toughest arguments in this thesis. Finally I would like to thank Robert Wilson and Robert Rupert for a number of conversations that provided me with important insights into the mistakes that a proponent of collective mentality is likely to make.

TABLE OF CONTENTS

Chapter:

I. I'VE GOT HALF A MIND TO RETHINK THE POSSIBILITY OF COLLECTIVE MENTALITY.....1

II. THE COLLECTIVE CONSCIOUS?.....41

III. I JUST CAN'T GET YOU OUT OF MY HEAD.....73

IV. COLLECTIVE REPRESENTATION?.....130

V. COLLECTIVE MENTALITY REVIVED!.....181

WORKS CITED 247

CHAPTER I

I'VE GOT HALF A MIND TO RETHINK THE POSSIBILITY OF COLLECTIVE MENTALITY

I have a cat named Nutmeg, and there is little that you could say to dissuade me from thinking that Nutmeg, like Fodor's (1987) Greycat (and unlike rocks, trees and spiral nebulae), has beliefs, desires, and a whole host of other mental states.[1] You might wonder what makes me so sure—and luckily I've got good reasons for *my belief*. First, when I go to sleep and Nutmeg finds her food bowl empty, she paws at my face and meows until I get out of bed and replenish her supply of cat food. Second, Nutmeg meows incessantly when I open the cabinet where I keep her kitty treats, and it's pretty clear that she both wants one of those delicious fish flavored delicacies and believes that if she meows at me she will get some of them. To put the point briefly, using "commonsense belief/desire psychology explains vastly more of the facts about [Nutmeg's] behaviour than any of the alternative theories available" (Fodor 1987, x).

The important thing to notice here is that these ascriptions of mental states to Nutmeg

[1] There are, of course, people who would deny this. Donald Davidson (1982) argues that the attribution of content to a subject requires substantial agreement between the attributor and the subject across a broad network of interrelated beliefs. This argument turns on holistic intuitions that I am unwilling to grant. I am inclined to think that there are other ways of individuating content that don't require such a substantial overlap. Stephen Stich (1979) offers another argument against animal beliefs that does not turn on such holistic considerations. His argument is based on the claim that the commonsense notion of belief requires that a belief has some specifiable content and that this content figures into the explanation of the systems behavior. I'm inclined to think that teleological theories of content (e.g., Millikan 1984) offer a promising response to these worries. However, defending this claim would take us far beyond the boundaries of this thesis.

are grounded on the same psychological considerations that I use when I ascribe beliefs and desires to my friend, Jacek, when he frowns while staring at the empty pickle jar in the fridge or complains about having no sardines left for breakfast. Of course, I do ascribe fewer, and simpler, beliefs and desires to Nutmeg than I ascribe to Jacek, commonsense psychology does a fairly good job of allowing me to predict and explain both Jacek's and Nutmeg's behavior—and this gives me very good reasons for thinking that both of them to have mental states.[2]

I have a lot to say about both Jacek's and Nutmeg's psychology (in fact, more than anyone should ever have to listen to). However, this isn't a thesis about Jacek's psychology nor is it a thesis about Nutmeg's psychology. It is a thesis about another sort of cognitive system to which we often ascribe mental states.[3] I'll refer to the sorts of systems with which I am concerned as collectivities.[4] The central contention of this thesis is that some ascriptions of mentality to collectivities ought to be understood *literally*, that is, they ought to be understood to refer to theoretical entities like beliefs, desires, and the like, in precisely the same way that ascriptions of mentality to Jacek and Nutmeg. Of course, the claim that

[2] At this point, there are two open options. You can either make the abductive inference to the conclusion that the mental states you ascribe to an organism are token-token identical to some physical state of that organism, or you can take that inference to be bogus. At this point, have little to say about this debate. My point here is merely that my evidence for the truth of the claim that Jacek has mental states is of a piece with the evidence for the truth of the claim that Nutmeg has them as well. How we understand the truth conditions for either of these organisms is, however, a further question to which I will return below.

[3] I use the term 'cognitive system' to refer to any information-processing system that possess the capacity to be in some mental state or other. Which sorts of things fall within the extension of 'cognitive system' is itself an interesting question—and one that I have many thoughts about. However, answering the question "what is a cognitive system?" falls beyond the scope of this thesis.

[4] I use the term 'collectivity' to pick out cognitive systems that are themselves constituted by other (preferably paradigmatic) cognitive systems. A few examples will be helpful for giving an idea of what I intend this term to pick out: sports teams (such as the 2005 Tarheel basketball team and the New Zealand All Blacks), corporations (such as Microsoft and Macintosh), herds and flocks, rioting crowds, the Communist party, the proletariat, anarchist collectives, avant-garde jazz ensembles, military units, and ant and bee colonies.

collectivities literally possess mental states in the same way that individuals do may seem to be a strange claim; however, it will help to note that we often ascribe mental states to collectivities in saying things like:

> Hewlett-Packard *believes* that loading Yahoo as the default search engine on its consumer PCs is the correct response to a similar *agreement* between Dell and Google.

More importantly, using commonsense psychology to ascribe mental states in cases like this is explanatorily useful for much the same reason that ascriptions of such states to Nutmeg and Jacek are explanatorily useful: in numerous cases where we want to predict and explain the behavior of a corporation such as Hewlett-Packard, using commonsense psychology proves to be a fairly reliable means for these predictions and explanations.[5]

Unfortunately, however, the instrumental values of prediction and explanation will never be sufficient to establish the literal truth of such ascriptions of mental states. As Daniel Dennett (1987a) notes, commonsense ascriptions of mental states to various purportedly cognitive systems do not form a natural kind; instead, they form a motley assortment of serious belief attributions, metaphors, facons de parler and other sorts of dubious ascriptions. So, we can't just look to the ascriptions actually allowed by commonsense psychology in order to demonstrate the existence of collective mental states. What we need is an answer to the following question: are commonsense attributions of mentality to collectivities more like attributions of such states to Nutmeg or Jacek or are they more like attributions of such states to simple thermostats (e.g., "it thinks it's colder that it really is") or to plants (e.g., "you

[5] NB: My claim is not that the sole end of commonsense psychology is prediction and explanation. It's an open and empirical question what sorts of phenomena commonsense psychology is directed towards and how commonsense ascriptions of mental states are used. Josh Knobe has recently adduced evidence suggesting that there might be other, perhaps moral, ends toward which commonsense psychology is directed. While I don't want to come down on either side of this issue at this point, I'm inclined to think that even if this is true, it would still be the case that a large number of our ascriptions of beliefs and desires will still take as their end the prediction and explanation of (both overt and covert) behavior.

should put the grass seeds in the freezer the night before you plant them; that way they'll think that winter is over and they'll start growing")? In this thesis I argue that we ought to construe at least some ascriptions of mental states to collectivities literally. However, I'll start with a quick look at the sorts of claims commonsense allowed by commonsense psychology in order to see what sorts of collectivities might qualify as cognitive systems.

1.1. The individualist dogma and commonsense psychology:

There is a commonplace dogma that holds that the mind cannot extend beyond the physical boundaries of an individual organism.[6] As Robert Wilson (2004, 3) articulates this dogma, "minds do not float free in the air or belong to larger, amorphous entities, such as groups, societies, or cultures. No, they are tightly coupled with individuals." From the standpoint of commonsense, it seems that individuals might even be identified by, or even be identical to their minds. And this seems to hold true even in the face of the intuitive dualism that pervades commonsense psychology (cf., Bloom 2004). That is to say, there is at least a pre-reflective intuition, prevalent in commonsense psychology, that there is psychological states supervene on neurological states. Most people seem to take it to be intuitively obvious that the correct way of studying the mind is by studying the brain. However, this is not just a commonsense mistake. In fact, cognitive science has also been unabashedly wedded to a focus on the individual—with some philosophers (cf., Fodor 1980 and 1991) going so far as to claim that psychology can only be practiced as a science of the individual. But what are we

[6] Difficulties can arise at this point; there can be some debate over what counts as an individual system. Cases like slime molds (*Physarum polycephalum*) and corals (phylum *Cnidaria*) push the boundaries of our intuitions about what counts as a single organism, as do colonies of termites (order *Isoptera*) ants (family *Formicdae*) and bees (superfamily *Apoidea*) that are sometimes studied as superorganisms. For the time being, I'll not get into any of these issues. For the purposes of this intuition all that I mean is a single member of a biological species (e.g., a tiger, a human, a marmoset, a cat, or a raccoon).

to say of this dogma? Like many pre-reflective intuitions in commonsense psychology, this one seems to be in tension with a number of the other intuitions that we find in commonsense psychology.

Let me begin to draw out this tension by considering some of the cognitive systems to which we ascribe mentality. The first thing to notice is that the cognitive systems with which we take ourselves to engage on a daily basis aren't limited to individuals; we also interact with collectivities. Corporations, institutions, states, nations, jazz ensembles, faculties, and other sorts of collective entities play a central role in helping us navigate our social world, as well as a central role in practical reasoning. We at least talk as though the faculty of the philosophy department is *considering* a tenure case—something that can have quite a substantial effect on an individual. We also talk as though a corporation like Shell Oil can *believe* that it needs a new environmental policy in order to respond to the criticisms of environmentalists. And I'm inclined to think that many of the avant-garde jazz ensembles that I listen to *want* to be innovative and provocative. As with the individuals to whom we ascribe mental states, many of these collective entities seem to engage in various sorts of actions and seem to do so on the basis of various sorts or intentional states. This presents us with at least a *prima facie* reason to attribute mentality to collectivities. And attribute mentality to collectivities we do.

> Consider a few quotes from various news sources:
>
> Israel *accuses* others of terrorism at the same time as it carries it out in the harshest forms"…"The Lebanese government *estimated* the damages at more than $500 million, not including loss of tourism and commerce (Mouawad and Erlanger 2006, emphasis mine).
>
> With the battle between Israel and the Lebanese militia Hezbollah raging, key Arab governments have taken the rare step of *blaming* Hezbollah, underscoring in part their growing *fear* of influence by the group's main sponsor, Iran (Fattah 2006, emphasis

mine)

> North Korea *said* Sunday that it was not bound by a United Nations Security Council resolution imposing weapons-related sanctions on it, and *insisted* it would 'bolster its war deterrent" in every way' (Reuters 2006, emphasis mine)

> Microsoft *fears* that Google could become a kind of operating system of the Internet in the same way that Windows is the dominant operating system of personal computing (Lohr and Hansell 2006).

> Whether Mr. Sokolof will be as successful this time is not so clear, but he certainly made McDonald's *angry* (Burros 1990)

In many cases, we are perfectly willing to accept it as true that the locus of a particular mental state is not the individual but a group, corporation or collective entity of some other sort.

Cases from contemporary fiction and contemporary film also provide data suggesting that intuitions about *the possibility* of collective mentality abound. Consider an example from Carson McCullers' *The ballad of the sad café*:

> Some eight or ten men had convened on the porch of Miss Amelia's store. They were silent and were indeed just waiting about. They themselves did not know what they were waiting for, but it was this: in times of tension, when some great action is impending, men gather and wait in this way. And after a time there will come a moment when all together they will act in unison, not from thought or from the will of any one man, but as though their instincts had merged together so that the decision belongs to no single one of them, but to the group as a whole. At such a time, no individual hesitates. And whether the matter will be settled peaceably, or whether the joint action will result in ransacking, violence, and crime, depends on destiny. (McCullers 1992)[7]

And another, from Robert Heinlein's *Methuselah's Children*:

> Since each of their egos was shared among many bodies, the death of one body involved no death for the ego. All memory experiences of that body remained intact, the personality associated with it was not lost, and the physical loss could be made up by letting a young native "marry" into the group. But a group ego, one of the personalities which spoke to the Earthmen, could not die, save possibly by the destruction of every body it lived in. They simply went on, apparently forever. (Heinlein 1941)

[7] Thanks to Bill Lycan for pointing me to this passage.

Thoughts about collective mentality are commonplace in contemporary science fiction—and the interesting thing to note is that they don't seem all that bizarre or far fetched. We can all make sense of the Borg of *Star Trek*, the bugs in Heinlein's *Starship troopers*, the Overmind in Clarke's *Childhoods end*, and the Precogs from Dick's "Minority Report". More importantly, we do so without questioning the ascription of mentality to various systems larger than the individuals that constitute these groups. Although this is *science fiction*, the fact that ascriptions of mentality to collectivities are so pervasive suggests that collective mentality is *at least* a possibility that we don't find too surprising—whether there are any actual collective cognitive systems is, of course, another question.

If all I had to go on was this data, it might be reasonable to say that I was using a biased sample. After all, newspaper headlines and modern fiction can often be sensationalist or rely on merely metaphorical turns of phrase. But, commonsense attributions of mentality to collectivities aren't reserved for the hyperbolic prose of newspapers, contemporary literature, and science fiction. In fact, recent social psychological data suggests that commonsense psychology is quite willing to attribute mentality to a number of actual collectivities. In a review of linguistic data, Bloom and Kelemen (1995, 25) found that "collective nouns, such as *family*, *bunch*, and *army*, refer to sets of objects that bear some salient and enduring relationship with one another, either by being spatially or perhaps physically connected like the grapes in a bunch, or by having more abstract social connections". Noting this diversity in the sorts of considerations that underwrite judgments of entativity,[8] Bloom and Kelemen (1995) argue that such judgments are best understood as grounded in the commonsense theories that we adopt in making sense of the world around us.

[8] By 'entativity' I mean to refer to those judgments about what counts as a single entity.

However, this just raises a question about what sorts of commonsense theory could underwrite a judgment that a particular system is capable of intentional action.

One option here stems from the psychological literature on theory of mind ascriptions. In a well-known experiment, Heider and Simmel (1944) presented volunteers with a short animation consisting of simple geometric shapes moving around the screen. When volunteers were told to 'write down what happened in the picture,' most of them offered interpretations of the animation in terms of the purposeful actions of animate beings. Heider and Simmel took these responses to suggest the presence of theory of mind mechanism that generates ascriptions of mental states in any case where such ascriptions facilitate explanations and predictions about the behavior of an entity. On the basis of this hypothesis, Paul Bloom and Csaba Veres (1999) have collected data suggesting that this system can also be brought on-line in order to facilitate the ascription of psychological states to some sorts of collectivities. Using computer simulations based on those that were used by Heider and Simmel, Bloom and Veres found that in conditions where subjects were presented with collections of objects *moving in an apparently unified way*, almost all subjects described the animation in terms of the intentional states of groups (e.g. 'the blue circles tried to stop the green triangles').[9] This seems to suggest that there are some conditions under which people are willing to attribute mentality to collectivities.

---

[9] Is there any reason to believe that subjects intended these ascriptions literally? Of course not; however, this does not go against the general point that I wish to make about commonsense psychology. The point is merely that commonsense psychology utilizes the same mechanisms for ascribing mental states to individuals and to collectivities. With this in mind, two further questions must be addressed then. First, does commonsense psychology take attributions of mentality to collectivities literally? Second, should cognitive science take them literally? I have little to say about the first question, though Arico, Fialla, and Nichols (unpublished data) have collected some evidence for the claim that ascriptions of mental states to collectivities are treated literally from the standpoint of commonsense psychology. The remainder of this thesis is dedicated to answering the second question.

We might wonder at this point, however, whether there are any actual cases where we see the behavior of some group of individuals as unified in the right way for us to track them using this sort of attribution. As we know, there are some intentional actions that can only be carried out collectively. Moreover, we typically make sense of these activities by ascribing intentional states to the collectivities in question. Consider the following examples:

- No single individual can play Steve Reich's *Music for 18 musicians* (according to Reich, the piece should be played with a *minimum* of 18 musicians—more musicians are preferable to prevent doubling on instruments). If this piece is to be played, it will be necessary for a group of individuals to intend to play it.

- No single individual can play a Balinese gamelan; a gamelan can only be played collectively. So, in order to play a gamelan, a number of people have to collectively intend to produce a single piece of music.

- In the King Crimson song "Frame by frame", from the album *Discipline*, Robert Fripp and Adrian Bellew play similar single note melodies on two guitars.[10] One guitar plays in 13/8, the other plays in 14/8. This creates an offset metric that grows and shrinks over 7 measures of 14/8. Playing the multi-meter in this song is something that neither Fripp nor Bellew could do on their own, though it is something that they intend to do together—and in fact they execute it perfectly as a joint activity.[11]

- No single individual is capable of running the Princeton Offense, the Flex offense, or of playing a zone defense in basketball; however, these are things that teams, under the direction of a knowledgeable coaching staff often intend to do.

- I can't carry a piano by myself; however, Carlo and I have moved a piano together. In order to successfully carry the piano, Carlo and I had to intend to do this together.

Of course, the mere fact that we explain these behaviors in collective terms from the standpoint of commonsense psychology doesn't, by itself, commit us to the actuality of collective mentality. These examples do, however, demonstrate that commonsense psychology is at least open to the possibility of collective intentional action. Noting this,

[10] Thanks to David Ripley for pointing me to this piece as well as helping me think through this example.

[11] Note, however, that a particularly proficient drummer could play the analogous multi-meter on her own. Unfortunately, the world has very drummers that are so proficient at playing their instruments.

philosophers, and theoretically minded social scientists, have argued that there are some actions that are possible only by way of collective action; some philosophers have even argued that collective intentions are thereby required to explain how such actions are possible. However, things in this area become quite complicated very quickly.

In an analysis of newspaper headlines containing ascription of intentional states to collectivities and individuals, Menon et al. (1999, 702), found that "while prevailing American theories hold that persons have stable properties that cause social outcomes and groups do not, the theories prevailing in Confucian influenced East-Asians cultures emphasize that groups have stable properties that cause social outcomes".[12] They suggest that while Americans are willing to engage in some ascriptions of mentality to collectivities, they are actually far less willing to do so than their Asian counterparts. Building on this data, as well as data of their own, Kashima et al. (2005, 149) have argued that there are two characteristics of systems that might underwrite the judgments of entativity that would allow for a literal understanding of different sorts of intentional actions. On the one hand, perceived internal consistency (i.e., the extent to which perceptions of individuals that belong to a group are likely to resemble one another in appearance and behavior) and perceived unalterability (i.e., the belief that the properties of a collectivity aren't changeable because it has some underlying essence) seem to play a key role in some judgments of entativity;[13] on the other hand, considerations of agency (Kashima et al. 2005, 150) are also at play, and sometimes seem to be doing all the work.[14] Kashima et al (2005) found that insofar as being

[12] Cf., Morris, Menon, and Ames (2001) for evidence suggesting that East Asians employ a conception of agency that allows a collectivity to count as a single entity and Kashima, et al. (1995) for evidence that suggests that such considerations of agency are capable of explaining the intentional actions of a collectivity.

[13] The thought here echoes considerations about natural kinds in folk biology that there is some essence to being an individual that is best understood in terms of some sort of internal mechanism (cf., Keil 1989)

a single entity is understood in terms of psychological essentialism (i.e., in terms of considerations of consistency and unalterability), individuals are perceived to be more entity-like than collectivities cross-culturally. Considerations of agency, however, are applied to individuals more often than collectivities in English-speaking and continental European cultures, but not in East Asian countries. (Kashima et al. 2005, 162). This suggests that people who are raised in East Asian cultures are far more willing to ascribe agency to collectivities than are Westerners.

Moreover, while there are significant differences in the number and sort of entities to which Americans and Asians are willing to concede agency, Menon et al. (1999, 702) have found that there are a number of cases on which the judgments of Americans and Asians seem to overlap. For example, almost everyone is willing to ascribe at least some mental states to collectivities. Recent data collected by Josh Knobe and Jesse Prinz (forthcoming) suggests that American volunteers are likely to ascribe a wide range of cognitive states to corporations. Knobe and Prinz presented subjects with sentences ascribing either cognitive states (e.g., beliefs, intentions, and desires) or phenomenal states (e.g., experiencing great joy, getting depressed, and vividly imagining) to corporations and asked them to judge the naturalness of the ascriptions. Volunteers found ascriptions of cognitive states far more natural than ascriptions of phenomenal states to collectivities.[15] In another study, where subjects were presented with sentences ascribing emotional states to corporations (e.g.,

[14] Note that 'agency' is not intended to pick out any philosophically robust kind; it is merely intended to pick out "the extent to which a social being is attributed mental states such as beliefs, desires, and intentions' (Kashima et al. 2005, 150).

[15] Subjects were asked rate sentences on a scale from 1 ('sounds weird') to 7 ('sounds natural'). "The mean ratings were as follows: *Non-phenomenal states:* 6.6: Deciding; 6.6: Wanting; 6.3: Intending; 6.1: Believing; 5.2: Knowing; *Phenomenal states:* 4.7: Experiencing a sudden urge; 3.7: Experiencing great joy; 2.7: Vividly imagining; 2.5: Getting depressed; 2.1: Feeling excruciating pain" (Knobe and Prinz, forthcoming).

Microsoft is upset) and ascriptions that contained both emotional terms and the word 'feeling' (e.g., Microsoft is feeling upset), Knobe and Prinz found that subjects even took the ascription of emotional states to collectivities to be natural so long as the emotional state was not overtly picked out as a phenomenal state.[16] Knobe and Prinz take this to demonstrate that commonsense psychology is willing to allow for that ascription of a large number of mental states to collectivities—what it precludes is the ascription of phenomenal states to collectivities.

In coordination with Hagop Sarkissian and Michael Bruno, I have also ran a similar survey of willingness to ascribe mental states to collectivities. We replicated Knobe and Prinz's findings; but we also found that although American subjects differ considerably in their willingness to ascribe mental states to individuals as opposed to groups, subjects in Hong Kong do not. That is, although Americans think that it is more often acceptable to ascribe mental states to individuals than groups, a similar difference is not present in East Asian volunteers. This suggests that the willingness to ascribe mentality to collectivities might be, at least partially, an artifact of cultural conditioning. But unfortunately, this can't be the end of the story.

We've known for a long time that commonsense psychology is willing to ascribe mental states to collectivities. However, the question has always been: how are we to make sense of commonsense ascriptions of mentality to collectivities? Although there is good empirical data suggesting that people tend to conceive of their ascriptions of mentality

[16] Using the same scale as before, the mean responses were:

	With 'Feeling'	Without 'Feeling'
Upset	1.9	5.3
Regret	2.8	6.1

.

literally as opposed to metaphorically (cf., Arico et al, forthcoming), this does not rule out the possibility that they are instrumentally useful but false claims or that they are hyperbolic assertions of some other form.[17]

1.2. Foundations for a theory of collective mentality.

So far, I've argued that commonsense psychology sometimes allows for the ascription of mentality to collectivities in order to explain apparently intentional actions. I've also argued that people often take such ascriptions literally. However, the mere fact that people often ascribe mentality to collectivities cannot, by itself, tell us whether collectivities ever do have mental states. Perhaps none of these ascriptions are true *in the same way* that ascriptions of mentality to my housemate or my cat are. And, at this point I have offered no resources for determining what facts about collectivities could possibly make these ascriptions true. If I am to answer questions about the literal truth of ascriptions of mental states to collectivities, I must offer an account of what it takes for a system to have genuinely mental states.

Thus, rather than offering an account of collective mentality, I must begin by offering a brief account of what it takes for a system to count as genuinely psychological. Once I've developed this account, it will offer the frame for an answer to questions about the possibility of collective mentality. In the absence of a more general theory of the mental, there would be no way to demonstrate that the mental states of a collectivity should be viewed as belonging to the same kind as the mental states of an individual. Thus, without a more general theory of

[17] Here's the relevant data from Arico et al (in prep). Subjects were asked to judge the literalness of the following sentences (on a scale of 1=figurative to 7=literal). 1) Some corporations <u>*want*</u> lower taxes; 2) Some millionaires <u>*want*</u> lower taxes; 3) Many corporations are <u>*overjoyed*</u> by a strong economy; and). Microsoft feels <u>*sad*</u> when it loses customers. With 67 participants, the means were as follows: 1: 6.12, 2: 6.21, 3. 4.33, 4. 3.25. There is no statistically significant difference between 1 and 2 (t =-.564, p=.575) and there's a pretty good correlation between them (r=.323, p = .008). However, there's a statistically significant difference between responses 1 & 3 (t=7.735, p<.001) and also between 1 & 4 (t=11.559, p<.001).

mentality, it would be unclear why the states of a collectivity should be seen as mental states at all. In the remainder of this chapter I offer an outline of a theory of mentality that will provide conditions for the literal construal of claims about collective mentality. The key point to note at this point is that the theory of collective mentality I develop, any attribution of mentality—be it to an individual or a group—will have to be held to the same standards. My contention is that once we examine our best philosophical and psychological theory of mental states, we find a theory that applies both to collectivities and to individuals. In developing a theory of collective mentality, I must, then, attend to the following sorts of considerations.

1) An adequate account of collective mentality must demonstrate that the domain of psychological explanation is not exhaustively specified by appeal to individuals. Instead, psychological generalizations apply to both individuals and collectivities.

2) An adequate account of collective mentality must demonstrate that psychological explanations are autonomous from facts about their realizers: individual mental states can be understood independently of their neurophysiological realizers; and, collective mental states can be understood independently of the mental states of the individuals that compose that collectivity.

3) An adequate account of collective mentality must distinguish between those systems that have genuine intentional mental states and those systems that merely behave as-if they had mental states (that is, we need a non-behaviorist account of mentality that applies to both individuals and collectivities). Any theory of mentality should distinguish between true believers and systems that are merely usefully described using mental terminology.

To begin with, it will help to get some structure on the table. The most promising view of the mind currently on offer suggests that the study of minds must occur at (at least) three levels of analysis. In his seminal work on the visual system, David Marr (1982) claims that an adequate theory of a cognitive system must explain phenomena at three distinct, though interrelated, levels of explanation. First, such a theory must explain what a system does as well as why it; Marr calls this the computational level of explanation. Second, such a

theory must explain the behavior of the system in terms of the representations over which it runs computations as well as the transformational rules governing the manipulation of representations; Marr calls this algorithmic level of explanation. Finally, such a theory must explain how the physical structure of a system is able to implement the algorithmic and computational structure of that system. Over the next several subsections, I will explain how these three sorts of explanations are relevant to our understanding of cognitive systems in general, as well as how they can be brought to bear on our understanding of collective mentality.

1.2.1 Computation and functions: I begin with the computational level of analysis for a cognitive system. According to Marr, such analyses are attempts to make sense of what a particular system does and why it does it, at a fairly high level of abstraction (Marr 1982, 20). This claim, however, fails to offer anything in the way of a model for giving a computational analysis of a particular system. After all, there are many, perhaps innumerably many ways of answering questions of what something does and why. However, we can get a more adequate idea about where to start by looking to the ascription of mental states in commonsense psychology. This does not, of course, mean that there will not be reason to revise and systematize commonsense psychology as scientific data are acquired. In fact, commonsense psychology might be wrong about a whole host of issues concerning cognitive systems; there may be some cognitive structures that fail to be adequately captured by commonsense psychology and there may be some cognitive structures that are not present as they are posited by commonsense psychology. However, commonsense psychology does suggest a number of avenues for inquiry into the cognitive structures that must be posited in order to explain the behavior of cognitive systems.

I, thus, propose to *start* by adopting a sort of intentional realism that posits internal representational states like beliefs and desires. Adopting this sort of intentional realism, then, opens up the possibility of adopting and account of psychology that is commonly known as functional analysis. Functional analysis begins by looking to the explanations offered by commonsense psychology, introducing modifications to the theory where necessary and attempting to make these explanations more systematic. We begin by noticing that the paradigmatic ascription of mental states like beliefs and desires occur where making such an attribution is the best explanation of a system's behavior. That is, we typically individuate mental states by their causal consequences. Another way of putting this point is to note that functional analysis occurs when we ask of part of a system *what role it plays* in the activity of the system as a whole. To see how such functional analyses work, it will help to start with some examples.

Consider the various ways that we might describe the parts of an internal combustion engine (cf., Fodor 1968). Adopting some terminology in such a description entail a commitment to the existence of particular structures, while other terminology entails a commitment only to functional characterizations of a structure. For example, referring to an engine component as a 'camshaft' carries with it a commitment to the existence of a cylindrical mechanism with a number of protruding lobes that are used to operate poppet valves. Thus, in finding that an engine contains a camshaft, one already learns a number of facts about the structure of the engine. However, if a device is merely referred to as a 'valve lifter' this carries with it only a commitment to functional characterization—there are many ways to lift a valve! Now, in speaking about valve lifters in general, there will be a lot of things that can be said about what a valve lifter *does*; however, being a valve lifter is not

reducible to the structural properties of an engine component *because* valve lifters are explicitly defined by their function. Similarly, what it is to be a poison is functionally defined (cf., Armstrong 1980).[18] Poisons are substances that have the function of causing a system to sicken or die when introduced and they have this function even when they fail to exhibit these causal powers (e.g., when blocked with an antidote).[19]

The important question for my purposes, however, is whether it is right to take mental states to be functionally characterized in the way that valve lifters and poisons are. The first thing to notice in making this claim is that many, if not all mental states are purposive. For example, if I want to write a song, this desire (when couple with the right sorts of beliefs and physical capacities) will cause me to pick up my guitar and start playing. There are, of course, many behaviors (picking up the guitar could be a getting-ready-to-write-a-song behavior, a straightening-up-the-living-room behavior, or even an ignoring-my-house-mate behavior) that are caused by many different psychological states, and this makes it more difficult to explain exactly what the function of a particular mental state is. Fortunately, we do know a more about how the function of a mental state is to be individuated.

We know, for example, that many of our mental states are intentional or representational. Beliefs, desires, and perceptions all represent the world as being a certain way. For example, I have a number of beliefs about the cup of coffee that I am currently drinking. I believe that it is starting to cool off, and that it has nice chocolate and nutty

[18] Note that in introducing this example Armstrong is not defending a functionalist account of the mind. However, his example of a poison nicely demonstrates a number of features of functional analysis. I use this example without following Armstrong's (1980) causal theory of the mind. I also disagree with Armstrong that it is a part of the *meaning* of concepts like POISON, BELIEF, and DESIRE that they have a particular causal structure.

[19] Note that this leaves a couple of empirical questions open. First, we can't determine *a priori* what is and what is not a poison. Second, it leaves open the mechanisms that make something poisonous. This will become important shortly.

overtones. I also desire that my coffee be hotter and less bitter. Each of these thoughts represents the world as being a certain way and each of these thoughts takes and intentional object (i.e., the cup of coffee in its current form). As Fodor (1980) puts the point, representations understood in this way have at least two dimensions.

First, mental states *qua* representations have semantic content. This allows for a distinction between the belief that this coffee has nice chocolate overtones and the belief this cup of coffee is getting too cold. I distinguish these thought by the fact that they are about different things—in one case, the content is, in part, the temperature of the coffee, and in the other the flavor, but not the temperature, is important. Second, mental states stand in some relation to their content. Thus, my belief that a cup of coffee has a particular flavor and my wish that it did are distinguished by way of the relation between my mental state and the world. One way of putting this point is in terms of direction of fit—beliefs are meant to fit metal states to the world and desires try to make the world fit them. This claim about direction of fit is, of course, not subtle enough to make the fine discriminations that we can make between different sorts of mental states. However, whatever the complete story is, there must be some difference in the way that different types of mental states relate to the world.

Another thing that has become a commonplace assumption about mental states is that they are not only semantically evaluable, but they are intimately tied to the production and control of behavior *in virtue of* their semantic content. My belief that this cup of coffee is too cold when coupled with my desire to have a nice warm cup of coffee will, *ceteris paribus*, cause me to get a new cup of coffee. My hopes concerning the well-being of migrant farm workers and my belief that attending a rally will make it more likely that people will take note of the lack of healthcare options for migrant workers will, *ceteris paribus*, cause me to

go to migrant farm worker's rallies. Moreover, these sorts of representational states facilitate the pursuit of a particular goal in a way that's flexible. As William James (1890) notes, we attribute a desire for food to a frog because when we prevent the frog from getting food by putting a glass barrier in its tank, it will modify it's behavior in an attempt to get around the barrier. This sort of state stands in sharp contrast to the merely metaphorical attribution of the desire to make it warmer in the house that I might attribute to my heating unit. After all, if I turn off the switch on the thermometer, the heater will not attempt to modify its behavior in a way that will allow it to fulfill this desire. This leads to the supposition that there are states of some systems, call these the mental states, that have the purpose of producing and modifying behavior.

Spelling out the nature of these mental terms as they are individuated by their function from within commonsense psychology is by no means the end of the story. However, it is at least a point at which we can begin to inquire into the nature of the mind. Provided there is some functional characterizations of a particular cognitive state (e.g., belief, desire, or visual experience), we come to a second sort of question: what needs to be the case for a system to execute these functions. In thinking through this issue, it helps to think of a function as an abstract entity that takes an input and uses an algorithm of some sort to map this input onto some output. Ned Block (1978), for example, claims that functionalism about the mind just is the thesis that "each type of mental state is a state consisting of a disposition to act in certain ways, *and to have certain mental states*, given certain sensory inputs and certain mental states." In explaining how a mental states can do this, however, it is necessary to move to Marr's algorithmic level of explanation.

1.2.2 Algorithms and RTM: At the algorithmic level of explanation, it is necessary to commit to both a range of representations that can be used by a system and to a set of transformational rules that operate over these representations. The functional characterizations that are developed at the computational level do not *entail* any particular theory about representations; however a functionalist theory of a cognitive system whose outputs are semantically evaluable as well as causally efficacious sits quite well with a representational theory of mind (hereafter RTM). In response to considerations about the semanticality and causal efficaciousness of human thought, RTM was self-consciously developed as a theory of cognitive systems that explains how the functionally characterized mental states of psychological explanation can be semantically evaluable as well as causally efficacious.

Developing a plausible representational theory of the mind is, of course, quite complicated. So, it will help to begin with a better idea about what the proponent of RTM actually claims. According to RTM, mental states are best understood—at the algorithmic level of explanation—as relations between a cognitive system and a mental representation. Fodor this formally as follows:

> For any organism O and any proposition P, there is a relation R and a mental representation MP such that: MP means that (expresses the proposition that) P; and O believes that P iff O bears R to MP (Fodor 1990, 16)

According to proponents of RTM, the algorithmic level of explanation for mental states should be seen as an attempt to discover the sorts of representations over which a cognitive structure operates as well as the syntactic transformations utilized by the system in order to carry out some function.

The most promising explanation for why we should think that the mind is a representational system, as supposed by proponents of RTM, turns on an analogy between thought and language. This argument begins from the assumption that thought, like language, is infinitely productive. That is, from a finite stock of primitive representations, we are able to construct an infinite number of complex thoughts. For example, just in virtue of having the constituent concepts, you can have the thought FRANK ZAPPA AND BENITO MUSSOLINI USED TO TANGO IN PARIS WHILE SMOKING CATNIP FROM A TWELVE FOOT BONG—though that thought probably had not ever crossed your mind before reading it. Fodor, thus, contends that any viable theory of mental states like beliefs and desires must be able to account for the boundlessness of thought. In a natural language, we have an easy story about how each sentence decomposes into sub-sentential components. From this stock of sub-sentential components, different sentences can be arranged by simple recursive rules, and the meaning of a complex sentence can be determined in a regular way by its constituent structure. All we need is a base of words, a set of syntactic rules, and a series of transformation rules, and you're off and running. The assumption is that if we had a story about how propositional mental states could be built out of things that are sub-propositional then we could have a parallel story for thought. If there were a language of thought that paralleled the structure of natural language, then we would have such a story. A language of thought is a structure of syntactic rules and mental representations as constituent semantic structures—so RTM follows.

RTM provides a story about the semanticality of thought. However, RTM must also offer a story that explains how semantic states can also have causal powers. On this point, Fodor has argued that we should think of the mind as computational system. His reasoning

here is that if the mind is a computer (call this the computational theory of mind, or CTM), this provides a story explaining how non-arbitrary content relations among causally related thoughts can be possible. The representational states of a computational system are capable of being transformed into other representational states or into output states merely in virtue of their formal properties because of the way that the computational structures of the system are organized. Now, if the mind is computational, then there have to be some mental particulars that have syntactical properties. Just as we see in a computer, transformation rules and data structure will be represented in the architecture of the system. In this case we would have a story about how the operations over syntactic primitives could give rise to semantically evaluable states. Elaborated in this way, thought is not just representational it is also computational in the sense that mental states understood in this way are symbolic (i.e., they are defined over representations) and they are formal (in that they apply to representations in virtue of their syntax).

With this computational theory of mind in hand, it is possible to turn to explanations at the level of the implementation (Marr 1982, 22) for these semantic and syntactic structures. The thought here is that it must be possible to move from a how-possibly story about the semantic and syntactic structures of a mind, to a how-actually story that explains how the sorts of computations that we ascribe to a system at other levels of analysis can be physically implemented in a system.

1.2.3 Implementation and realization: One approach to questions about implementation is to avoid them and adopt the instrumentalist project suggested by Daniel Dennett (1978a, 1987a, 1987b, 1991b). Dennett has spent much of his career attempting to undermine 'industrial strength realism'. In order to achieve this, Dennett distinguishes

between two claims might be made about the existence of mental states. Given that we can't directly observe mental states, and that we have to infer their existence, there are two ways in which this inference can be carried through (cf., Reichenbach 1938). First, they mental states might be *illata*: independently existing entities whose existence is inferred from observable phenomena, but which are themselves unobserved. Second, they might be *abstracta*: abstract objects that exist only within a theoretical framework—the existence of which is settled by way of theoretical convention. Dennett (1991b) contends that mental states are *abstracta* and that we need not be concerned for the purposes of psychological theorizing about whether these states are implemented at the level of neural architecture. We are interested in such *abstracta* as beliefs and desires because, and only because they allow for the prediction and explanation of behavior, empathetic responses to others, the organization of memories, and the interpretation of emotions (Dennett, 1991b).

Adopting a view that treats mental states as *abstracta* turns on understanding the ascription of mental states exclusively in terms of adopting the intentional stance. To a first approximation, adopting the intentional stance is a matter of treating a system whose behavior you want to predict as a rational agent with beliefs, desires, and other mental states exhibiting intentionality. Dennett claims that a system whose behavior is predictable on the assumption of rationality, and whose behavior cannot—for practical purposes—be explained merely in terms of its physical structure, is *in the fullest sense of the word* a believer. [20]

[20] There are, of course, a number of other arguments that can get you to this point. Dennett often argues in the following way. Recent neuroscience suggests that the brain is an inherently plastic system (cf., Churchland 1979, and Ramachandran 1993) and that the brain structures that produce any complex behavior are likely to be distributed across multiple heterogeneous brain regions (cf., Clark 1989, and McClelland and Rumelhardt 1986). Given these facts about the human brain, it is quite likely that the neural structures that could realize beliefs and desires (if there are any) are likely to be extremely plastic and distributed across multiple brain structures especially since the constituents of beliefs and desires are likely to be tied to particular long term and working memories (cf., Prinz 2002 and Barsalou 1987) of particular agents—and our best neuroscience suggests that memories are multiply distributed if anything is (cf., Cabeza and Nyberg 2000 for a review). Thus,

But why should we believe Dennett on this point? So far as I can see it, we have two choices. Either we can follow Dennett (1978a, 1987a, 1987b, 1991b) and take mental states to be the abstracta that we use in ascribing mental states, or we can take the behavioral regularities that are predicted and explained by way of our mental state ascriptions to be evidence for the presence of some underlying causal mechanism that gives rise to such states.[21] Dennett is certainly right that when we find some system for which the intentional stance works, we endeavor to interpret some of its internal states as internal representations (Dennett 1978a). However, there are some very good reasons for taking mental states to be *illata* (cf., Lewis 1972, Fodor 1968, 1989, and Lycan 1987, et al).

As Jerry Fodor (1987, 16) puts the point, "We have no reason to doubt that it is possible to have a scientific psychology that vindicates commonsense belief/desire explanation." However, Dennett's claim that beliefs and desires are *abstracta*—when he is pushing more industrial strength versions of his instrumentalism—is partly grounded on the empirical claim that generalizations applicable at the neurophysiological level of explanation will not be sufficient to justify the sort of isomorphism psychological realism requires between kinds in commonsense psychology and kinds in neurophysiology. The psychological realist, however, has a response to this claim. This leaves industrial strength instrumentalism in the following awkward position: if there is no way to vindicate the

---

the tokens of a particular type of belief are unlikely to have enough structural properties in common to explain why they are tokens of that type. Now, if our best neuroscience finds no way of mapping all the tokens of a particular belief to underlying neural structures, then so much the worse for the inner cause story of mentation. Moreover, Bill Lycan (1988, 518-519) has suggested two other reasons for Dennett's instrumentalism: 1) Dennett's objections to the language-of-thought (the most plausible inner cause theory) and 2) Dennett's implicit commitment to verificationism about meaning. Dennett concedes these as his reason for instrumentalism, though he thinks that appealing to verification conditions in the absence of an underlying causal mechanism is innocuous (cf., Dennett 1988, 543).

[21] NB: If I were to adopt the former strategy, my work here would be done. Provided that there are cases in which the behavior of a collectivity were best predicted in terms of intentional ascriptions, that collectivity would have mental states. On Dennett's brand of instrumentalism, collective mentality follows straight away!

functional patterns of commonsense psychology by appeal to structural realizers *of some sort*, then there is little reason to think that we wouldn't be better served by adopting patterns of explanation that were couched in terms of neurophysiological states (or whatever the relevant patterns are) and abandoning all talk of beliefs and desires. But, as Fodor (1987, 10) puts the point, "we can't give them up *because we don't know how to*. So maybe we had better try to hold on to them".[22]

Given that all parties to this dispute recognize that it is an empirical question whether mental states are realized in a way that allows them to count as *illata, w*e cannot say, *a priori*, that the relevant sorts of states won't be found in the brain or in whatever else happens to realize a mind. Moreover, because there is currently no overwhelming evidence on either side of this issue, it strikes me as reasonable to look for some story about the sorts of states that we are tracking with our belief talk that will allow them to be viewed as *illata*. This story will probably be told, at the end of the day, in terms of the architectural features of a system that facilitate computations over intentional states in a way that yields beliefs. What this class includes is, at least currently, not easy to settle. Moreover, the mere fact that we have a difficult time articulating the neurophysiological realizers of particular beliefs, for example, doesn't mean that there is no interesting class of realizers that cluster as a natural kind at both neurological and psychological levels of explanation. And even Dennett, in some moods, agrees with at least this claim.

However, Dennett also argues that even if you want to defend a view of mental states as *illata* that are grounded on certain sorts of computational structures, there are serious

[22] Dennett is, of course, unclear about his position on this point. In some cases, Dennett would agree entirely with this sort of position. Dennett has, indeed, at points exhibited a sort of eliminativist tendency. However, Much of Dennett's work is also grounded on humanist assumptions that preclude the possibility of eliminativism about mental states.

worries about how the brain could possibly implement a semantic engine. Although the mind seems to be a semantic engine, any architecture of the brain seems only to be able to realize a syntactic engine (Dennett 1987b). Neural structures just are not capable of doing anything more than discriminate structural, temporal, and physical features of inputs. Moreover, the brain is an entirely mechanical system whose activities are governed by the syntactic features of inputs by way of (likely) incredibly simple transformation rules. However, if CTM is to be an adequate theory of minds, then there must be a story about how the brain manages to get from syntax to semantics. Given that syntax cannot, by itself, determine semantics, this seems to generate an unbridgeable gap. However, Dennett also notes that there is good reason to think that a purely syntactic system could be designed in such a way that it approximates a semantic engine.

Promising strategies for such approximations emerge when we consider analogies to other cases of approximations in biological systems. Consider the animal that needs to know when it has found and eaten food. In many cases this organism will settle for a friction-in-the-throat-followed-by-stretched-stomach-detector, a mechanical system that can be tricked but that works pretty well in its normal environment (cf., Dennett 1987b). To consider another philosophically commonplace example of such approximations, we can note that magnetotactic bacteria succeed in avoiding deadly oxygen rich waters without oxygen detectors. In their natural environments, these bacteria utilize a set of magnetosomes that ensure that they are constantly impelled towards magnetic north. This mechanical system that can be tricked by placing a southern-dwelling bacteria in northern waters; however, it does well enough to get these bacteria around in their natural environments. Borrowing from this sort of model, Dennett claims that if we are to explain how to get semantics from syntax, “the

system has to be put together as a bag of tricks that functions to pick out and type classify stimuli, filtering out irrelevant data, in the end *seeming* to discriminating meanings by actually discriminating things (no-doubt tokens of wildly disjunctive types) that reliably covary with meanings" (Dennett 1987b, 63).

The only way in which we could possibly specify mental states as *illata*, then, is by doing sub-personal cognitive psychology on the design specifications for a cognitive system. Making the ontology of belief an empirical question, however, we are faced with a potentially troubling result. If we find that the brain does not include systems that can facilitate the computations required for beliefs, we must worry that the system is not actually a believer. To put the point another way, if we posit black boxes that cannot be causally sustained by mechanisms internal to a system, a theory that ascribes beliefs to that system has got to be mistaken!

This argument does not, however, tell against computationalism *per se* without a number of additional premises. Surely a computational model that offers an account of the mind that doesn't refer to the world probably isn't going to be viable; however, this does leave open at least one version of computationalism open. If we want to have a syntactic engine, that is at least virtually a semantic engine, then there will have to be some syntactic relations that reliably covary with semantic relations (Dennett 1987b, 63). However, this requires constantly checking outside the system to see how the internal states of a cognitive system operate in that system's natural environments, determining how it responds to different stimuli as well as whether the states of that system actually covary in the right way with states of the environment. It is only in this way that we begin to make sense of both the

function of mental representations as well as the requisite flexibility that must be present in mental states.

At this point, we see that the way to start with the interpretation of a system as a cognitive system is to work from the level of commonsense mental ascriptions, use such ascriptions to construct an account of the various functional tasks that a system can undertake, and then to see whether there are any syntactic structures at place in the system that covary with the semantic states ascribed to the system. This however, suggests that we cannot make assumptions about what sorts of systems will have semantic states. It is only by working through what it takes to be a mental state (at the level of psychology) and then checking for the right sorts of isomorphism with some realizers that we can ascribe genuinely mental states to a system. Now, if there are good reasons to attribute mentality to collectivities from the standpoint of psychology, the next question will be whether there are any states of the system that stand in the right relation to semantic states. Surely this is not a question that can be answered by dogma—only doing the empirical inquiry can answer the question about whether collective mentality is possible.

1.2.4 The autonomy of commonsense psychology: Insofar as commonsense attributions of mentality are concerned, we do not typically seem to care about the implementation of cognitive states; this brings us to a final component of an adequate theory of mentality: how can psychological explanations be autonomous from claims about their realization. Fortunately, taking mental states to be functional kinds suggests an initial story about the autonomy of psychological explanation. My claim, here, follows Dennett's in noting that Laplacean Martians who could predict the movement of every physical particle in

the universe would still be missing something perfectly objective.[23] Dennett explains this in terms of the patterns in human behavior that can be described only by adopting commonsense functional psychology. I agree with Dennett that the Martians wouldn't be able to see the right sort of counterfactual stabilities.[24] While they might be able to compute some counterfactuals, they would not be able to see the indefinitely many unique patterns of physical motions that could be substituted for particular physical realizers without perturbing the patterns of human behavior.

Let me try to spell this point out with an analogy. Consider the question: What makes something a carburetor? One way to answer this question is by an appeal to the fact that its operation corresponds to a function detailed by the theory of internal combustion engines. Now, suppose we want to know what the structure of a particular carburetor is. There will be an account of the physical parts out of which a particular carburetor is made; however, before we can begin to investigate the relevant mechanisms in an engine, we have to have a theory about what carburetors are. Otherwise, we would have no criteria for determining which parts of the engine constitute the carburetor. That is, we need to have a theory of carburetors that is stable enough and projectable enough to pick out a carburetor in any internal combustion engine that we approach—even in cases where the particular mechanism that is doing the carburetion is one that we haven't encountered before.

---

[23] I am quite fond of Dennett's argument on this point. However, he typically denies functionalist theories about the mind; I, on the other hand, think that there is no reason why we cannot adopt Dennett's argument on this point without denying functionalism.

[24] Both Frank Jackson and David Braddon-Mitchel argue that there is no reason to suppose that the Martians wouldn't be able to compute the counterfactuals. However, the point here is a bit more subtle. In order to be able to compute *the right sorts* of counterfactual stabilities, the Martians would have to have the capacity to track the relevant class of behaviors that constitute a particular sort of intentional state. However, because *at the physical level* these states are quite heterogeneous, they wouldn't be classified as belonging to a particular kind *except* by way of psychological explanation. See the next four paragraphs for an elaboration of this point.

Similarly, if we want a genuine science of psychology, we need to know is how to generalize over psychological states in a way that is stable and projectable. However, the only way that we can get to these sorts of generalizations is to begin at the level of psychological explanation.[25] Had we not adopted a view of functional psychological, we never would have picked out the kind at the neural level in the first place. The reason here is that the similarities at the neural level (e.g., similarities in the sorts and density of cells or in the range of tasks for which different areas show activity) do not always recapitulate similarities at the psychological level, so were we to start at the level of neurophysiology, we might end up carving up the world in a radically different way. Neurophysiological states just don't cluster into the right sorts of patterns for us to start with generalizations about them and infer upwards to psychology.

To put a finer point on this claim, it's a well-known worry about functional magnetic resonance imaging (fMRI) that neuroanatomical localization is highly variable across subjects. Merely looking to the areas that happen to be active at a particular time is never sufficient for determining what the function of a region is. It is only by deciding on a functional task before hand (hence the f in fMRI) that it's possible to understand the tasks to which a particular region of cortex is dedicated. However, once a task is functionally specified, we find that particular regions of the brain are active for particular sorts of tasks in a single individual. Now, while our current imaging techniques are insufficient to figure out *precisely* what areas are active in which tasks, it is plausible to think that we will eventually be able to discriminate distinct areas of cortex that are dedicated to particular sorts of

[25] The exact starting point here is going to vary from case to case. In some cases (e.g., beliefs and desires), the best place to start is with folk psychological ascriptions. In these cases, the intuitions will have to be made rigorous by constructing a theory of such states. In other cases, the relevant sorts of phenomena will be scientific psychological phenomena (e.g., attention, long term memory, or semantic memory) that do not have clear equivalences in folk psychology.

functional tasks, but it is only on the assumption of functional characterization that the data collected in neurological studies is interpretable.

However, things get even worse for a theory that starts at the neurological level. If functionalism is true, then there are lots of ways of realizing such kinds that don't happen to be instantiated in the actual world right now. And here's the important point. While we do get token identities between psychological states and their structural realizers, this does not entail much of anything in the way of a reductive story about psychology. As Fodor (1968) puts the point, functional kinds (e.g., psychological states) are not easily seen as being capable of being micro-analyzed in any way; after all, the mere fact that we have identified a certain mousetrap with its physical structure does not commit us to thinking that all mousetraps have to be built like that—otherwise it would be impossible to build a better mousetrap.

1.3. Functionalism, CTM and collective mentality

At this point I've collected all the tools necessary for constructing a theory of collective mentality that shows how they are analogous to the mental states of other cognitive systems. I'll start to develop this theory of collective mentality by considering an argument offered by D. H. M. Brooks (1986) in his paper "Group minds". Brooks argues that accepting functionalism entails at least some cases where we would be warranted in attributing mental states to collectivities. I'm inclined to think that the sort of story he tells is insufficient in the end, however it does provide a foundation, and with some elaboration this view becomes incredibly plausible.

If materialism is true,[26] then there will be some supervenience relation between the mental and the physical. What this relationship is, however, is far from immediately obvious. One way of looking at this relationship is in terms of type identities between mental states and brain states. However, alternatively, and far more plausibly, we can adopt a functionalist view with token identities between mental states and brain states.[27] Type identity seems too strong to adequately capture our standard understanding of the mental. After all, we should expect no psychological change in a person where one of her neurons is replaced with an artificial neuron (Brooks 1986, 456). But if replacing one neuron with an artificial neuron maintains psychological identity, there is no reason to suppose that a system consisting exclusively of artificial neurons having the same functional properties as neurons and the same relational properties as your neurons would have different psychological properties from you.[28] But, artificial neurons can be built in a variety of ways given that the relevant function of a neuron is merely to be on/off-signaling devices. So, anything that can function as an on/off-signaling device can be used as an artificial neuron. A person can function as an

[26] I think that you can get away without buying this supposition provided that you're willing to concede a supervenience relation between the mental and the physical (which even Descartes was willing to do). There are, of course, deep questions here that would turn on a particular sort of reading of Descartes dualism—and answering these questions would lead us quite far a field of my main line of argument. However, this is Brooks' (1986) argument and he presupposes materialism.

[27] I have serious reservations about the truth of this claim. To begin with, there is reason to think, with Burge (1979 and 1986), that the content of mental states depends not just on these molecular identities but also on facts about the social histories in which the relevant concepts were learned. Moreover, if the extended mind thesis pushed by Clark (1997), Clark and Chalmers (1998), Clark and Wilson (forthcoming), and Wilson (1995, 2004) is correct, then the environment (both physical and social) of an entity will also be relevant to the identity of mental states because according to this view mental states are not to be understood just as states of brains but instead as the computational states that facilitate various sorts of actions by a system—some of which may be extended or embedded in various ways that are not wholly dependent on the neural state of the system. I'll return to these worries throughout the thesis.

[28] This argument, of course, comes far too quickly. Although a commitment to functionalism does entail the truth of the proposition that a system whose functional characteristics were identical to yours would be psychologically identical to you, Brooks claim only follows if you suppose that there would be no difference in the functional characteristics of a system that consisted exclusively of artificial neurons. This, of course, has not been established—and Brooks does not attempt to establish it.

on/off-signaling device, so a person can be an artificial neuron. So, there is no reason to suppose that a system that consisted of a number of persons arranged so as to replicate the relational and functional architecture of the neurons in your brain would have different psychological properties from the individual that such a system replicates.

As Brooks (1986, 457ff) puts the point, if we were to collect everyone in the nation of China and create a Brain City that replicated the functional architecture of your brain, there is no reason to suppose that Brain City would lack any of the psychological properties that you possess. In fact, Brain City could even have a 'drink' that had the same effect of it that alcohol has on us provided that we could introduce runners into the system that would "dash through the appropriate parts of the city doing the analogue of whatever it is that alcohol molecules do, damping down neurone response levels or changing the signal relations or whatever" (Brooks 1986, 457). The thought is spelled out simply as follows: If functionalism is the correct account of the mind, then it is not the biological properties of neurons that allows them to realize mentation, rather it is their functional properties; and if a group of people can have the same functional properties as a person then they can realize a group mind that has all of the same psychological properties as that person does.

Sure enough, this is one way to defend collective mentality. However, there are other ways of spelling out functional analysis of mentality that provide equally plausible accounts of the mentality of collectivities—indeed accounts that will diverge in interesting ways from this model. While we might choose, with Brooks, to look to the implementation level for an account of the relation between psychological states and their realizers (In fact, this is the way that much of the literature on collective intentionality has gone),[29] there is reason to

[29] I leave this literature to the side; however, I return to it briefly below.

think that there might be some systems were what we should pay attention to the *functional organization* of a system at the level of computational structures and the algorithms that they carry out. While looking to the implementation level might suggest some places where collective mentality is possible, in order to get to the most interesting cases of collective mentality in the actual world it will help to look at a variety of other levels of explanation. In order to get a handle on how this might be possible, it will help to look to a distinction between functional *roles* and the *occupants* of those roles, between programs and realizers, or between software and hardware. Now, we might rest content with a picture of the mind that operates just in terms of two-level explanations, looking to see which sorts of physical structures underlie which sorts of psychological functions and then trying to find analogous structures to see if a system can have mental states. The worry, however, is that this picture fails to capture all of the relevant similarities between mental states.[30]

Consider what it takes for something to count as the program that we know as Mozilla Firefox. To begin with, we can find a LINUX version of Firefox, a MAC OS X version, a Windows version, etc., and each of these programs counts as Firefox in virtue of the sorts of computations and algorithms that the program is running. Even though the program is running on different platforms, we identify the program in virtue of what the system can do. Now, a functionalist might assume that there is exactly one way to make sense of the relation between the mental and the physical. The functionalist might claim that the only relevant analogies between cognitive systems are at the implementation level—but doing so is surely a mistake (cf., Lycan 1987). Brooks supposes that *the best way* to defend

[30] I do not, here, mean to claim that Brooks (1986) is committed to two-levelism. What I intend to suggest here is that there is no reason to think that this is the only way in which collective mentality can be realized. Numerous attempts to explicate the relations between collective intentionality and individual psychological states have focused on this sort of explanation—I find none of these stories compelling. I turn to some of the problems with this approach in subsequent chapters.

the hypothesis of collective mentality is by showing that the functional role of belief, for example could be realized in terms of artificial neurons. However, this is not the only level of explanation to which we might look if we are to develop a story of collective mentality. In fact, for many of the collectivities to which we might be willing to attribute mentality, there is a far more promising story to be told. As Bill Lycan has correctly pointed out, it only makes sense to say that something is a role as opposed to an occupant *modulo* a particular level of explanation. Thus, where an individual person is concerned, it might make sense to say that a language of thought is the software and the wetware is the hardware, there might be also be reason to say of a corporation that the marketing plan is the software and the individuals in various departments at the corporation are the hardware. All of this will depend on what sort of explanation we are looking for. Let me, then, turn to an elaboration of the claim that there might be another way to look at the realization relationship in collectivities.

In the remainder of this thesis, I adopt the sort of position that Lycan (1987) refers to as homuncular functionalism, or homunctionalism. The key claim of homunctionalism is best put in terms of Marvin Minsky's (1988) notion of a society of the mind. The thought is that the functional architecture of the mind is best understood as having a similar structure to a corporate hierarchy. There are various different divisions of the mind, each of which is dedicated to a particular sort of computational task. Consider the visual representation of a beer bottle being thrown at your face as you watch Patti Smith play the last show ever at CBGBs. In this case, there will be a system dedicated to detecting motion, a system dedicated to constructing representations of objects (presumably out of the representations of edges and colors that have been constructed by simpler systems), and a system connecting these

perceptual representations to an action system, and probably a whole host of other systems beyond these. On the homunctionalist theory of mind, the way to explain the functioning of each of these homunculi is by an appeal to the more specialized homunculi that constitute it, and by detailing the behaviors of these homunculi to explain how they produce a corporate output rather than going all the way down to the level of neurology. Lycan (1987) contends that mental states are type-identical with the property of having such-and-such an institutionally characterized state obtaining in one (or more) of one's appropriate homunctional departments or sub-agencies.[31]

Now, suppose that we find cases where we are willing to ascribe mentality to collectivities from the standpoint of commonsense psychology. Further, suppose that we develop an account of what it is to have a particular sort of mental state in terms of the functional characteristics of such a state. At this point, the important questions focus on the how this function is actually realized in the system. We have garnered some tools at this point for thinking about these functions in terms of the sort of information that they are able to process. What we will need to look for is a functional decomposition of the particular sort of mental state that we are attributing to the system. What we will do at this point is try to spell out a sort of boxology, or corporate hierarchy for the relevant sort of belief.

Assuming that it is possible to develop such a picture, we will then look not to the level of physical mechanisms, but instead to the computational processes that need to take place in order for a particular sort of mental state to be possible. This story will be told in terms of the passing of information from homunculus to homunculus. Provided that there are the right sort of homunculi in a system and provided that these homunculi are able to pass the

[31] Lycan spells this out in the case of pain: To be in pain of type T is for one's sub…sub-personal Φ-er to be in a characteristic state St(Φ), or for a characteristic activity At(Φ) to be going on in one's Φ-er

right sort of information to one another, this will yield a system that can possess genuinely mental state. At this point it will be a further question what the realizers are for the function of each of the relevant homunculi. This question will either be answered in terms of a further homuncular decomposition, or at some point just a story about the brute physical primitives.

This is the important point: for something to count as a legitimate psychological process, all that needs to be the case is that the process is realized on some sort of *representational mechanism* that is running the right sorts of *computations*. For collective mentality to be possible, all that will need to be the case is that there is an isomorphism between the computational processes that occur in individuals when we attribute a psychological state ψ to them and the computational processes that occur in collectivities when we attribute ψ to them. Here also lies the key to the story about the autonomy of psychological explanation from neurology. Neurological explanations are irrelevant to psychological explanations *except* (and this is a very important consideration) in so far as the neurological explanations suggest that the psychological story can't be realized in the system in question. So long as the homunctionalist decomposition comes out right, psychology should be happy to concede mentality regardless of how the computations happen to be realized. Thus, we should concede that a collectivity has mental states to the extent that it exhibits the right sort of homuncular decomposition, regardless of the sorts of states that are possessed by the individuals that compose the collectivity.

If we have reason to suppose that the right sorts of computations are being carried out by some subsystem (or sub-subsystem) of a collectivity, then we have good reason to say that our ascriptions of collective mentality are warranted. Alternatively, if the system is behaviorally identical to a system that has a particular mental state but there are no such

computations going on, then we have no reason to think that the system in question is a cognitive system, and thus we have reason to suppose that our attributions of collective mentality are not warranted. It at least seems possible that there could be a collectivity that engaged in computations that were functionally equivalent to the computations that individuals engage in. Thus, it seems like collective mentality is possible.

1. 4. Concluding remarks:

There are three points that I want to make clear about this preliminary sketch of the view I wish to defending. First, what I am concerned about is the functional architecture of cognitive systems. While there are a number of views that have been defended in recent years about the nature of collective intentionality (e.g., John Searle 1990a, 1995; Raimo Tuomella 1992; Margaret Gilbert 1987, 1989), these views have typically been concerned with explanations at the level of the implementation of collective intentions by individual psychological states of various sorts. I do not want to argue that such explanations are unimportant. In fact, I am inclined to think that it will, at some point, be quite important to determine how it is that the individuals who compose a collectivity can implement collective intentions. This is true in just the same way that it will be important to know how the brain realizes mental states in individuals. However, we do not, at least for the purposes of doing psychology, need to be committed to any particular story about the implementation of collective mental states.

Second, we need to be careful with the sorts of states with which we concern ourselves in the analysis of collective mentality. If we start with the sorts of cases that are typical of the collective intentions literature, we find ourselves entangled in a number of

debates that are better left to the side. Margaret Gilbert, for example, often starts with cases like *deciding to take a walk with someone*. However, if we start with these sorts of cases, we will focus on facts about the individuals and what it takes for them to intend to walk together rather than facts about the functional organization of the system in question. This leads to a number of serious worries about how it is that you can get from facts about the psychological properties of aggregations of individuals to facts about genuine collective states. However, if we start with more complicated cases where the corporate architecture is more salient, the functionalist picture that I want to defend becomes much clearer—we see how the psychological properties of the group are actually isomorphic in the right way to the psychological facts about individuals. This being done, it is easier to see that the actual psychological states of individuals aren't really what are at issue in defending collective mentality. My contention is that we need to have the right sort of functionally characterized system in any case where we find collective mentality. The problem with the previous literature on the topic is that it's just harder to see this sort of organization in simpler systems.

Finally, I've argued mental states form a natural kind from the standpoint of psychology and cognitive science. We should not expect there to be one account of individual mental states and another completely distinct account of collective psychological states—at least not so long as we are attempting to develop a theory of *collective psychology*. Once we know what it is to have a mental state, we can apply this theory either to the individual or to the group and the explanation will be the same sort of explanation in both cases. In the same way that we don't need a story about the properties of individual neurons to explain the psychological properties of an individual, we don't need to look to facts about

the individuals that compose a group (e.g., whether we have 'we-intentions') to see if there are collective intentions. We start from the psychological posits and then we check to see if the system in question has the right sort of functional organization to give rise to the mental state in question.

In the remainder of this thesis, I argue that there are reasons not just for thinking that such an account of collective psychology is not merely a metaphysical possibility, but that there are real world cases of collective cognition. But, that's a long and arduous road that must take us through the snags and snarls of a number of serious objections to the possibility of such systems in the actual world.

Chapter II:

THE COLLECTIVE CONSCIOUS?

It used to be taken as obvious that there were cognitive social phenomena that needed to be accounted for by any science of the mind. At the end of the nineteenth century, numerous accounts of the mentality of crowds, for example, appeared in the foundational documents of social psychology; Emile Durkheim argued for the possibility of collective representations as a way to make sense of census data and suicide rates; and, a number of biologists argued that we should understand social insects as superorganisms. However, such appeals are rarely offered these days; when they are, things don't go too well. Even if our best theory of mental states allows for collective mentality, the argument in the previous chapter is bound to leave philosophers and non-philosophers alike unpersuaded. I have argued that accepting functionalism about mental states entails that there are no *a priori* reasons to rule out the possibility of collective mental states. However, a pernicious argument waits in the wings, intent on undercutting any position that purports to demonstrate the possibility of collective mentality.

Although there is a substantial consensus regarding the functionalist view of the mind in philosophy and cognitive science, there are detractors. When faced with the possibility of collective mentality, philosophers, and non-philosophers tend to turn to worries about the impossibility of collective consciousness. They claim that since there is nothing that it's like *to be a collectivity* there must be something wrong with this project of justifying collective

mentality. The argument, in brief, runs as follows. Functional and representational capacities are not sufficient to distinguish genuinely cognitive systems from computational systems that are not genuinely cognitive. What's left out in such explanations is the fact that cognitive systems are conscious and merely computational systems are not, and consciousness is a necessary condition on mentality. So, since collectivities can't be conscious, they can't be cognitive systems.

There are two ways in which this argument can be answered. First, it is possible to resist the argument by demonstrating that consciousness is not a necessary condition on mentality. If it is possible for a system to have mental states without having conscious states, then the impossibility of collective consciousness will not impugn collective mentality. Second, it is possible to argue that our most plausible account of consciousness strongly suggests that a collectivity could be conscious. Most of my arguments in this chapter are directed at establishing the former claim. However, I also claim that collective consciousness is possible, even if highly unlikely in the actual world. However, before developing these claims, I turn to the ways in which the argument from the absences of collective consciousness might be elaborated. In this chapter, I argue that none of these arguments impugn the possibility of collective mentality.

2.1 There are things collectivities can't do:

The first worry that might be suggested concerns the possibility of limitations on the sorts of states that collectivities might be able to exhibit. While people possess myriad mental capacities, it's difficult to imagine a collectivity that could possess all of these. However, the thought goes, if a collection of humans is to count as minded, it must be capable of having at

least the same sorts of mental states that we find in the human beings that compose the collectivity; otherwise there would be no reason to suppose additional mental states beyond those possessed by the individuals that compose the collectivity. So, if there are mental states that collectivities can't possess but that their constituents can, the collectivities must not have minds. But there are clearly states that can be possessed by the individuals that compose a collectivity but not by the collectivity itself. For example, every member of a human collectivity could have the capacity to enjoy a sunset; however, it's difficult to imagine what it could mean to say that a collectivity has the capacity to enjoy that same sunset without appealing to the phenomenal states of the individuals who enjoy it. A proponent of this objection would claim that any system to which we ought to be willing to concede mentality will have to be able to experience such qualitative states as enjoying the sunset. However, if every qualitative experience of enjoyment is localized in an individual, and if ascriptions of enjoyment to collectivities should always be read distributively, then collectivities aren't the sort of thing to which we ought to be willing to concede mentality.

This objection is misguided on a number of levels. First, there are very few sorts of organisms that take enjoyment in watching the sunset. It may just be humans that watch sunsets for the purposes of enjoyment; and if it's not just humans, it's probably just people and *some* tribes of bonobos. So, if *particular* qualitative experiences are a necessary condition on mentality, then humans, and perhaps *some* bonobos, are the only *cognitive* systems that we know of—and this is such a bizarre claim that it's not worth adopting. This point is, of course, rhetorical. However, the rhetoric generalizes in an incredibly robust way. We know that there are a number of mental states that can occur in human systems that cannot occur in some simple systems. Humans have the capacity for normative reflection;

scorpions and badgers probably don't. When I get a phone call from a friend inviting me over for dinner, I can decide whether I want to go or not; but when a scorpion detects vibrations in the air with its trichobothria and in the sand with its pedipalps it has no choice, it just moves toward the prey (and sometimes to its death). The scorpion is immediately pulled toward the food by the vibration sensations; it doesn't decide to act. In fact, it's probably true that all scorpion activity is driven by pushmi-pullyu representations—they get around in the world without reflection, decision or higher-order cognitive, states. However, it seems quite reasonable to me to say that such states are mental states.

Moreover, there are a number of capacities that are not possessed by all members of *Homo sapiens sapiens*. We know of lots of mental disorders that make it impossible for a person to be in certain sorts of mental states. Some people are achromats, seeing the world only in black and white. Others are autistic and don't have the capacity to attribute complex mental states to others that are different from their own. The cases go on and on. Now, it would seem crazy to rule out all of these organisms as cognitive systems. So, to put the response briefly, we ought to recognize that mentality is not an all or nothing affair.

Perhaps collectivities will lack qualia. However, even this is an open question that will turn on what our best theory of what qualia are in humans. While it's probably true that collectivities will only be able to possess the sorts of mental states that are exhaustively explained in terms of their computational structure, whatever those happen to be, I'm inclined to think that a representational theory of qualia is probably the right sort of view, I don't want to come down on that issue at this time. Instead, I'll just say that we need to be careful to specify precisely what claim we're making when we say that collective mentality is possible. We need to be precise about what sorts of mental states we mean to be talking about

and we need to be careful to specify the computational structure that underwrites this attribution of a mental state (or mental states). Thus, I begin with the reasonably untendentious assumption that there are no collectivities around these parts that exhibit conscious states. However, without further argument, this does not, by itself, offer a good reason to deny mentality to collectivities.

This first argument can, thus, be set aside, and I suggest that we follow Rob Wilson (2004) in noting that in order for collective mentality to be possible, it will only have to be possible that there could be a collectivity that had at least one sort of psychological state. Since there is a spectrum of mentality running from systems that possess a wide range of mental states and systems that possess relatively few, we shouldn't begin by asking whether there are any group minds, but we should, instead focus on the question: can any collectivity have beliefs (or desires, or memories, or perceptions, or emotions)? There are, however, other objections lurking in the area.

2.2. Collectivities can't be self-conscious

A more substantial argument against the possibility of collective mentality based on the lack of conscious states rests on a purported connection between the capacity for conscious thought and the capacity for thoughts that you are conscious of as being *your own* (cf., Rosenthal, 1986). There are, of course, many ways to spell out this connection. The most promising of which seems to be grounded on the claim, possibly Kantian in origin,[32] that only a self-conscious organism could have the sort of conceptual representations that constitute thought. Building on this intuition, Jay Rosenberg (1986, 10) has argued that the

[32] David Landy (unpublished manuscript) has argued that the Kant's "Transcendental deduction" in the *Critique of pure reason* contains an argument that requires self-consciousness for the acquisition of object concepts.

"conceptual representation of an objective world is possible only for *self-conscious* subjects". But this position leaves us with a rather significant worry. After all, it seems a bit strange to think that a collectivity could be the subject of it's own thoughts. But if it is true that collectivities can't be self-conscious, they won't be able to engage in conceptual representation. And if collectivities can't engage in conceptual representation, then the range of possible representations that are attributable to collectivities might be too narrow to be psychologically interesting.[33]

Fortunately, there is a fairly straightforward response to this worry. To begin, we need to be careful to specify what sort of self-consciousness is necessary for conceptual thinking. Rosenberg, for example, claims that having a representation of a self is a necessary condition on the possession of object concepts because object concepts require a subjective ordering of representations that picks out a series of impressions as constituting a representation of a single thing. This does require a representation of a self, but it's a relatively thin representation—what Kant might call a transcendental unity of apperception. To put the point briefly, an 'I think' representation must accompany every representation of an object. But, there is no reason to assume that it would be impossible for a collectivity to be structured in such a way that it could have such a self-concept. Let me sketch briefly what such a collective self could be.

First, there is a thin sense of 'self' under which any system that is in the business of self-preservation has to be able to distinguish itself from other systems. As Dennett (1989,

[33] Robert Rupert (2005, n4) offers this worry as a possible objection to collective mentality. He argues as follows: "It is often thought that a mental representation of the self plays a special role in the life of a mind, particularly in self-consciousness. Admittedly, many group systems have names, written on letterhead or painted on signs. It is another matter, though, to show that such representations play the role of a concept of one's self." Rupert does not develop this objection, however, because he thinks that other worries are far more pressing for accounts of collective mentality.

1991) puts the point, nobody can preserve the whole world, so even incredibly simple systems have to distinguish self from other. Some collectivities are clearly in the business of self-preservation, so will clearly have at least this sort of minimal self.[34] Unfortunately, this minimal concept of the self won't be enough to answer the objection. After all, merely being able to preserve one's self doesn't require any conception of one's self at all. However, thinking in terms of what a collectivity must do in the service of self-preservation points the way to a thicker conception of the self that can be possessed by a collectivity. The thing to note is that many collectivities monitor their own behavior and modify it in light of current circumstances—providing conditions under which a collective self-concept may be manifested.

To put the point another way, collectivities are not one and all Darwinian systems (Dennett 1996, 84-85) whose behavior is unreflective and static. Nor are they all Skinnerian systems (Dennett 1996, 85-88) that modify their behavior in response to stimuli by way of some sort of dumb feedback mechanism operating in accordance with Thorndike's Law of Effect. In fact, some collectivities even appear to be able to decouple indicative from imperative representations (cf., Millikan 1984, 1989, 1996) in a way that marks them off as Popperian systems (Dennett 1996, 88-93) that are able to allow their hypotheses to die in their stead by preselecting behaviors on the basis of internal models. Some collectivities might even reach the level of Gregorian systems (Dennett 1996, 99-101) that are able to engage in meta-representation to the extent that they can genuinely ask if they are correctly modeling the world in a way that produces the optimal response to the circumstances at hand.

[34] Remember, even an anarchist collective that the CIA is attempting to dismantle has to try to preserve itself to some extent if it is to be stable enough to make it to Heiligendamm to protest the next G8 summit. Even clearer cases of self-preservation come in the cases of large multinational corporations (e.g., Microsoft), Universities, cultures, and the like.

In order to make this claim clear, consider what happens in the case of an E. Coli outbreak linked to the monster-burgers at Burgerzilla.

In the face of such an E. Coli outbreak, Burgerzilla must defend its interests, but it doesn't want to defend the meat at every burger joint in town. So Burgerzilla has to respond in a way that is driven by its own interests in self-preservation (suggesting that Burgerzilla has at least a minimal self). In responding to the outbreak, Burgerzilla will attend to the fluctuation of its profits (ignoring for all intents and purposes the profit margin at BurgerTown, for example) in response to the E. Coli outbreak. Burgerzilla will also pay close attention to where the E. Coli infected meat came from as well as what it will take to ensure that there won't be any more infected meat sold at Burgerzilla. But most importantly, Burgerzilla has to respond to this crisis in a way that will facilitate a continued presence in the monster-burger market, there will have to be a public response demonstrating that Burgerzilla is committed to preventing such an outbreak in the future.[35]

In responding to the crisis, suppose some department or division of Burgerzilla, Inc. monitors both public opinion and the internal organization of the corporation—call it Public Relations.[36] When faced with the E. Coli outbreak, the PR department would see that public opinion about Burgerzilla is declining, but it might also see other trends. The PR department might also come to realize, when it collects data on the public perception of Burgerzilla via phone and internet surveys (in true Gregorian fashion), that for a long time Burgerzilla has

[35] Note, however, that it won't be enough to show that the individual members of Burgerzilla's board of directors are committed to preventing E. Coli outbreaks. After all, boards of directors change often: some people die, other people leave Burgerzilla to go to BurgerTown (or Halliburton), new members are added, etc. However, Burgerzilla must demonstrate that *it* is committed to preserving the health of its clientele if it is to preserve its profit margin.

[36] Note, this need not be the only task that this division or department undertakes. Although the clearest case in the vicinity is one in which there is such a distinct department, a case in which the owner or CEO of Burgerzilla Inc. takes this to be her job, or some other case is equally plausible and equally a case in which the relevant sort of self-concept will be possessed by the collectivity.

been seen as a 'dive' burger joint. In response to this perception, the PR department at Burgerzilla might attempt to create a new public view of what it means to get a burger from Burgerzilla. But in order to do this Burgerzilla will have to be able to develop a model of how it is to govern its behavior as well as how it is likely to be perceived by others. This, then, would require both a model of how Burgerzilla is to understand itself as well as a meta-representational model of how it is to be understood by others.

In attempting to demonstrate its commitment to the health of its customers, suppose that the PR department of Burgerzilla writes press releases, recommend modifications of Burgerzilla's mission statement, and in general engages in and overall restructuring of Burgerzilla. Burgerzilla might even place a full-page add in the New York Times stating that they will now be using only the finest Kobe beef. Burgerzilla might change the appearance of its restaurants from the drab purple and brown interiors that have always been the hallmark of a Burgerzilla, to a sleek, bright, red and yellow with fancy new menus designed by the finest graphic designers that money could buy. And most importantly, Burgerzilla might make a *self-conscious decision* to be perceived as the safest burger-joint in the world—even offering a new tag-line on their commercials: "we've gone from the last place you'd wanna buy a burger to the first place you'd think to buy a burger!"

All of these things seem like reasonable things to expect from Burgerzilla. However, all of these moves require an incredible amount of internal monitoring that yields much more than a minimal self. Burgerzilla has to represent itself in various ways, recognize that Burgerzilla exists above and beyond the members of Burgerzilla, and it has to act in such a way that its actions will be seen as the actions of Burgerzilla. This is at least enough to yield an apperceptive self-representation for a collectivity, and in order to meet Rosenberg's

version of this objection that's all that's required. There is more to say about the nature of collective selves; however, this thesis is not the place to address such issues.

2.3 Is phenomenal consciousness necessary for mentality?

This brings me to the most difficult argument to address, as well as the argument most likely to occur to the philosopher. The argument rests on a worry generated by one of Ned Block's familiar thought experiments. Block (1978) asks us to imagine a case where that the nation of China is forced by a ruthless leader—a 'true believer' who has recently converted to functionalism—to implement a person's functional architecture just for an hour.[37] In a massive philosophical undertaking, every person in China is given a two-way radio that is connected to some other radios and to a body that looks (from the outside) just like a human body. Each person is then asked to carry out a relatively simple task. For example, a person might be told that if she sees a Φ projected on an overhead screen attached to a satellite, she should send radio signal ψ. If all goes well, the two-way radios will be wired in a way allows the nation of China to be in the same functional state as some person. Now, if functionalism were a true and complete theory of the mind, such a system would implement a person's functional architecture and would thereby have mental states. However, while the functionalist would take a properly organized group to possess mental states *in just the same way as an individual does*, Block thinks that such a homunculi-headed

[37] After all people do have a tendency to get bored quite easily and we can't expect the nation of China, even under such a ruthless leader, to stay focused for any longer than maybe an hour.

system is probably *not* the sort of thing that we should be willing to say is capable of mentation.[38]

But is Block right? Should we be unwilling to attribute mentality to such a homunculi-headed system?[39] Block argues that homunculi-headed systems are missing an important class of states that should be possessed by a system capable of mentation (i.e., qualitative states, raw feels, etc). Borrowing a phrase popularized by Nagel (1974),[40] there's nothing it's like to be the nation of China—and Block claims that if there is nothing that it's like to be a system, then the system probably isn't capable of mentation. In the absence of qualitative states, Block contends that this example of the nation of China ought to be seen as providing a *prima facie* doubt about the plausibility of functionalism. Block does not, of course, purport to defend this intuition that a group could not have phenomenal mental states; rather, the *prima facie* implausibility of qualitative states in such a system is supposed to do all of the work.

I'm not sold here, and I don't think anyone else should be either! As I see it, Block's nation of China example can be read in at least two ways. The first, which seems most

[38] It is important to keep in mind here that Block is just remaining true to one of the fundamental commitments of functionalism: what goes for one system goes for any system that is functionally identical to that system. So, if it turns out that there is one system that is functionally identical to a second system, but the first system possess a property that is not possessed by the second, then said property will not be exhaustively explained in terms of the functional states of the system.

[39] Note that this sort of argument could just as easily be developed as an argument against the possibility of collective mentality in general. Given that there is a rather dominant intuition against the possibility of consciousness in any collectivity (cf., Knobe and Prinz, forthcoming), it could just as easily be the case that Block's intuition could be marshaled in an attack on the possibility of collective mentality simpliciter. If his argument works at all, it is meant to show that the mere fact that a collectivity is functionally equivalent to an individual is insufficient to establish its capacity for mentality. So, while I address Block's version of the argument, it is meant to generalize to any claim about collective mentality whatsoever.

[40] To the best of my knowledge, the phrase 'what it's like' as it is used by Nagel is first adopted by the psychologist B.A. Farrell (1950) in a paper called "Experience". Unlike Nagel, Farrell considers the case of 'what it's like to be a bat' in order to motivate a form of eliminativism about sensations and other qualitative states given the impossibility of a third-person, public criteria for the truth of claims about what it's like.

consistent with the rest of Block's work, is to take the nation of China case as suggesting strong *prima facie* reasons against the ascription of *qualitative* mental states based on functional organization. On this reading, Block doesn't intend to make any argument *whatsoever* against purely representational states in homunculi-headed systems. In fact, at points Block (1978, 306) does claim that homunculi-headed systems would at least have some mental states:

> Propositional attitudes are an example. Perhaps psychological theory will identify remembering that P with having 'stored' a sentence-like object that expresses the proposition that P (Fodor 1975). Then if one of the little men has put a certain sentence-like object in "storage," we may have reason for regarding the system as remembering that P.

Unfortunately, reading Block's argument in this way reduces his claim to an intuition about the possibility of a collectivity possessing qualitative states—though it is a perfectly common intuition (cf., Knobe and Prinz, forthcoming; Huebner, Sarkissian and Bruno, under review). However, it is not clear that this is an intuition that our best philosophical and psychological theory of the mind should lead us to retain.

To begin with, imagine what would happen if a team of cognitive scientists decided to run a series of experiments on the nation of China system (while it was implementing the functional architecture of a single individual). Perhaps they would run some tests to see whether or not the nation of China were capable of having the qualitative experience of hearing unresolved dissonance. In order to test this, the team of cognitive scientists would broadcast a series of consonant chord progressions to the nation of China and a series of chord progressions containing unresolved dissonance. Suppose that the homunculi-headed system engages in reports of having a particular qualitative experience—suppose that upon probing, the system reports feeling uncomfortable when presented with unresolved

dissonance and the system reports that the consonant chord progressions sound pleasant and appealing. Suppose, moreover, that the nation of China engages in precisely the sort of behavior the cognitive scientists would expect from a system that was experiencing unresolved dissonance for the first time (e.g., the reporting system—whatever it may be—outputs the request not to be presented with such stimuli again, without provocation system outputs a yuck response, etc). It seems that in this case we have the best evidence that we could possibly have for the claim that this system is having that sort of qualitative experience.[41]

The point is this. Even when we try to figure out whether another person is having a particular qualitative state, the best that we can do is to pay attention to her overt behavior. But given that we've got a team of cognitive scientists involved, we can do even better than that. Suppose that while broadcasting the chord progressions, various subsystems of the nation of China could be monitored for activity (for example, by paying attention to the expenditure of energy in areas of the country that have been assigned particular roles) to see if there were differences in the processing relevant to the parts of the nation of China that have been dedicated to attention and the processing of emotional stimuli (a sort of macro-scale EEG). Suppose that the team of cognitive scientists finds that there is an increase in activity in the areas that are intended to process acoustical stimuli and associated affective

[41] This move is a bit too quick. After all, as Block (1981) argues, knowledge of the mechanisms that give rise to a particular sort of behavior are relevant to distinguishing systems that are genuinely cognitive from those that are behaviorally indistinguishable from, but not actually cognitive systems. Block contends that unless a systems processes information in a similar way to paradigmatic cognitive systems (i.e., humans)—from the standpoint of cognitive psychology—we will have reason to suppose that the system in question is not a cognitive system but merely behaviorally identical to a cognitive system. The realization of mental states does, however, allow for variance in etiology provided that information is processed in relevantly similar ways for various cognitive systems. I am inclined to think that certain sorts of collectivities do process information in a way that is captured by the kinds laid out by cognitive psychology. However, much of this argument will have to wait until later Chapters. For now, I just ask that you assume that the sort of information processing that occurs in an individual and in this particular collectivity is functionally equivalent.

responses. This would then suggest that the covert behavior of the system was functionally equivalent to the covert behavior that we find in the processing of such stimuli for humans. If all of this were to be the case, as it would in Block's case *ex hypothesi*, we would have evidence from overt and covert behavior that the system was in fact having the qualitative experience. And, as Lycan (1987, 27) puts the point "in the presence of such behavior, a skeptic would have to come up with substantial defeating evidence in order to overrule the presumption of genuine…qualitative states."

It might, however, turn out that there are no actual collectivities that are capable of experiencing such qualitative states. Qualitative states might presuppose a sort of unity that cannot be possessed by collectivities in the actual world. However, even if it is impossible to attribute qualitative states to other collectivities, this won't, by itself, be a problem for the hypothesis that functionalism allows for the possibility of collective mental states. Of course, we would have to be careful in ascribing various states to collectivities, making sure not to ascribe states with phenomenal content; however, the lack of some mental states needn't worry us at all about the possibility of collective mental states in general.

There is, however, another strain in Block's argument that proves far more troublesome for the proponent of collective mentality. At some points, Block makes the stronger claim that "there is a prima facie doubt whether [the nation of China] has *any mental states at all*" (1978, 278 *emphasis mine*). While Block would not typically claim that qualitative consciousness is a necessary condition on mentality,[42] there is good reason to

[42] In fact, Block (cf., 2003) typically argues that qualitative states are a sort of mental paint (sensory qualia that are vehicles of mental representation) or mental oil (sensory qualia that are not vehicles of mental representation) that can be distinguished from the *merely* representational states of an organism. Block typically argues for no more than the claim that a computational *cum* functional view of the mind is *not sufficient* to make sense of all of human mental life. Instead, something more has to be added to make sense of our intuitions in absent qualia cases (Block 1978, 1980) and inverted qualia cases (Block 1990).

think that only a stronger sort of intuition could underwrite an argument against the mentality of groups. To put the point succinctly, the fact that a system lacks qualitative consciousness does not imply anything about the systems capacity for mentation *unless* there is some reason to think that qualitative consciousness is a necessary condition on mentality *simpliciter*.[43]

This understanding of Block's position is made more plausible by the fact that during this period Block found many of the intuitions that underwrite Searle's Chinese room thought experiment quite palatable. For example, Block (1980a) offers a thought experiment in which a person and her homunculi-headed doppelganger are asked a series of questions via a two-way television system in an experimental paradigm styled after the Turing test (Turing 1950). Block supposes that the person and her doppelganger will, given that they are functionally equivalent, respond to interrogation in a perfectly indistinguishable manner. Contra functionalism, however, Block (1980a, 261) notes that although the person would understand the interrogator's questions and reply to them in a way that would express her own thoughts, we cannot say the same of her homunculi-headed doppelganger. The absence of understanding in this case is supposed, then, to demonstrate that the "homunculus-headed system seems as lacking in thought as in qualia, and so any argument against functionalism based on such an example could as well be couched in terms of absent thought as well as absent qualia" (Block 1980, 261). These claims are further developed against the nation of China case in "Troubles with functionalism" where Block claims that:

[43] Block (personal communication) has noted the sort of view that is pressed in his papers on homunculi-headed systems is not indicative of his considered view on the subject. Block believes, and has believed for a long time that functionalism is true as an account of *most* mental states. His considered position is that functionalism fails in the case of qualia—for qualia, however, an identity theory is needed. Block does acknowledge, however, that his position may have wavered in the 1970s when his papers on absent qualia were written.

> There is a prima facie doubt whether it [the nation of China] has *any mental states at all* (1978, 278 *emphasis mine*).
>
> The Absent Qualia Argument rested on an appeal to the intuition that the homunculi-headed simulations *lacked mentality*, or at least qualia (1978, 281, *emphasis mine*).

Keeping these claims in mind, Block might make the claim that homunculi-headed systems *lack mentality simpliciter*. However, the only reason that Block has given for this claim is that it seems *prima facie* implausible that the nation of China could have qualitative states. If this is to be, as it is billed, a demonstration that functionalism is too liberal (i.e., that it attributes mentality to too many entities), then the lack of qualitative states will have to be sufficient to demonstrate the absence of mentality *simpliciter*. Unfortunately, all that Block has offered by way of argument on this point is the intuition that homunculi-headed simulations lack qualia and on this basis he has claimed that "there is no independent reason to believe in the mentality of the homunculi-head, and I know of no way of explaining away the absurdity of the conclusion that *it has mentality*." (1978, 282, *emphasis mine*). Now, supposing we concede that there's nothing it's like for the nation of China to be the homunculi-headed system that it is, what would motivate the claim that such a system lack mentality *simpliciter*? The claim here rests on the thought that possessing qualitative states of consciousness is a necessary condition on possessing intentional mental states; and there is a philosophical option in this vicinity that Block, or someone who wanted to push Block's argument further, might be willing to consider.

John Searle (e.g., 1992) defends this stronger position as follows.[44] To begin with, it is perhaps one of the least contentious claims in the philosophy of mind that many mental states are intentional. Intentional states (e.g., my belief that it is too hot outside, my desire to

[44] Thank you to Felipe de Brigard for his assistance in understanding Searle's argument.

drink another cup of coffee, and my hope that the next album I hear at the café alleviates my desire to dig out my eyes with a sugar spoon) are all directed at, or about something. In fact, many philosophers agree that anything that is going to count as a mental state of a system must exhibit intentionality—that is, it must be about something (i.e., the temperature, a cup of coffee, and the next album that they put on at the café).[45] Now, supposing that mental states are intentional, there is a question about what this intentionality amounts to.

Searle (1990, 586) contends that mental states must have intrinsic intentionality rather than as-if intentionality. This means that mental states must have conditions of satisfaction (e.g., truth conditions in the case of beliefs) that are intrinsic to the state rather than relative to an interpreter.[46] This, however, merely raises the problem of what could account for the fact that mental states have this sort of intentionality. Searle (1990, 587) argues that intrinsic intentionality of mental states can only be understood in terms of the *aspectual shape* of the satisfaction conditions of thoughts. Briefly, the intentional content of a thought is always intensional. That is, human thought exhibits what Quine calls opacity, in the sense that thoughts are always entertained under a particular description. Searle argues that opacity results from the fact that every thought is entertained from some perspective and under some aspect (and *eo ipso* not from other perspectives or under other aspects). Searle (1990, 587) contends that this aspectual character of thought cannot be captured by third-person

[45] A bit of a qualification is required here. If there are, conscious states that take the form of what Ned Block (2003) has called 'mental oil', these states would be states of a system that don't have any representational content and thereby aren't intentional. For the purposes of this section, I don't think that much turns on the plausibility of mental oil. What needs to be established is that the nation of China has no mental states *simpliciter*, not merely that it has no purely qualitative states and if it were true that there are purely qualitative states this would not have any bearing at all on the presence, or lack there of, of intentional mental states.

[46] In order to make sense of the notion of interpreter relativity, consider the sort of intentionality that the words on this page have. These words have intentionality in the sense that you or I can interpret them to be about Searle's views on consciousness; however, in the absence of an understanding of the English language and in the absence of and understanding of the ways in these symbols represent words the symbols are utterly meaningless. Although these words can succeed in being about Searle's theory of consciousness, they can do so only as interpreted (cf., Searle 1980, 199).

descriptions but can only be entertained as first-personal states with aspectual character. Finally, on the basis of these considerations, Searle (1990, 588) argues that any mental state must at least be 'in principle accessible' to consciousness because otherwise the state that is being picked out would just be a neurophysiological state that played some role in the production of some behavior and not a truly mental state. Thus, what it means for a belief to be a tacit belief, for example, is for it to be a belief that you can become conscious of having. In the absence of this possibility, Searle holds that the state cannot be a mental state.

Building on this argument, it is possible to develop a stronger version of Block's initial claims about the mentality of the nation of China. Suppose we've conceded for the sake of argument that "there is a *prima facie* doubt whether there is anything which it is like to be [the nation of China]" (Block 1978, 278). If there is nothing that it is like to be the nation of China, then there is nothing that it's like for the nation of China to be conscious of entertaining a particular thought. If there is nothing that it is like for the nation of China to be conscious of entertaining a particular thought, then there is no first personal aspectual character for any state of the nation China. But if there is no first personal aspectual character for any state of the nation China, then no state of the nation of China exhibits opacity. If no state of the nation of China exhibits opacity, then no state of the nation of China are intensional; and without intensionality none of the states of the nation of China are intrinsically intentional. But mental states have to be intrinsically intentional, so the nation of China has no mental states.

What, then, is to be said of this argument? To begin with Block (1980b, 425) argues that "the burden of proof lies with Searle to show that the intuition that the cognitive homunculi head has no intentionality (an intuition that I and many others do not share) is not

due to doctrinal hostility to the symbol-manipulation account of intentionality." Block's considered judgment on the issue is that there is good reason to think that functionalism *does* give us a powerful account of *most* of human mentality (pesky qualia excluded). The position here follows Fodor (1968), Dennett (1978), Cummins (1975), and Lycan (1981), and it is of a piece with the position that I adopted in chapter 1—claiming that some version of homuncular decomposition or functional analysis is the correct methodology for the practice of cognitive science.[47] Building upon this methodological assumption, Block (1995, 418) argues that the best criticism of Searle's Chinese room argument—and presumably of the sort of Searlesque argument that I've just developed—comes in the form of what Searle (1980) refers to as the systems reply to the Chinese room case. This reply holds that although there may be no individual component of the Chinese room that can properly be held to speak Chinese, this all by itself is insufficient to entail much of anything at all about the cognitive system as a whole's mental characteristics. But more must be said about why one should accept the systems reply.

Block defends the systems reply as follows. First, he notes that just as we can't reason from "'Bill has never sold uranium to North Korea' to 'Bill's company has never sold uranium to North Korea'…we cannot reason from 'Bill does not understand Chinese' to 'The system of which Bill is a part does not understand Chinese'" (Block 1995, 418). Second, he argues that there is no reason not to construe the system in question (composed of the person, the translation manuals, and the input/output doors) as a cognitive system—and this is the really important point for us. If functional decomposition offers the correct

[47] Here's how Block (1995) puts the point: "Think of this homunculus as being composed of smaller and stupider homunculi, and each of these being composed of still smaller and still stupider homunculi until you reach a level of completely mechanical homunculi." Faced with a particular ability that a person exhibits, cognitive scientists then attempt to spell out the mechanisms that realize that particular ability by breaking down the task into simpler systems that could do the work.

explanation of any systems cognitive behavior, then *a person's* cognitive behavior will be best explained in terms of subsystems and subroutines jointly capable of producing her behavior, though none of them is capable of doing so individually. Similarly, if functional decomposition offers the correct explanation of any systems cognitive behavior, perhaps *the system that Searle calls the Chinese room* will exhibit cognitive behavior that is best explained in terms of a subsystems and subroutines jointly capable of producing its behavior, though none of them is capable of doing so individually. Similar considerations apply *mutatus mutandus* to the nation of China case that we are here considering.

In concluding his argument against Searle, Block (1995, 420) argues that "to the extent that we think of the English system as implementing a Chinese system, that will be because we find the symbol-manipulation theory of the mind plausible as an empirical theory." At this point, the dispute then becomes not a dispute over intuition, but a dispute over what is the best theory of human mentality from the standpoint of an informed cognitive science. Block contends that the most plausible account of human mentality will be spelled out in terms of some form homuncular functionalism—aside from qualia, which he believes will rest on an identity theory of mind.

However, if the systems reply works as an argument against Searle's Chinese room argument, then it will also work as an argument against the Searlesque reading of Block's nation of China example. The systems argument for collective mentality would hold that there are intentional states exhibited by the collectivity that are distinct from the intentional states of the individuals. Now, if functional decomposition offers the correct explanation of any systems cognitive behavior, then perhaps *the nation of China* will exhibit cognitive behavior that is best explained in terms of a subsystems and subroutines jointly capable of

producing its behavior, though none of them is capable of doing so individually. Block's argument against Searle thus provides reason to think that the nation of China is a cognitive system as well. Since the nation of China could, if properly organized, have some sorts of intentional mental states, then collective mentality can be saved from this worry about consciousness.

This response is not, of course, a knockdown argument against Searle's claim that consciousness is required for mentality *simpliciter*. In fact, Searle will contend that this sort of response misses the point given that if fails to address his concerns about the necessity of every mental state having an aspectual shape and a first person character. This is, after all, the reason that Searle is not going to be willing to accept functionalism and RTM as an adequate theory of the mind. Given his commitment to the first-personal nature of psychology Searle (1990b) thinks that starting from these sorts of functional explanations will prevent us from giving any adequate scientific account of mentality precisely because all mental states are supposed to be conscious (or at least available to consciousness). However, there is more to say in response to Searle's requirement that all mental states must be conscious.

At this point, it will help to get clear on what, precisely, the connection is supposed to be between consciousness and intentionality. After all, to deny that there are unconscious processes in human cognition should strike everyone as wildly implausible. And, in fact, this is not the possibility Searle wants to deny. Searle's argument is meant to demonstrate that any unconscious state that is *genuinely intentional* is the sort of state that about which one *could be conscious* (Searle 1992, 153). The primary claim that Searle (1992, 132) wants to advance is that "only a being that could have conscious intentional states could have

intentional states at all, and every unconscious intentional state is at least potentially conscious." However, there are a number of points here that beg for clarification.

First, it isn't clear what it means for a state to be potentially conscious. There are at least two different ways in which this possibility could be articulated. Metaphysical possibility won't work; in fact it is metaphysically possible that the states of a magnetotactic bacteria's magnetosomes could be conscious (in some possible world), or that I could be conscious of the secretion of hormones in my pituitary gland in some other possible world, and Searle doesn't want *these* states to count as possibly consciousness.[48] Searle's primary concern is with the laws of psychology, so following Uriah Kriegel we would do well to construe the relevant range of possibilities as psychological possibilities, that is, a mental state "M is potentially conscious iff there is a possible world W, such that the laws of psychology in W are the same as in the actual world, and M is conscious in W" (Kreigel 2003, 275). Second, there are at least two ways of reading Searle's claim about the necessity of *consciousness* for intentionality (cf., Block 1990b). Searle can either mean that every genuinely mental state is potentially conscious in the sense that there are worlds consistent with the laws of psychology in which this state is *accessed by reasoning and reporting mechanisms*, or he can mean that there are worlds consistent with the laws of psychology in which there is *something it's like* to be in this state.

Suppose we adopt the first reading of potentially conscious. I've offered a number of arguments both in this chapter, as well as in Chapter 1, for the claim that the laws of psychology do not preclude the possibility of reasoning and reporting mechanisms in a collectivity. It is, of course, an open and empirical question whether there are any

[48] That a state is "metaphysically-possibly conscious appears to be a purely logical property (or perhaps a "metaphysical" property) of it, rather than a genuinely psychological property" (Kriegel 2003, 274).

collectivities that actually have the relevant sorts of mechanisms for accessing their representational states (and indeed whether collectivities can have representational states at all). However, the homunctionalist position that underwrites my argument for the possibility of collective mentality suggests that if individuals have such functionally specified mechanisms for reasoning and reporting, then a properly organized collectivity could just as easily have such mechanisms. So, it would at least be psychologically possible for a collectivity to have conscious states, and there would be no reason to claim that collectivities could not have intentional states even on Searle's picture.

Searle must, then, adopt the '*what it's like*' sense of consciousness. However, If higher-order representationalist accounts of consciousness succeed (which I'm inclined to think that they do), then the 'what it's like' of a particular mental state can be specified in terms of the higher-order monitoring systems used to attend to the internal states of the psychological system. However, there's no reason to suppose that such mechanisms could not be present in a collectivity, and if such mechanisms can be present in a collectivity, the problem of 'what it's like' for a system to be in a particular states reduces to a special case of consciousness as access by reasoning and reporting mechanisms. But if this is the case then there is no reason to claim that collective consciousness is psychologically impossible, in which case there is no reason to claim that collectivities couldn't have intentional states, even on Searle's picture.

Searle won't concede the higher-order representationalist theory of consciousness, and there is, thus, a stronger version of Searle's argument that must be addressed. Searle contends that what distinguishes a genuinely mental state from a derivatively representational

state is that genuinely mental states have an intrinsic aspectual shape.[49] However, since mental states *are nothing but neurophysiological events*, and since there is no intrinsic aspectual shape for merely physical states of a system, Searle contends that the only thing that could make my unconscious belief a belief about Superman and not about Clark Kent is the way in which this belief is understood by a conscious system: were this to be a conscious belief, it would have the aspectual shape, for me, of being about Superman.

According to Searle the psychologically impossibility of collective mentality depends on the intuition that a collective representation cannot have an aspectual shape *for the collectivity*. More importantly, the sorts of collectivities that seem most promising as avenues of inquiry for claims about collective mentality don't seem to be the sorts of things that could have conscious states. These collectivities consist of various people, connected only by informational and causal relations, engaging in a variety of functionally specified tasks that lead to various sorts of collective actions. Searle would contend that each person understands what she is doing as contributing to the goals of the group (in fact, she will likely explain her own actions by saying that 'we are trying to Φ); however, he would also claim that there is no understanding at the level of the collectivity. Searle is thus inclined to claim that the content of any collective representation will be derived from the representations of the individuals that compose the collectivity, and this rules out genuinely mental collective representations. Fortunately, there are deep problems with this Searle's reliance on first-personal states in making sense of the cognitive content of a particular representation.

The first response to this stronger version of the argument is to resist Searle's claim that you need to have first-person consciousness in order to make sense of aspectual shape. In

[49] Searle is not particularly clear about what aspectual shape is: however, it is something like this. Psychologically speaking, my belief that Louis Lane loves Superman is a belief about Louis Lane and Superman, it's not a belief about Louis Lane and Clark Kent, and this is true even if this belief is unconscious.

order for a representation to have an aspectual shape, it must be able to be represented-as something. Thus, we have the capacity to represent a duck-rabbit either as a duck or as a rabbit. We have the capacity to represent Clark Kent either as Clark Kent or as Superman. There is some introspective data that seems to suggest that it is because these things appear to us to be a particular way that they have the aspectual shape that they do, and it is on this basis that Searle claims that aspectual shape is only possible within a conscious system. However, although Searle claims that there is no way for a non-conscious system to be able to represent something as having a particular aspectual shape, it's not at all clear whether this is true. Sure enough, we haven't yet worked out what sorts of computational structures can give rise to a representation with a particular aspectual shape. However, as David Chalmers (1996, 360n10) notes, there is no reason to suppose that the storage of information about a person in a database could occur either under their name or social security number. I don't find this particular claim of Chalmers' compelling. However, The sentiment is quite promising.

Computational theories of vision based on feature detection (e.g. Marr 1992 and Edelman 1999) have retained a great deal of prominence in cognitive science. By studying the ways in which the human visual system responds to various sorts of stimuli, computational neuroscientists have been able to design computational models of vision that succumb to the same sorts of illusions to which persons succumb (Cf., Purves and Lotto 2003 for a review of the evidence). It doesn't seem too unlikely that computational systems could be constructed that would behave exactly like a person does when faced with a duck-rabbit. On the supposition that we could build a computational visual system that could attend to various features of a display (cf., Ullman and Sali 2000 for some promising suggestions here) and then 'saccade' over the image in the right way to produce the switch, it's not at all clear

to me why we should be unwilling to say that this system is not represent the stimuli as having a particular aspectual shape.

Moreover, and more importantly, if there is a naturalistically plausible theory of intentional content, that is spelled out in terms of the relation between the computational states of an organism and particular feature of the world that matter to that system (whether the system be the organism or some subcomponent of the organism), such a model will also provide us with an account of the aspectual shape of mental states without an appeal to consciousness. Searle, of course, denies this possibility. He claims that the only way in which a cognitive system could possibly be appeared to both in a way that respects aspectual shape and that explains how the particular aspectual shape of a representation could matter to that subject is by way of conscious representation. However, it's not clear why he believes this.

Consider a system whose representational states stand in nomic relations to the features that distinguish the various different aspects of representation from one another. Now, suppose that the features of the representations that are represented determine the categorization behavior of a system in a way that either allows the system to succeed or fail in some intentional task that the system happens to care about. Now, regardless of whether the aspect under which a representation occurs is cognitively available to a system, it's behavior will be determined by the aspectual shape of the representation and it's having such a representational state will matter to a system. The important point here is one that is not often noticed by proponents of wide content. In order for it to be true that a system is representing something under a particular aspect, it need not be the case that the system is aware of representing it as such. If it is true that nomic relations between features and states of a system are what is constitutive of content, then whether I take myself to be representing

the watery stuff as water or as twater, as a duck or a rabbit, as superman or Clark Kent, it's the nomic relations that exhaustively determine the content. The thoroughgoing naturalist has to claim that introspective failures are possible, and that the production of a representation, in a way that is important for psychology, is not a matter of what we happen to *think that we are representing*, but *what we actually are representing*. Given that introspection is a process anterior to the process of producing representations, these processes can come apart, although it is rarely the case that they actually do.[50] This is not, of course, to say that I committed to the truth of some naturalistic theory of representation. In fact, it might, in the end, turn out that naturalistic theories of representation are a miserable failure. However, CTM, a theory plausibly grounded on a naturalistic semantic theory seems to be a rather successful theory of the mind for the time being; and in the absence of a adequate successor theory, it seems perfectly reasonable to see how far this research program can be pushed.

Searle, of course would remain unmoved, claiming that the relevant 'sort' of aspectual shape hadn't been captured. However, as Chalmers (1996, 360n10) puts the point, Searle might claim that "the only true aspectual shape is *phenomenal* aspectual shape; but this would seem to trivialize the argument." However, if Searle is going to hold out and claim that this sort of phenomenal aspectual shape is the only thing that can possibly underwrite intentional content, there is another response to his worry. At this point the best way to respond to Searle is by demonstrating the falsity of the claim that the phenomenal states could not be present in collectivities.

[50] Such a distinction between representation and introspective representation might seem strange. However, consider bizarre neurological breakdowns like Anton's syndrome (in which a person is cortically blind but has an introspective representation of sight), and Cotard's delusion (in which a person who is alive has an introspective representation of being dead). Mechanisms of representation and introspection are typically in synch. However, the interface can, and does, break down on very rare occasions. This being the case, the mere fact that introspection does not happen to access the aspectual shape of a representation does not, all by itself, entail that the representation does not have a particular representational content.

For his argument to be conclusive, Searle needs to demonstrate not merely that there is a logically possible world where something that's functionally equivalent to me has no conscious states. He needs to make the stronger claim that all worlds that are consistent with the laws of psychology are worlds in which there is nothing that it's like to be a collectivity. Although thought experiments like Block's 'Nation of China' and 'homunculi headed robot', and Searle's 'Chinese room' might be sufficient to demonstrate that the lack of consciousness *is possible* in a system that is functionally isomorphic to me, they aren't strong enough to demonstrate that no collectivity could possibly be conscious.

Note that the intuition that no collectivity could possibly be conscious is nothing more than an intuition. But it's an intuition that's not even universal. Although Knobe and Prinz (forthcoming) found that Americans are typically uncomfortable with the attribution of conscious states to groups, Huebner et al (under review) found that East Asians were significantly more willing to attribute conscious states to groups. It's true that these data are far from conclusive; however, it does give us some reason to worry about the intuition that no group could be conscious. Additionally, the mere fact that it's hard to imagine how the nation of China could give rise to phenomenal experience shouldn't trouble us either. As Lycan (1987) points out, it's also really hard to imagine how a mass of neurons, skin, blood, bones and chemicals could give rise to phenomenal experience. These counter-intuitions, however, merely scratch the surface of the problems with Searle's project.

Most importantly, the modal force of these absent qualia intuitions is nowhere near as secure as Searle might hope. Consider a version of the 'fading qualia' thought experiment offered by David Chalmers (1996, 253ff). Suppose absent qualia are nomologically possible in a system made entirely of homunculi instead of neurons. Now, consider the intermediate

cases between Chalmers and homunculi-headed Chalmers (with one neuron replaced with a homunculus, two neurons replaced with homunculi, all the way up to the point where Chalmers is a homunculi-headed system). At each point, the system is functionally isomorphic to Chalmers and shares all of his behavioral dispositions. Now, if it is it's true (as the absent qualia intuition holds) that there is nothing that it's like to be homunculi-headed Chalmers, we are left with a couple of options: "Either 1) consciousness gradually fades over the series of cases before disappearing, or 2) somewhere along the way consciousness suddenly blinks out, although the preceding case had rich conscious experience" (Chalmers 1996, 255).

First, let's note that (2) doesn't seem too promising. After all, any point you pick for the disappearance of qualia will be entirely arbitrary (Chalmers 1996, 255). Why think that consciousness blinks out of existence after replacing 75% of the neurons in a system rather than replacing 1%. It doesn't seem as though there is any straightforward reason to suppose that any point along the continuum of cases suggests a promising point for ruling out consciousness. Moreover, (1) doesn't fair much better. Consider an intermediate case between Chalmers and homunculi-headed Chalmers (call him half-Chalmers). If we are allowing for fading qualia, half-Chalmers will see faded colors where we see vivid ones, and the subtle distinctions between similar qualitative states will have begun to collapse (Chalmers 1996, 256). The important thing to notice about half-Chalmers is that he will be systematically mistaken in his reports of his qualitative states. He will claim that today feels a lot hotter than yesterday *even though he experiences no differences in the temperatures between yesterday and today*. More importantly, if a functional theory of belief is right, he will even believe that today feels hotter than yesterday! "Here we have a being whose

rational processes are functioning and who is in fact *conscious*, but who is utterly wrong about his own conscious experience" (Chalmers 1996, 257).

Here's the problem with fading qualia. It is a commonplace of functional psychology and functional neuroscience that subjects are able to introspect upon their qualitative experiences and give veridical reports thereof. In fact, it might even be true by definition that qualia are introspectable.[51] Moreover, we have reason to believe that it's an empirical fact about consciousness that subjects are capable of making correct judgments about at least some of their qualitative experiences—but half-Chalmers is systematically mistaken about *all of his conscious experiences.* Abandoning this claim about consciousness is possible. However, doing so has costs. Unless we can expect the reports of a person about her conscious experiences to adequately represent her qualitative states *at least most of the time*, this will force us to abandon any use of first-person reports within psychology—and this would be a bad result for Searle. Alternatively, we could claim that the only psychology worth doing is armchair reflection on one's own mental states. But this would hardly count as psychology and it sure as hell wouldn't count as science.

Searle could (and in fact he has cf., 1992, 66-67) object to this worry by claiming that fading qualia are accompanied by changes in the propositional attitudes of the person who is being changed. According to Searle, although the various versions of Chalmers would be able to recognize the changes in his qualitative states, his continued reporting of changes in the temperature would be out of his control; although he would recognize that it didn't feel any different today than yesterday, he would hear himself saying "it feels hotter today than yesterday". However, as Chalmers (1996, 258) correctly notes:

[51] I'm inclined to follow Bill Lycan (2001) in taking qualia to be the introspectable phenomenal features that characteristically inhere in sensory experience.

> An organization-preserving change from neurons to [homunculi] simply does not change enough to effect such a remarkable change in the content and structure of one's cognitive states. A twist in experience from red to blue is one thing. But a change form ["it feels hotter today than yesterday"] to "Oh no! I seem to be stuck in a bad horror movie!" is of a different order of magnitude.

Supposing that being so radically disjoined from your conscious experience seems far less plausible than supposing that every intervening Chalmers between regular Chalmers and homunculi-headed Chalmers has conscious states. But if this is possible, and if there is no reason to suppose that consciousness could just disappear from a system on the basis of a small change, there is no reason to suppose that homunculi-headed system lacks consciousness. Now, given that Block has acknowledged that the nation of China is a special case of a homunculi-headed system, there is no reason to suppose that the nation of China couldn't have conscious states. But if this is true, then even the strongest version of Searle's arguments does not rule out the possibility of collective mentality.

Of course, at the end of the day these arguments will not be surprising to Searle. In fact, one can always adopt Searle's picture of the mind; however, the costs of doing so are remarkably high. First, Searle has a rather unscientific view of what is for something to count as a mental state. Ever since Freud (and probably earlier) we've been inclined to think that there are some mental processes that you just can't be conscious of (e.g., you're desire to sleep with your mother and kill your father, or the fact that the American dream is really just an artifact of a massive propaganda machine). Moreover, continuity with animal models and evolutionary history requires that we attribute some mental states to systems that may or may not be conscious, and the sort of sub-personal psychology that's carried out in much of cognitive science (especially in cases of priming and attention) require appealing to systems that are below the level of conscious awareness.

Second, Searle's position requires us to abandon any hope of having a naturalistic semantics. So far as Searle is concerned, the aspectual shape of thought grounded in the conscious perspectivality is the only way to make sense of primary intentionality. This rules out *a priori* any account of mentality that grounds mental content on ontogeny, phylogeny, or computational structure. Searle would not, of course, see this as an objection, and one can surely decide to abandon RTM and CTM, however, doing so requires offering an alternative account of mental content that can answer concerns about systematicity, productivity, compositionality, and other syntactic qualities of thought. Now, it's unclear to me exactly where we should go with these exclusively first person methods for studying consciousness. While I am inclined to think that there is some role to be played by first person conscious reports in cognitive science, such claims only seem to make sense as embedded in a field of third-personal results.[52] As it turns out, these are all options (or bullets, as the case might be) that Searle is willing to take (or bite). I'm not! And neither are most cognitive scientists. I thus propose at this point to leave Searle's arguments to the side.

Before moving on to my argument that there is actually good reason to suppose that research programs investigating collective mentality will offer both interesting explanations of phenomena in the actual world and interesting research projects for a more collective cognitive science, I must address some more lurking worries. While I have suggested in this chapter that worries about the consciousness of collectivities need not worry us, there are further objections to both the autonomy of collective psychological explanation and the possibility of collective representation. In the following two chapters I'll turn to each of these worries.

[52] I do have arguments for this claim; and I offer them elsewhere. However, I cannot defend them here.

Chapter III:

I JUST CAN'T GET YOU OUT OF MY HEAD

John Searle claims that any theory of collective intentionality must "be consistent with the fact that society consists of nothing but individuals. Since society consists entirely of individuals, there cannot be a group mind or group consciousness" (Searle 1990a, 404). Holistically inclined social scientists and philosophers of social science do, of course, maintain that we have good reason to appeal to collective mental states. In fact, some claims about social ontology and social epistemology seem to require the possibility of collective mental states that are distributed across individual agents. However, the intuition that collective mentality is not a viable option cuts deep, and at least one prominent philosopher of social science has argued that:

> Treating society as an organism, even metaphorically, and taking latent functions seriously force the holist to make difficult choices. He must either opt for Durkheim's '*âme collective*'—the group mind—to explain how society arranges institutions to meet its needs, or embrace a Darwinian evolutionary view, according to which all long lasting social institutions arose through variation and selection for their beneficial functions…the individualist considers either of these alternatives unattractive enough to reject holism (Rosenberg 1988, 134).

To put the point briefly, there is a prevalent intuition that any theory that's as 'offensive to the intellect' as the existence of collective mentality cannot possibly be right. However, this apparent offensiveness to the intellect is insufficient, by itself, to undercut the possibility of collective mentality. In this chapter I address one of the primary philosophical objections to

collective mentality, the thought that any appeal to collective mentality would be better cashed out in terms of aggregations of individual mental states in their social context.

3.1 A bit of history:

Auguste Comte developed one early attempt to establish the autonomy of collective psychological explanation, appealing to collective phenomena that appear to obey objectively specifiable laws distinct from the laws of individual psychology. He offered few arguments for his claim; however, Comte took the emerging sciences of social psychology and sociology to require generalizations over collective phenomena that were autonomous from generalizations over psychological states of individuals. John Stuart Mill demurred, and spent Chapter Six of *A system of logic* attempting to refute this claim. Mill famously argues that there are no collective psychological phenomena that cannot be better explained by appeal to individual psychology, so long as all of the relevant social relations and contextual facts are taken into consideration. Mill's arguments against Comte provide a foundation for the most troubling objections to the possibility of collective mentality.

Mill begins from what he sees as an uncontentious starting point: the assumption that individual behavior is law governed. Consider what would happen were I to put a cockroach in your coffee mug and ask you to drink out of it. My money is on your refusing to drink the coffee, and my reason for being so sure about this is that there are incredibly robust psychological generalizations about human behavior in response to cockroaches grounded on the human disgust responses (cf., Rozin & Nemeroff, 1990). There are, however, a minority of people who may still be willing to drink from the mug—maybe you're one of them. However, even in this case, we typically assume that there is some other psychological fact

that explains why your disgust response to the cockroach has been swamped. Mill's overriding principle, in the pursuit of explaining human behavior, is that human action is always explicable in terms of reasons for action. Unless psychological explanation is doomed from the beginning because there are chaotic psychological drives that are neither predictable nor explicable, Mill claims that there must be psychological explanations for all human behavior. On the basis of such considerations, Mill is unabashed in his defense of the lawfulness of human behavior. In fact, he goes so far as to say that:

> given the motives which are present to an individual's mind, and given likewise the character and disposition of the individual, the manner in which he will act might be unerringly inferred; that if we knew the person thoroughly, and knew all the inducements which are acting upon him, we could foretell his conduct with as much certainty as we can predict any physical event (1843/1988, 23).

This is not, of course, to say that we actually have—or in fact that we ever will have—the tools required to produce such accurate descriptions of the motives, character, dispositions, etc, of individuals. Perhaps psychology will never be able to advance these *absolutely certain* predictions and explanations. However, Mill claims that on the basis of the relative accuracy of psychological predictions, it is at least possible that psychology could produce such predictions and explanations.

Mill does, however, recognize that it is likely that a science of psychology will always, because of human epistemic frailty, be cashed out in terms of *ceteris paribus* laws (what Mill (1843/1988, 31) calls empirical laws). But we do well enough with these. If, for example, I know that Granny refuses to talk about the French before she's had a stiff martini, and I try to talk to her about Comte, I can predict that she'll put up her index finger, walk to the kitchen and pour herself a stiff martini, *ceteris paribus*. However, even here I have to keep in mind the fact that there might be other beliefs and desires that I've failed to

acknowledge in Granny. Perhaps Granny also believes that Comte was Italian, or that the French have poisoned her Gin, or that the earth is about to be hit by a giant meteorite. In any of these cases, we won't actually be able to predict Granny's behavior. However, this is only because of our ignorance of a range of facts that happen to be pertinent to our psychological explanation of Granny's behavior.

The *ceteris paribus* nature of the laws of psychology does not, however, impugn their status *as laws* for Mill. In part, this is because Mill takes the *ceteris paribus* nature of psychological laws to be a result of human epistemic limitations rather than a fact about the psychological phenomena itself.[53] In this regard, Mill takes psychology to be analogous to the science of tideology, in which "circumstances of a local or casual nature, such as the configuration of the bottom of the ocean, the degree of confinement from shores, the direction of the wind, &c., influence in many or in all places the height and time of the tide" (Mill 1843/1988, 31). Developing a precise tideology, thus, must include all of these factors; likewise a precise psychology must rely on a number of ever-changing and hyper-local facts about the systems in which particular humans happen to find themselves.[54] Thus, when we study the psychology of individuals, we necessarily rely on approximations and idealizations in order to do the predictive work. However, people are relatively similar to one another and circumstances aren't really all that variable, at least concerning the issues we typically care

[53] Mill actually believes that no fundamental laws are ceteris paribus laws. He even goes so far as to claim that ceteris paribus laws amount to nothing more than the lowest sort of empirical law (Mill 1843/1988).

[54] It's actually worse than this. Mill argues that "the impressions and actions of human beings are not solely the result of their present circumstances, but the joint result of those circumstances and the characters of the individuals; and the agencies which determine human character are so numerous to be diversified (nothing which has happened to the person throughout life being without its portion of influence,) that in the aggregate they are never in any two cases exactly similar. Hence, even if our science of human nature were theoretically perfect, that is, if we could calculate any character as we can calculate the orbit of any planet, *from given data*; still, as the data are never all given, nor ever precisely alike in different cases, we could neither make positive predictions, nor lay down universal propositions." (Mill 1843/1988, 33)

about for the purposes of prediction and explanation, and this allows us to generate approximate generalizations—which Mill thinks will be good enough for the purposes of social psychology (Mill 1843/1988, 34).

This predictability of human psychology, however, leads Mill to question the relationship between psychological and neurological laws. Mill recognizes that many of the predictions required in order to develop a completely lawful theory of human action will be facts about her neurophysiology. In fact, the analogy to tideology suggests that a completely articulated human psychology will appeal to circumstances of a local or casual nature, such as the configuration of a person's brain and the influence on her brain from outside forces influence in many or in all places the behavior of that individual. But if this is true, then the generalizations of psychology begin to look useful *merely* because of our epistemic limitations—and, a better explanation of human behavior might be the one that appeals to these lower-level causal processes. However, Mill argues, there are compelling reasons to take psychological laws to be more than useful shorthand for generalizations over neurological states.

Consider what would happen were we to come to be aware of all of the uniformities obtaining between psychological states and physiological states. Even in this case, we would still have to admit that there is at least a difference in mode of access to facts about human psychology and facts about physiology suggesting that generalizations over psychological states and generalizations over neurological states are about different sorts of things. We are, as a matter of fact:

> wholly ignorant of the characteristics of these nervous states; we know not, and at present have no means of knowing, in what respects one of them differs from another; and our only mode of studying their successions and co-existences of the mental states of which they are supposed to be the generators or causes. The successions,

> therefore, which obtain among mental phenomena do not admit of being deduced from the physiological laws of our nervous organization; and all real knowledge of them must continue, for a long time at least, if not always, to be sought in the direct study, by observation and experiment, of the mental successions themselves. Since, therefore, the order of our mental phenomena must be studied in those phenomena, and not inferred from the laws of any phenomena more general, there is a distinct and separate Science of Mind. (Mill 1843/1988, 37)

The thought, here, is that we come to know about psychological states of persons in the absence of any knowledge about their neurophysiology. Moreover, we recognize immediately that facts about human psychology could remain stable across counterfactual variations in states on which these psychological states are realized. Psychology is meant to explain the regularities in human behavior, and such explanations do not require any *particular* story about the realizers for these states (though they may make some of these stories more or less plausible).

Given that there are significant differences in our mode of access to psychological and physiological facts, Mill argues that we have strong *prima facie* reason for thinking that psychological explanations ought to remain autonomous from physiological explanations. Perhaps more interestingly, Mill thinks that the only way we have of studying physiological states *qua* realizers of the psychological states is by first looking at the psychological facts that explain a person's behavior, and then trying to see if there are interesting similarities at the level of realizers. Mill thinks that unless we start with an account of psychological kinds, we'll have no way to way to construct a psychological story even from a complete science of neurology. On the basis of such considerations, Mill argued for an autonomous science of psychology to be pursued by way of experimental methods. Unfortunately, Mill thinks that things don't go so well for attributions of mentality to collectivities.

Mill claims that when we examine social structures, we find that there is nothing new

in large-scale a social phenomenon that is not already present in a fully articulated account of the psychological states of the individuals. As Mill puts the point, "the effect produced, in social phenomena, by any complex set of circumstances, amounts precisely to the sum of the effects of circumstances taken singly" (Mill 1843/1988, 83). Whereas the science of psychology requires the emergence of new phenomena such as beliefs, desires, and emotions that are not specifiable in terms of the aggregation of the underlying neurological phenomena, there is nothing new in kind that is introduced by appeal to states of a collectivity. In defense of this position, Mill offers an argument on the basis of a commitment to empiricism and a commitment to deductive relations between the laws of psychology and laws about social phenomena—I turn now to these arguments.

3.2 Mill's argument for reduction:

Mill claims that the empirical investigation of social phenomena never rests on the observation of emergent entities (Mill 1843/1988, 65). In observing groups of people, we find aggregations of people, each of whose behavior is describable in terms of her psychological states. Mill, thereby, claims that empiricist considerations always militate against reifying collective mental states. When we explain the behavior of a rioting crowd, for example, we do so in terms of the rioting individuals: some people are setting things on fire, others are breaking windows, others are turning over cars, and still others are throwing rocks at the police. There's a prominent intuition here that there's nothing more to explain about the behavior of the crowd *per se* than what's specified by facts about the behavior of individuals and the relations between these individuals as they come together in a crowd. Building on this intuition, Mill argues that in every case where a social scientist posits a

social phenomena, such posits are shorthand for, and are conceptually entailed by a complex set of facts about the individual psychological states that constitute and produce that social phenomenon. To put the point briefly, since we don't encounter social phenomena *per se* in the world, claims about these social phenomena must be conceptually reducible to psychological phenomena.

One way of reading Mill's argument for the reduction of collective psychological posits to individual psychological explanations is grounded on the claim that laws ranging over collective psychological phenomena will always be reducible to laws of individual human psychology. On the basis of his commitment to individual behavior being completely determined by psychological laws, Mill argues that if we had a complete story of individual behavior as it occurs *in the social world*, there would be nothing more to explain at the level of the collective psychology once the laws of individual psychology are applied. This would make collective psychological phenomena at best, redundant and at worst, explanatorily superfluous. Thus, while it might be true that interesting local circumstances arise in groups, these circumstances don't generate any new phenomena that require autonomous social laws above and beyond the laws of individual psychology and an account of the circumstances that constrain the behavior of individuals within a particular collectivity. Were there autonomous psychological states of collectivities, we would expect new laws for social phenomena distinct from the laws of individual psychology—and we just don't seem to need such laws in order to explain human behavior. As we approach rock-bottom explanations of human behavior, Mill contends that we'll find that "human beings in society have no properties but those which are derived from, and may be resolved into, the laws of the nature of individual man" (Mill 1843/1988, 65). Perhaps an example of such a reduction will be

useful here.

At the foundation of the science of political economy, a social science in which Mill was thoroughly invested, Adam Smith (1776) argued for the claim that apparently large-scale phenomena such as fluctuations in industrial revenue could be explained in terms of the aggregation of self-interested psychological states of individuals. Given that all economic behavior is grounded on the pursuit of a greater gain over a smaller gain (cf., Mill 1843/1988, 105), Mill claims that we can deductively explain the apparently emergent phenomena of market trends. However, most cases of the emergence of social phenomena are much more complicated that this (because of the variety of factors that contribute to their production—some of which are psychological, some of which are not), and for this reason, their deduction is far less obvious—though Mill contends, there is no case where it is obvious that this deduction is impossible or even improbable. This is why, in most cases, we have to start by positing ceteris paribus laws about social phenomena and explain how these phenomena are likely given what we know about human psychology and contingent facts about the particular humans that happen to constitute the relevant collectivity. This is less than complete reduction of these social laws, but it's good enough, claims Mill, to demonstrate the possibility of such reductions.

3.3 Initial troubles for the Millian argument:

There are a number of ways in which this idea of inter-theoretical reduction can be spelled out. When Mill makes the claim that the laws of the social sciences will always be reducible to the laws of psychology, he likely had something like the following in mind. Suppose that there is a law of collective psychology that takes the following form. If a

corporation, C, fears a decrease in its profit margin. D, it will change its marketing strategy, M, ceteris paribus. If Mill's thesis about the reduction of collective psychology to individual psychology is correct, then we will need a series of bridge laws such that:

1) $Cx \leftrightarrow P_1x$
2) $Dx \leftrightarrow P_2x$
3) $Mx \leftrightarrow P_3x$

where 'P_1x', 'P_2y', and 'P_3x' are predicates of psychology. These laws are called bridge laws because they contain predicates of both the higher-level collective psychology and the lower-level individual psychology and are thus capable of bridging the gap between the higher-level and the lower-level sciences. However, to complete the reduction, there must also be a law of psychology such that:

4) $P_1x \bullet P_2x \rightarrow P_3x$

If this sort of picture is correct, then any law that includes an apparently collective psychological phenomena will be related to a law at the level of individual psychology such that if we knew all of the laws of psychology and all of the bridge laws, we could, thereby, explain the apparently collective psychological states of a system in terms of the psychological regularities that underwrite that collective behavior. Such reductions were the wildest dreams of the positivists. However, there are serious, and well-known problems with adopting such a reductionist project.

On the reasonably untendentious assumption that laws must range over natural kind predicates,[55] the sort of reduction posited here comes out to be far too strong. So, suppose that the predicates picked out in the antecedents and consequents in (1) through (4) are

[55] If you're worried about the natural kind talk here, you can feel free to replace it with something more amenable to your empiricist proclivities. I'm inclined to think that the same argument will go through even on the minimal assumption that the predicates used in scientific laws must be projectable predicates rather than gerrymandered, gruesome predicates. I'll run it in terms of natural kinds; however, since I'm borrowing the argument from Fodor (1980) and that's the way that he does it.

supposed to be natural kind predicates. While it might indeed be true that it's possible to spell out notions like 'profit margin', 'corporation' and, 'marketing strategy' at the level of economics, sociology or cultural anthropology, it seems reasonable to assume that these notions will not be realized on anything that looks like a natural kind at the level of individual psychology. After all, these are *functional kinds* if anything is. So, while an individual may play a key role in designing a marketing strategy in response to a change in profit margin, she does so only in her role as a member of a corporation.

To put the criticism another way, there are numerous psychological attitudes that any particular individual within the corporation can adopt toward the development of a marketing plan without affecting the eventual outcome. Given that this is the case, we actually do find the sorts of counterfactual stabilities that Mill claimed were absent at the level of collective psychology. Suppose someone who is involved in the production of a marketing plan for Wine and Co. is lambasted by a supervisor for failing to wear a Hawaiian shirt on the second Friday of the month. She might, given her everlasting hatred of tropical climates, adopt the policy of attempting to undercut the corporation by producing the worst marketing plan she can possibly muster. Fortunately for Wine and Co., however, there are structures in place to mitigate breakdowns in the functional architecture of the corporation. Despite her best efforts to the contrary, a viable marketing plan may emerge because her supervisors might see that her version of the marketing plan, recognize that it looks miserable, and thereby redirect the project so as to allow another person to produce a more viable plan for increasing profits.[56] Alternatively, the same person could decide that she ought to become more of 'team player'

[56] If V.S. Ramachandran (1988) is right, then something similar occurs across a broad range of neurological breakdowns in individuals. Areas of cytoarchitecture that would typically play one role in the functional architecture of a person are recruited in order to do the work of some area that is damaged, and so if failing to perform it's standard function. Perhaps the most interesting cases here are Ramachandran's studies on phantom limb pain.

and might decide that producing a viable marketing plan is the best way to begin—in the end producing a relevantly similar and viable marketing plan.

Psychological states of corporations, if there are such things, are often produced in a way that is resistant to local breakdowns; they, thereby, exhibit counterfactual stability across variations in the psychological states of the individuals that compose these collectivities. The distribution of cognitive tasks across a number of individual psychological systems, at least as this occurs in the most interesting cases of collective mental states, facilitates the realization of a particular psychological state of a collectivity on a variety of different individual psychological bases. Thus, just as we should be unwilling to engage in a straightforward type reduction of the mental states of an individual to her neural architecture because the kinds that are present at the psychological level are multiply realized on different, heterogeneous neural structures, we should be unwilling to engage in a straightforward type reduction of collective psychological states to individual psychological states because these states are multiply realized on a variety of heterogeneous individual psychological states. The important point here is that the counterfactual stability of things like corporate beliefs, marketing plans, and the like guarantees that they will not be straightforwardly reducible by conceptual means to the natural kinds of individual psychology.

Any attempt to reduce collective psychological phenomena to individual psychological phenomena is going to be faced with the fact that although collective entities like a decreased profit margin are well behaved at the level of collective psychology, the psychological states that realize the movement of capital and the decisions to act on the movement of capital will be wildly disjunctive and completely unprojectable at the level of

individual psychology. There is a nearly infinite set of ways in which profit margins can fluctuate, and as such, the psychological attitudes directed towards these fluctuations would themselves be nearly infinite. The point is a familiar point against the straightforward reduction of the laws of one science to the laws of another science. This suggests, however, that the initial way of spelling out Mill's reductive project is likely to fail; yet the claim that we should expect the *a priori* deduction of the laws of the special sciences to something more primitive continues to have a great deal of currency.

3.4 Reduction and supervenience:

While the conceptual reduction of the social sciences to psychology is unlikely, there is a weaker reading of Mill's reductionist claims on which they might succeed. On almost anyone's account of collective behavior, unless there are changes in individual psychological states or environmental conditions, there will be no change in the states of the collectivity that these individuals compose. However, if Mill is right to claim that the social sciences must be reduced to psychology, then every social fact must admit of some sort of reductive explanation; a more promising understanding of such reductions is to see them as requiring nothing more than each of social fact being explained entirely in terms of simpler entities. Perhaps the most promising way of spelling out such reductive explanations in the case of mental entities is suggested by David Chalmers (1996; Jackson 1998; Jackson and Chalmers 2001).[57]

Chalmers argues that if you want to reduce B-properties of one type to A-properties of another, then it will be a minimal condition on such reductions that B-properties supervene

[57] Chalmers argues that if materialism is true every fact will admit of a reductive explanation in physical terms, and although he's not immediately concerned with the reduction of the social to the psychological, his story is helpful in spelling out an alternative account of reduction by way of supervenience.

on A-properties. Spelling out this notion of supervenience, Chalmers notes that B-properties supervene on A-properties just in case no two possible situations are identical with respect to their A-properties while differing in their B-properties. There are, however, a number of ways in which the relevant notion of possibility can be understood. Suppose someone wanted to explain why all worlds that are identical with regard to their physical properties would also be identical with regard to their biological properties. In offering an explanation of such similarities across worlds, one might begin with a notion of nomological or natural possibility that could constrain reductive explanations to all and only worlds that are identical to ours as regards the natural laws. Spelling out supervenience according to this notion of possibility, entails that any *naturally possible* situations with the same A-properties will have the same B-properties (Chalmers 1996). However, as Chalmers notes, this is a weak notion of possibility so it will not entail the sort of reductions that the Mill hoped for. A nomological relation between the social facts and the psychological facts cannot assure the deductive entailments that Mill requires between collective psychological explanations and individual psychological explanations.

The problem is this. If B-properties (e.g., the facts of the collective psychology) are only nomologically supervenient on A-properties (e.g., the facts of individual psychology), then it is possible to conceive of a world in which the A-facts hold but the B-facts don't; and, provided conceivability is a good guide to possibility, this suggests that a world in which facts about individual psychology are the same and facts about collective psychology are different is metaphysically possible. The proponent of Millian deductions must recognize that such explanations require a stronger sort of necessity.[58] Nomological supervenience is not

[58] I am inclined to think that these sorts of deductions are unlikely to be forthcoming in any form. In fact, very few people (perhaps only Jackson, Chalmers, and Joe Levine) are convinced that a priori deductions are

strong enough to guarantee that the introduction of B-facts into a world doesn't offer something new that requires their own explanation in B-terms.

What Mill needs is a notion of supervenience that *guarantees* that if the collective psychological facts supervene on the individual psychological facts, *any* two situations that are identical in individual psychological facts will *necessarily* be identical in their collective psychological facts. To put this point another way, if Mill's argument is to be successful, the supervenience relationship between collective psychological facts and individual psychological facts must be sufficient to guarantee that the individual psychological facts deductively entail the collective psychological facts. But, in order to guarantee this, we have to opt for a much stronger interpretation of the supervenience relationship, what Chalmers (1996) refers to as logical supervenience. Logical supervenience is the claim that B-properties supervene on A-properties just in case no two logically possible situations are identical with respect to their A-properties but different with respect to their B-properties. If this sort of relation obtains, then once God creates a world in which the A-properties are fixed, she doesn't have any more work to do in fixing the B-properties. The B-properties are fixed as a matter of *logical necessity* once the A-properties are fixed and there is no conceivable world in which the B-facts differ while the A-facts remain the same. Suppose, for a moment, that this sort of relation obtains between the facts of collective psychology and the facts of individual psychology. How, then, would explanatory reductions work within this framework.

Jackson and Chalmers argue that there are *a posteriori* identities that obtain between the facts that articulated in macro-scientific explanations and their subvenience bases. This

---

required for explanation. However, this research project does seem to have some affinities with Mill's reductive project, so I'll entertain these possibilities.

thought seems, at least initially, quite plausible: the explanation of macroscopic phenomena takes the form of an analysis of the mechanisms that give rise to a particular sort of macroscopic phenomena. Thus, for example, if we want to explain why a car is accelerating slower than it typically does, the explanation should not be given at the level of the whole car; instead, the phenomena should be explained at a lower explanatory level by appeal to facts about clogged fuel injectors, bad spark plugs, or even old spark plug wires.

Similarly, as Dennett is fond of pointing out, when a person fails to behave rationally, for example when a person with Anton's syndrome is obviously blind but denies being so, we explain her behavior not at the level of psychological phenomena but in terms of facts about her neurology; this, for example, might occur by way of an appeal to the sort of damage that has occurred to her occipital. Finally, if we want to explain why NC State fails to defeat UNC at basketball even though they're playing to sort of offense that should cause problems for UNC, we'll appeal to the athletic abilities of UNC's players and to the breakdown of the Princeton Offense because of the particular mistakes made by the members of the NC State team.

Many explanations clearly do take the form of explanation in terms of simpler entities. In fact, this is precisely the sort of explanatory model on which the homuncular functionalism that I advanced in Chapter 1 is grounded. However, the truly astonishing claim advanced by Jackson and Chalmers, and the claim that is required for Mill's reductive project, is that these sorts of explanations rest on *conceptual truths* that allow for an upward derivation of the macroscopic facts from their microscopic realizers. The reason for such a claim is that unless there is some good reason for thinking that the macroscopic facts

logically supervene on the physical facts, we will not have fully explained the macroscopic facts. Here, in brief, is the sort of argument Jackson and Chalmers typically offer:

1) If materialism is true, every fact will (eventually) admit of reductive explanation in physical terms;

2) Reductive explanation of B-facts in terms of A-facts requires logical supervenience of the B-facts on the A-facts;

3) Phenomenal facts don't logically supervene on the physical facts;

4) So, phenomenal facts don't admit of reductive explanation in physical terms;

5) So, materialism is false.[59]

There are, of course, a number of contentious premises in this argument—and I'm not going to argue against this position (Bill Lycan (2003) has done a nice job of pointing to a number of problems with this argument—so I'll leave that to him). However, there is a version of this argument that captures Mill's reductive intuition about collective psychology. Although Mill would not follow Jackson and Chalmers in their claim that physicalism is false, he would be concerned to avoid a similar untoward conclusion about empiricism. Mill's empiricist version of this argument takes something of the following form:

1) If empiricism is true, then every psychological fact will (eventually) admit of reductive explanation in terms of individual psychologies;

2) Reductive explanation of B-facts in terms of A-facts requires logical supervenience of the B-facts on the A-facts;

3) But, empiricism is true;

4) So, collective psychological facts admit of reductive explanation in terms of individual psychological facts;[60]

[59] A version of this reconstruction of this argument occurs in Lycan (2003)

[60] This relies on a special case of (2): unless collective psychological facts logically supervene on individual psychological facts they won't admit of reductive explanation in terms of individual psychological facts.

5) So, collective psychological facts logically supervene on individual psychological facts.

As with the Jackson and Chalmers argument, there are a variety of places at which to resist this argument. However, before turning to the ways in which one might resist this argument, let me turn briefly to the sorts of arguments that might be marshaled in favor of this reductive picture of explanation.

3.5. Some stabs at definitions:

There are a couple of ways in which the attempt to move upward from the claims of individual psychology to the claims of collective psychology can be carried our. First, it could be the case that collective psychological phenomena are definable in terms of individual psychological states. For example, ascriptions of psychological states to a collectivity might be read distributively, or as summative claims about the members of a collectivity (cf., Gilbert 1989). According to such an analysis of a collective belief attribution,

A collectivity C believes that P iff all or most of the members of C believe that P.

This, however, is an insufficient model of collective belief attribution. After all, if every individual in a collectivity *privately* believes that P, but no one ever shares her belief that P with any other member of the collectivity, we would be unwilling to say *of the collectivity* that it believed that P on this basis. Consider the case of a hiring decision in a philosophy department. If everyone in the department believed that hiring a candidate would be good idea, but everyone kept this belief private, we would be unwilling to say of the department that it thinks that this hire is a good idea—even though every member of the department thinks that it's a good idea. We might, then, modify this definition to say something like:

> A collectivity, C, believes that P iff 1) most members of C believe that P, and 2) it's common knowledge in C that most members of C believe that P.

However, this claim has problems as well. The most troubling worry is that information can sneak in under the common knowledge condition that is irrelevant to the belief of the collectivity *qua* belief of the collectivity. So, for example, it could be the case that every member of a philosophy department hiring committee believes that the fall semester begins in late August, and it could also be the case that it's common knowledge among the members of that committee that all of the members of the committee believe that the fall semester begins in late August, and this belief could be so utterly irrelevant to any decision that the committee makes about hiring that it would seem strange to attribute to the committee a belief that the fall semester begins in late August.[61] Perhaps this suggests that summative models of collective beliefs are too weak to count as viable analyses of collective belief attributions.[62]

Moreover, there are numerous cases where we claim that a group believes that P when none of the individual member of that collectivity believe that P. Again, consider the case of a hiring decision in a philosophy department. It could be the case that there is no member of the faculty who believes that the present candidate is the best candidate to hire. However, it might still be the case that every member of the faculty is willing to assent to the claim 'the philosophy department believes that this is the best candidate for the job' qua

[61] The reasoning here is grounded on the thought that we shouldn't attribute to the hiring committee *qua* hiring committee any beliefs that are irrelevant to hiring.

[62] Moreover, as Bill Lycan (personal correspondence) has pointed out, this analysis ignores all of the relevant power relations at play in such a group. It is possible that most of the members of the hiring committee believe that person A is the right candidate to hire, and that it is common knowledge that most of the member of the committee believes that person A is the right candidate to hire, but that the department chair and the head of the hiring committee believe that person A should not be hired. In this case, the fact that the majority believe that person A should be hired, even coupled with the common knowledge constraint, is insufficient to tell us anything interesting about how the hiring committee is likely to behave.

member of the department even if she is unwilling to assent to this claim as a privately held belief. This suggests that summative models are too strong (as well as too weak) to count as analyses of collective belief attributions.

One final approach to offering the sort of definitions that are required by the Millian begins by noting that the theoretical terms of psychology can be defined as the occupants of a particular causal role, or at least in terms of the typical causes and effects of that particular sort of mental phenomena (cf., Lewis, 1972; Armstrong, 1980). According to this view, the theoretical terms of the ascription of psychological states to collectivities can be implicitly defined within the theory in which they occur. As such, these theoretical terms can always be replaced in any explanation by their definientia (e.g., by a suitable Ramsification of the terms), providing a way of fully explaining the presence of some collective psychological phenomena by way of the simpler entities of the psychological sciences. On this approach, we begin by supposing that a particular mental state of a collectivity (e.g., the intention to run the Princeton Offense) is going to be defined as whatever fills the causal role of getting people in the positions that are required in order to run the Princeton Offense and getting them to move in the ways that are constitutive of the Princeton Offense. We then suppose that empirical investigation into the cases in which the Princeton Offense is run will reveal that the occupant of this role is a set of individual beliefs about where each of the individuals need to be. It will then follow, straightforwardly, that since the collective intention to run the Princeton Offence is whatever fills the role of getting the individual players in the right positions, and since what gets the individual players in the right position is their individual beliefs, these individual beliefs just *are* the intention to run the Princeton Offense.

There are, however, serious problems with this position as well. First, the mere fact that the two levels of explanation pick out the same phenomena does not entail that the higher level of explanation should be abandoned. Even if it were true that the psychological state of a collectivity and the psychological states of the individuals that composed that collectivity were identical, there might still be reasons to retain both sorts of explanatory structures (and this is something that even Mill wouldn't deny!). In fact, offering the sort of analysis that is required by this causal model might even provide us with reasons to retain collective mental states in our ontology. I agree wholeheartedly that our commonsense understanding of mental states is likely be vindicated, though perhaps with substantial revision, by the cognitive sciences. However, even if there are token identities between an individual's mental state and the state of her brain, this does not thereby commit us to eliminativism about the mental.

As I've already noted, there are substantial counterfactual stabilities both in the case of individual and collective psychology, that are not present in the explanations that appeal to the realizers of these states. Moreover, commonsense psychology is extremely useful in prediction and explanation—and it's useful precisely because it generalizes across individuals and across collectivities. When we want to explain why a person ducks as a basketball is thrown at her face, our best bet is to appeal to the psychological state of the individual (she didn't want to get hit in the face with a basketball because she thought it would hurt) rather than to her neurological state. In the same way, if I want to explain why UNC is likely to struggle in a game against Georgetown in the same way that it struggled in a game against NC State, I will be much better off if I explain this prediction by noting that UNC plays a transition offense and both NC State and Georgetown play a Princeton Offense.

The explanation here will turn on the ways in which a Princeton Offense is used to slow down the tempo and keep the score low (which is incredibly detrimental, if executed correctly, to a transition offense). Moreover, I can offer this explanation without regard to the particular players on either team. However, just as I might appeal to facts about a particular person's neurology in order to explain why the standard psychological explanations fail, I will also be able to appeal to particular facts about particular players in order to explain why UNC would have no difficulty dispensing with Northwestern even though they too play the Princeton Offense. The point here is that the ascription of an intention to run the Princeton Offense does a lot of work so far as our predictions about upcoming basketball games are concerned, and that it generalizes in a way that appealing to facts about particular players psychologies doesn't.

Second, it is unlikely that the relevant sorts of identities will have the form required by the Millian model. In order for this sort of model to be correct, it will have to be true a priori that a particular mental state is identical to whatever has the right sort of causal structure. If materialism is true, then it will be true that there are some physical realizers for every mental state. However, it would be an interesting discovery that there is some reasonably homogeneous class of neurological states that will count as an individual's intention play her role in running the Princeton Offense. In fact, it seems fairly unreasonable to assume that there will be such a unified class of occupants for this role. Things get even more ugly when we start to consider the heterogeneity of the class of individual states that could implement the collective intention to run the Princeton Offense. Though there will assuredly be some class of individual mental states that give rise to this collective intention, what the relation is between these states at the level of individual psychology is an empirical

question in a way that is not allowed by this sort of causal theory of mind. Although it may turn out that there is some sort of identity to be found between the mental states of a collectivity and the mental states of the individuals that compose that collectivity, there is little reason to think that this is something that we can figure out without investigating the phenomena in question.

Finally, as Michael Bratman (1987) and John Searle (1990a) and argued, no set of psychological states of the form 'I believe P', even when supplemented with some sort of common knowledge criteria, can ever be summed up to produce a psychological state of the form 'we believe P'. The problem is that in any case where we find an attribution of a mental state to a collectivity, we find that even though there are individual intentions that can be used to explain at any point what is happening in a group, these intentions are often *derived from* collective intentions. Thus, when a basketball team tries to run the Princeton Offense, the play might start with four players outside the arc and one player inside. The players will then keep the ball in constant motion until a player at the post tries to make a backdoor cut (hoping for a bounce-pass so that she can take a lay-up) or until a defensive mismatch (for example, when the opponent packs the paint to prevent backdoor cuts) allows a for a three-point shot from the perimeter. Now, while it might be the case that every individual on the team has the belief "we should run the Princeton Offense" this individual belief is something that has to be derived from the coordination that occurs between these players as a team.[63] An individual cannot run the Princeton Offense by herself. Moreover, even if every individual has the belief that she should run the Princeton Offense (and the corresponding desire to do so) this will not suffice to produce a Princeton Offense. Finally, even if every

[63] NB: there are ways of developing this position into an argument for individualism. I return to this in the next section.

individual has the belief that they should run the Princeton Offense (and the corresponding desire to do so), and every individual has the belief that every other individual has the relevant beliefs and desire to run the Princeton offense, this will not suffice for running the Princeton Offense unless there has been a coordinating decision by the team as a whole to do so. Since, individual beliefs by themselves cannot insure the relevant sort of coordination, they're not going to be sufficient to explain the collective intention.

At this point, we have good reason to think that there is not a straightforward way in which we can define the psychological states of the collectivity in terms of the psychological states of individuals. However, the reduction of the collective psychology to individual psychology does not require explicit definition of social phenomena in terms of the psychological states of individuals. Instead, if there is a reductive story to be told about collective psychological phenomena, it may be told in terms of functional analyses or functional decompositions of the relevant collective psychological phenomena into facts about the individual psychological states that compose the collective state. I turn now to this approach.

3.6. Functional analyses:

Instead of appealing to reductions by way of explicit or implicit definitions of collective psychological phenomena in terms of individual psychological states, one might turn to the response that is often given to concerns about multiple realizability: some form of functional explanation. There is a broad consensus in the philosophy of mind that individual psychological states are best understood by attending to the contribution that they make to the functioning of a system as a whole, facilitating the *flexible* engagement with the system's

environment. On this sort of view, mental states are understood as types that are explicable only in terms of abstract, functional characterization, though it is often thought that they are token-token identical to physical processes. At the level of collective psychological phenomena, we might then say that although an intention to run the Princeton Offense must be specified functionally, there will be individual psychological realizers for each instance of these collective phenomena.

A crucial feature of a mental state, characterized functionally, is that it has the capacity to contribution to the goal-directed behavior (or disposition to behave) of a system. Moreover, in explaining the behavior of a system, our best bet is to attribute to it the intentional states that are likely to give rise to a particular sort of behavior.[64] The difference between a belief and a desire, for example, is—at least to a first approximation—spelled out in terms of the different roles that each plays in allowing a system to engage in various sorts of behaviors. The most promising forms of functionalism start by recognizing some intentionally characterized phenomena (e.g., Roy Williams' belief that repeated shifts in the line-up will facilitate an up-tempo game that will wear down the less athletic Georgetown) and then breaking it down into simpler components until a purely mechanistic explanation is reached (cf., Dennett 1978b, 80). This view takes a cognitive system to be a complex entity consisting of a number of subroutines, each of which is capable of carrying out some task or other in the service of the person-level phenomena in question. In the case of William's belief, long term and working memory structures (e.g., his memories of past performances of the Tarheels and his memories of things that Dean Smith taught him when he was a young

[64] This is not, of course, to say that every mental state has large-scale effects on the behavior of a system. There are idle and inconsequential thoughts—it's just that they're a whole lot harder to track from the outside.

assistant coach) and perceptual structures (e.g., his observation of the way that his team is moving tonight) among other systems play some role in producing Williams' behavior.

While it might be true that the best explanation of Williams' behavior is that he has a belief that repeated shifts in the line-up will facilitate an up-tempo game that will wear down the less athletic Georgetown, a robust functional explanation in psychology would attempt to offer some characterization of this intentional phenomena in a way that explains it in terms of the functioning of simpler entities. Thus, rather than taking Williams' belief to be a black box that's meant to be left unopened, functional psychology attempts to open this black box and offer some account of why this mental state functions in the way that it does. However, at this point we need to tread carefully. In the case of a person, or in the case of other biological organisms, these decompositions will eventually be resolved into dumb mechanisms such as neurons that can only be in a state of firing or not firing. However, this is not the only way that things can go. Functionalism is a topic-neutral theory. The functionalist is, thus, able to remain completely non-committal about the sorts of entities that fill any particular causal role. Provided that there are some intentional phenomena that can be attributed to a collectivity, and provided that there is some decomposition into the simpler entities that is homologous to the sorts of decompositions that we find in a paradigmatic cognitive system, there is no reason to think that functional characterizations of the mind will apply only to individuals and not to collectivities. Moreover, the fact that the most promising model of functional analysis that we find for mental states is offered in terms of what Marvin Minsky calls a 'society of mind', it seems unreasonable to assume a priori that there are no collective mental states that have functional architectures of the same sort that we find in individual mental states.

More importantly, suppose that there is some functional state of a collectivity that we call a belief, a desire, or intention. Suppose further that this functional state is, at one level of analysis, realized on the mental states of the individuals that compose that collectivity. This fact would give us no more reason to eliminate the collective psychological phenomena than we have to eliminate individual psychological states because, at some level of analysis, they are realized on neuronal states. The reason that such functional models of the mind are so promising is that they are explanatorily useful without having to make any claims about what sorts of entities happen to realize that functional state at some level of analysis. Even if it is likely that there will be token-token reductions of collective mental states to individual mental states—this just isn't a problem.

3.7. Reduction and modal intuitions:[65]

There is one more route by which the Millian might try to offer a reductive account of the relation between individual mental states and collective mental states. This is a strategy recently advanced by Frank Jackson and David Chalmers. In spite of the worries that mid-twentieth century philosophy raised against the *a priori*, Jackson and Chalmers argue that the only way to avoid mystery in our scientific explanations (or at least the only way to justify scientific explanations by appeal to something other than an appeal to faith that the correlations established in the sciences will secure the identities required for the defense of physicalism) is by way of *a priori* entailments between physical truths and ordinary macroscopic truths. According to Jackson and Chalmers, the best way to go about securing these *a priori* entailments is by way of conceptual analysis.

The story goes something like this. Suppose we want to establish that 'water =

[65] Thanks to Jacek Brzozowski for help on the development of the argument in this section.

H_2O'. We begin by noting that when provided with sufficient information and sufficient time for reflection, ordinary subjects have the capacity to identify the extensions on their concepts. Much of this capability is the result of considering "a concept's extension within hypothetical scenarios, and noting regularities that emerge" (Jackson and Chalmers 2001, 322). Although what emerges through this sort of conceptual analysis is unlikely to be anything like an explicit definition, it does give us important information about the features of a particular kind of thing that allow us to use concept to use a concept to apply to that thing. Through a process of reflection on the places in which she uses the concept, the fearless conceptual analyzer can come to note that WATER refers to the stuff that we typically find in lakes and rivers, the stuff that comes out of the tap at home, and the stuff in the glass from which I am currently drinking. It's just a matter of competent use of the concept that it discriminates the things that are *water* from the things that aren't. In fact, fearless conceptual analyzers will even know that if they were to come across a substance that looked, smelled, tasted and behaved in all ways just like water, dripping on her head from the pipes in her favorite bar, that this too was water. The important thing about our concepts is that we have the capacity to consider where they apply both in actual situations and in counterfactual situations.

What we see here is that there is at least one characteristic reference fixing property for the 'water' role. Now, it's *a priori* true that 'water' is the actual 'watery stuff of my experience', the stuff that falls from the sky, drips from my tap, and runs through the coffee maker. This is just a matter of the meaning of the concept WATER. So, it's true *a priori* that the actual occupant of the 'water' role is water. Next, however, we need to establish that it is actually H_2O that fills the 'water' role. But, this is a contingent fact about our world. Fortunately, however, chemical science has the capacity to figure out how it is that all of the

stuff that plays the 'water' role are one in the same kind of stuff at the level of chemistry—it turns out that all of these things are H_2O, and that's a contingent matter of fact about our world. The stuff that plays the water role in the actual world is H_2O. Now, if we know (by way of scientific evidence) that the actual occupant of the 'water' role is H_2O, and we know that the actual occupant of the 'water' role is water, then we can infer *a priori* by the transitivity of identity that water is H_2O.

Suppose, then, that we want to establish an analogous case with a collective mental state. We would need to show that when provided with sufficient information and sufficient time for reflection, ordinary subjects have the capacity to identify the extensions on their collective mental concepts in such a way that a fearless conceptual analyzer would come to note that 'collective belief' refers to the characteristic outputs of deliberations and investigations within collectivities, the coordinating force of these decisions, etc. In order to establish this, there would need to be at least one characteristic reference fixing property for the 'collective belief' role. If someone could succeed here, it would be *a priori* true that 'collective beliefs' are the actual 'collective belief-ish phenomena of my experience'—whatever these happen to be. This would just be a matter of the meaning of 'collective belief'. So, it would be true *a priori* that the actual occupant of the 'collective belief' role is whatever the natural class of phenomena that underwrite our claims about collective beliefs happens to be. Next, we would need to establish that it is actually individual psychological states that fill the collective belief role. This, however, would still be a contingent fact about our world. The things that play the 'collective belief' role in the actual world would be individual psychological states. Now, if we know that the actual occupant of the 'collective belief' role is collective belief, and we know that the actual occupant of the 'collective belief'

role is individual psychological states, then we can infer *a priori* by the transitivity of identity that collective beliefs are individual psychological states.

As it turns out, much of the literature of collective intentionality is quite amenable to this form of argument. Consider one promising reductive picture of collective intentionality: Bratman's 'shared intentions'. Bratman begins from a couple of methodological assumptions. First, he assumes that intentions are a distinctive sort of attitude that is integral to our understanding of ourselves as agents—a sort of attitude that facilitates planning and practical reasoning (Bratman 1993, 97). Second, he assumes that although it is clear that we often do attribute mental states to collectivities, the only way in which we can make sense of intentions is by attributing them to individuals in such a way that they allow for apparently collective behaviors. Bratman acknowledges straight away that an individual can't run the Princeton Offense, play a Balinese gamelan, or play Steve Reich's music for 18 musicians. However, Bratman claims that by making sense of the intentional states of each of the individuals that compose a collectivity, it is possible to make sense of each of these behaviors in a way that doesn't suppose any sort of collective mental state. Bratman argues that the key role of an intention is to coordinate and constrain behaviors in such a way that they lead to intentional actions. If I intend to roll a cigarette, I thereby commit to a range of actions that will constrain my behaviors in various ways. I commit to taking my tobacco out of my pocket, pulling out a rolling paper, pinching out the appropriate amount of tobacco, rolling the cigarette, and licking the adhesive strip to close the cigarette. This is how my behaviors over the next couple of minutes are coordinated by my intention to roll a cigarette. In the same way, if we intend to stage a work stoppage in order to protest pay inequalities, we *thereby* commit to various sorts of constraints on our behavior as well as to ways in which

we will coordinate each of our individual behaviors. With this shared intention, we commit to leaving our work-stations at the same point, to refusing to return to work until some suitable result has been achieved, to not trying to undercut one another, and to preventing scabs from entering the work area (just to name a few constraints on behavior). Provided this understanding Bratman then asks what it would take for such a shared intention to be possible. His account of how we share an intention, then, takes the following form. We intend to Φ iff

1) (a) I intend that we Φ, and (b) you intend that we Φ;

2) I intend that we Φ in accordance with and because of (a) and (b) and we have meshing subplans in accordance with (a) and (b),[66] and you intend that we Φ in accordance with and because of (a) and (b) and we have meshing subplans in accordance with (a) and (b); and,

3) (1) and (2) are common knowledge between us (cf., Bratman 1993, 106).

Bratman contends that in any case where there is a collective intention, we will find such intentional states in the individual.

Bratman, thus, agrees that there are collective intentional phenomena that must be explained. He concedes that there are collective intentions that are individuated by their ability to coordinate the actions of the various individuals that compose a collectivity. In fact, Bratman seems to think that it is a conceptual truth that there is such a reference fixing property for 'shared intentions'. But if this is a conceptual truth, then it will be *a priori* true that 'shared intentions' are the actual shared intentional phenomena of our experience, and this is just a matter of the meaning of 'shared intention'. So, it will be true *a priori* that the

[66] Bratman (1993, 105-6) claims that this is a condition on ensuring that the coordinating feature of a collective intention is met. While it need not be the case that the persons who compose a collectivity have all of the same plans underlying a collective intention (I might have the subplan of making a big scene when I stop working and you might have the plan to stop working quietly and just sit down for a cup of coffee), it must be the case that our plans mesh in the sense of not preventing the intended action from occurring.

actual occupant of the 'shared intention' role is shared intentional phenomena of our experience. Next, Bratman contends that we can establish that it's actually individual psychological states that fill the shared intention role. Having provided the account that Bratman offers, we could look to the world and see if there were such individual intentions and this would be an empirical matter. If Bratman is right, then the things that play the 'shared intention' role in the actual world will be these sorts of individual intentions. And, if the actual occupant of the 'shared intention' role is shared intention, and the actual occupant of the 'shared intention' role is individual intentions, then we can infer *a priori* by the transitivity of identity that shared intentions are a particular individual intentions. This allows Bratman to make two claims. First, it allows him to claim that there are shared intentions but they are reducible to individual psychological states. Second, it allows him to claim that these shared intentions are not properties of a collective mind, but properties of the individuals that compose the collectivity. However, it's not at all clear that Bratman can have everything he wants here.

Although he speaks disparagingly of collective minds and their ilk, noting that "a shared intention is not an attitude in the mind of some superagent" (Bratman 1993, 99), it's not at all clear that he gets this claim for free. As David Velleman (1997, 38) notes, the claim that there are no collective minds is not as obvious as Bratman seems to think. "Whether there are collective minds depends on whether there are collective mental states…Hence we cannot rule out the possibility of collective intentions on the grounds that there are no collective minds; the direction of logical dependence goes the other way". However, recognizing this makes a huge difference to the way in which we interpret the relevant reductions. The mere fact that something is reducible to something else doesn't, all by itself,

rule out the possibility of things existing at both levels of explanation. The fact, if it is a fact, that individual mental states are reducible (at least token-token) to neurological states does not, all by itself, entail that we should eliminate individual mental states from our ontology.

The problem is that there is always an extra step required in order to argue from the claim that two things are of the same sort to the claim that one of the things ought to be eliminated. Consider the case of individual beliefs. Suppose that there is some analogous story to be told about the way in which individual beliefs are to be reduced (at least token-token) to states of a person's neurology (as they no doubt will if the physicalist is right). The fact that there is a reductive story to be told about individual beliefs *would not* force us toward the eliminativist position unless we had some commitment to retaining only the most basic, or primitive explanatory structures in our ontology. At this point, we would be left with atoms and the void—likely an incredibly unpromising strategy for the practice of psychology. To put the point another way, even if at the end of the day all that we find at the basement level of physics is atoms and the void, the fact that we find the sorts of entailments proposed by this explanatory model is neither necessary nor sufficient for allowing these higher-level explanatory structures into our ontology. To put the point another way, there is always a question of 'location' versus 'elimination' from our ontology (Jackson, 1998); however, at least on this picture, the fact that we find the *a priori* entailments they actually mandate the retention of such phenomena in our ontology. Just as Jackson (1998) wants to retain semantic properties and solidity even though they are not concepts of basement level physics, this argument, by itself, would not warrant elimination. Thus, I claim that even if someone were to develop such a reductive account of collective mental states, we would have no need to be troubled by these reductions. At the end of the day what really matters are

the counterfactual stabilities at the level of collective psychology; and no reductive story is going to take those away.

3.8 Another reason not to worry about reduction.

Recall that my reconstruction of Mill's empiricist argument in favor of reduction takes the following form:

1) If empiricism is true, then every psychological fact will (eventually) admit of reductive explanation in terms of individual psychologies;

2) Reductive explanation of B-facts in terms of A-facts requires logical supervenience of the B-facts on the A-facts;

3) But, empiricism is true;

4) So, collective psychological facts admit of reductive explanation in terms of individual psychological facts;[67]

5) So, collective psychological facts logically supervene on individual psychological facts.

However, there are substantial reasons for resisting both (1) and (2). Premise (2) seems *crazy*! Getting to the point where one thinks that reductive explanation requires anything as strong as logical supervenience requires a whole lot of theoretical apparatus that very few people have been willing to adopt and against which there are very strong arguments. I'll not offer the arguments here (see Lycan 2003); however, I do find them incredibly compelling and sufficient to suggest that such an argument need not bother us in the first place. However, even if this argument were supplemented with a much weaker notion of supervenience, there is still a lot to say about the adoption of premise (1).

[67] This relies on a special case of (2): unless collective psychological facts logically supervene on individual psychological facts they won't admit of reductive explanation in terms of individual psychological facts.

As I noted earlier, Millian arguments turn on the claim that we don't perceive anything as a collectivity but only as aggregates of individuals. However, this point is neither as clear nor as intuitively obvious as it may seem. As Bloom and Kelemen (1995) have demonstrated, young children have the capacity to acquire a novel collective terms (i.e., terms referring to a group or collective entity) in a way that facilitates the use the term to refer to entities that have permeable physical boundaries. Although there is an overwhelming tendency towards the early acquisition of terms that refer to whole objects, we come to learn very early on that some 'objects' are distributed. The clearest examples of such terms refer to things like flocks (of birds), herds (of antelope), and bunches (of grapes). This does not, however, demonstrate that we are capable of immediately seeing social groups as individual entities. However, Bloom and Veres (1999) have argued that the same system that's dedicated to theory of mind attributions for individuals can be brought on-line in order to drive the attribution of mental states to collectivities, and—this is the important point—such judgments may be able to underwrite judgments about whether a particular collectivity counts as a single entity or not.

In fact, there is good reason to think that many of the judgments that underwrite a prejudice towards experiencing the world in terms of individuals are deeply indebted to the particular cultural conditions under which these judgments arise.[68] Nisbet et al. (2001) review evidence suggesting that although Western subjects typically think in analytic terms that readily suggest the reduction of collective phenomena to individual phenomena, this is not a culturally universal tendency. East Asians, for example, tend to be more holistic in their analysis of human activity, often taking collectivities to be the primary locus of practical

[68] The following two paragraphs draw, with substantial revision, from Huebner, Bruno, and Sarkissian (under review).

activity. Following Roger Ames (1994), we might, then, distinguish two senses of 'individual'. Given predominantly Western predilections, 'individual' is typically taken to refer to single, indivisible entities that come to be members of various collectivities in virtue of adopting some particular psychological attitude toward that collectivity. On this understanding of 'individual', considerations of autonomy, independence, equality, privacy, and freedom play the key role in determining which systems are capable of practical action. However, 'individual' can also be understood as a locus or focal point within a web of social relations. On this conception, 'individual' refers to the unique focal points of social relations that are both abstractions from collective structures as well as the loci of practical activity that collectively determine the properties of collectivities. On such a view, individuality is achieved not in opposition to one's social relationships but by way of the distinct roles that are occupied within a collectivity.

That such diverse conceptions of individuals exist is further bolstered by a number of recent studies. Menon et al. (1999, 702) suggest that "while prevailing American theories hold that persons have stable properties that cause social outcomes and groups do not, the theories prevailing in Confucian influenced East Asian cultures emphasize that groups have stable properties that cause social outcomes". Moreover, Morris, Menon, and Ames (2001) have provided evidence suggesting that East Asian subjects employ conceptions of agency that are highly social when reasoning about what gets to count as entity. Building upon this foundation of individualism/collectivism research, Kashima et al. (2005) investigated differences in the attribution of entativity by East Asian and Western subjects. They found at least two sorts of considerations that drive judgments of entativity: psychological essentialism and agency. Psychological essentialism includes both perceived internal

consistency (i.e. the extent to which perceptions of individuals that belong to a group are likely to resemble one another in appearance and behavior) and perceived unalterability (i.e, the belief that the properties of a collectivity aren't changeable because it has some underlying essence). Kashima et al (2005) found that insofar as being a single entity is understood in terms of psychological essentialism, individuals are perceived to be more entity-like than collectivities cross-culturally. However, when considerations of agency are adopted, there is significant reason to think that individuals are perceived as entities more often than collectivities are only under the sway of Western ideologies (Kashima et al. 2005, 162). In other words, East Asian subjects are far more willing to attribute agency to collectivities than are Western subjects, and on this basis East Asian subjects are far more willing to class collectivities as single entities than are their Western counterparts.

To put the point more brusquely: it's only from a uniquely western and liberal standpoint that this sort of individualism makes sense. Thinking that individuals are the only sort of intentional systems that there are requires a peculiar act of reification of an abstract entity, the individual person (cf., Nietzsche 1887/1998).[69] The positing of an '*âme collective*' need not require the positing of a new substance that emerges above and beyond the individuals that compose that collectivity. That's just to say, claiming that there must be a particular sort of physical substrate underlying every sort of mental act is a presupposition that we need not adopt. In fact, the assumption that individual, physically bounded subjects

[69] Note that I am not making the claim here that individual human animals are abstractions from collectivities. Rather, my claim here is that in tracking something as an intentional system we rely on a particular sort of theoretical commitment to claims about which sorts of systems are capable of intentional action. I am inclined to think that there are very strong evolutionary pressures (e.g., seeing another critter as a mate or as a threat) that would militate in favor of seeing physically bounded entities as intentional actors. However, such pressures are likely to be significantly weaker militating in favor of seeing a collectivity as an intentional system. For this reason, the primary factors that are operating in this case will be social pressures that will vary across cultures and even across a variety of social milieu. I suggest that this gives us very strong reasons to be skeptical of our intuition about collective mentality—whether or intuitions are strongly individualist, strongly collectivist, or multileveled and layered in some way.

always exist as the locus of any sort of practical activity is far from intuitively obvious, especially once we come to realize that many of the sorts of actions about which we are concerned are defined functionally in a way that doesn't allow them to be reduced straightforwardly or translated without loss into claims about their constituents, and this has been the standard criticism of this aspect of Mill's picture for a long time.

Maurice Mandelbaum (1955, 307), for example, argues that it's impossible to understand the actions of human beings as social organisms except on the assumption that some "concepts which are used to refer to the forms of organization of a society cannot be reduced without remainder to concepts which only refer to the thoughts and actions of specific individuals." In defense of his claim, Mandelbaum offers the example of making a withdrawal from a bank. Although there are a number of aspects of this procedure that can be explained in terms of individual beliefs and behaviors, there are also a number of things that beg for an explanation that overruns these psychological states of individuals. For example, part of the explanation of the procedure will probably be spelled out in terms of filling out a slip of paper and handing it to another person in order to get her to hand me some notes and some coins; however, this is far from the whole story. If we are to explain the procedure of making a withdrawal from a bank, we will have to make appeal to both the institution of banking and to the social roles that are produced within that institution. In order for something to count as a withdrawal it has to be a transaction that occurs between a 'customer' and a 'teller' within the confines of a 'bank' or similar economic institution.

Mandelbaum advances this as a claim about concepts, but the point of his argument cuts deeper. Not only is it true that we can't engage in these sorts of conceptual reductions, it's also true that we can't make sense of the occupants of such social roles unless we posit

some sorts of institutional structures to which these individuals can belong. Now, while it might be true that there is one way to tell the story that starts from the individuals and builds up, it's not at all clear that doing so doesn't require a particular set of philosophical presuppositions that we're not warranted in adopting. The problem here is that any appeal to particular facts about particular people will neglect the fact that there are a number of ways in which something can play the right functional role, as specified within the institutional structure of banking, that are not specifiable in a way that will facilitate the reduction of these terms to the psychological states of particular individuals. More importantly, this suggests that there are many cases in which the our experience the world is best understood in terms of individuals playing roles within various sorts of collective structures. What it is to be a bank-teller is something that can only be spelled out in terms of functional roles that make reference to higher-level institutions. Just as we need to make an appeal to a theory of combustion engines if we are to develop a theory of carburetors, we must appeal to institutions such as banks if we are to develop a theory of bank-tellers, and it is precisely on this point that Mill's theory founders. To put the point briefly, people actually do experience the world as consisting of collectivities (though the extent to which they do so is, to a significant degree, a matter of cultural convention).

Here's how this bears on premise (1). Mill argues that his commitment to empiricism require that we perceive only individuals as cognitive systems and that we don't perceive collectivities as cognitive systems. However, it is reasonable, in light of the evidence that I've adduced in this section, to think that this just isn't the case. Bloom and Kelemen's (1995) data suggests that even young children have the capacity to track collectivities as entities with permeable physical boundaries and Bloom and Veres' (1999) data suggests that

theory of mind attributions are deployed in order to make sense of the behavior of both individuals and collectivities. The important thing to notice here is that so long as Mill allows psychological states into his empiricist ontology, he will also be forced to allow collective psychological states as well. After all, we don't *see* psychological states. Instead, we perceive such states by way of a low-level, typically nonconscious inference from the behavior of a system to its intentional states. However, this data suggests that systems are in place in all of us that allow such inferences for both individuals and collectivities.

Moreover, even if there is a tendency amongst Westerners to take 'individual' to refer to single, indivisible entities, the data collected by Menon et al. (1999, 702), Morris, Menon, and Ames' (2001), and Kashima et al. (2005) suggests that the individualist frame of reference is the result of a cultural tradition that supposes that there must be a particular sort of physical substrate underlying every sort of mental act. Fortunately, this supposition is far from intuitively obvious, especially once we come to realize that many of the sorts of actions about which we are concerned are defined functionally in a way that doesn't allow them to be reduced straightforwardly or translated without loss into claims about their constituents. Keeping these facts in mind, it seems far from obvious that a commitment to empiricism requires a commitment to individualism. At the end of the day, Mill's commitment to empiricism is thus orthogonal to any claim about collective mentality.

3.9 Superfluity arguments: a first attempt

Mill's intuition that collective psychological explanations are not autonomous from individual psychological explanations has retained a great deal of currency. In searching for a legitimate sociological method, Max Weber followed Mill in arguing for methodological

individualism in sociology. Although he acknowledges that we often treat "social collectivities, such as states, associations, business corporations, foundations, as if they were individual persons"(Weber 1914/1968, 13); however, "in sociological work these collectivities must be treated as solely the resultants and modes of organization of the particular acts of individual persons, since these alone can be treated as agents in a course of subjectively understandable action" (Weber 1914/1968, 13). Floyd Allport voices a similar sentiment at the foundation of social psychology when he claims that "there is no psychology of groups which is not essentially and entirely a psychology of individuals"; and given that this is the case, "Social psychology must not be placed in contradistinction to the psychology of the individual; it is a part of the psychology of the individual, whose behavior it studies in relation to that sector of his environment comprised by his fellows" (Allport, 1924, p. 4; cited in Kashima et al. 2005, 148; emphasis in original).

Moreover, debates in mid-twentieth century philosophy of social science also centered on the possibility of offering reductive explanations for commonsense and scientific attributions of mental states to collectivities. Friedrich Hayek (1942), building on the arguments offered by Weber, claims: 1) that there are distinctively psychological facts that can't be accounted for in physical terms, and 2) that these psychological facts about individuals provide the only foundation on which sociological explanation in terms of human practical activity can be grounded. Hayek claims, appealing to such psychological facts requires that all attempts at sociological explanation:

> start from what men think and mean to do, from the fact that the individuals which compose society are guided in their actions by a classification of things or events in a system of sense qualities and concepts which has a common structure and which we know because we, too, are men, and that the concrete knowledge which different individuals possess will differ in important respects (Hayek 1942, 283).

The question, then, is: how might this reductionist argument be spelled out in a way that demonstrates sensitivity to the commonsense willingness to attribute mental states to collectivities while developing some reason why an adequate cognitive science need not allow for such states?

The important thing to notice, here, is that a strong metaphysical thesis, like logical supervenience, is not required in order to impugn collective mentality from the standpoint of cognitive science. Even if there could be collective mental states in some far off corner of the universe, unless there is some reason to think that collective mentality is a possibility in *this world*, in the world we experience, there is no reason to think that scientific explanations ought ever to be couched in terms of collective mentality. Thus, the most troubling sort of argument I now face is the argument that there is no good reason to think that we'll find collective psychological phenomena in our world. If the explanatory structures that I've been attempting to develop thus far are not useful for the cognitive sciences, I might as well throw in the towel. And unfortunately, there are a number of cases in which one might initially think that collective mentality is the way to go, but in the end, it isn't going to be explanatorily useful in the cognitive sciences

In line with this sort of worry, Robert Wilson (2001 and 2005) has offered one of the most compelling arguments against collective mentality. Wilson argues that the proponent of collective mentality must show that some collectivities exhibit *at least one* paradigmatically psychological ability or process (though it's likely that if a system exhibits one such state it will also exhibit more). Wilson (2001, S266) also notes that although a clear definition of 'psychological' is unlikely to be forthcoming; however, he believes that we have a good idea what the paradigmatic cases of psychological states are. These include perception, memory,

imagination, attention, motivation, consciousness, problem solving, believing, desiring, intending, trying, fearing, willing, and hoping. Wilson claims that if we find some collectivities that possess some of these states, then we will have good reason to suppose that there are collective mental states. Wilson, however, believes that we will find no cases in which any of these mental states should be ascribed to collectivities once we are clear about the details of the psychologies of the individuals that compose a collectivity.

Wilson begins by noting that psychological and biological scientists often ascribe mental states in a merely figurative sense. This means that we need to be careful not to treat ascriptions as literally true unless we have a good reason to suppose that the phenomena in question are best understood *as states of the collectivities* rather than states of individuals. Wilson argues that our ascription of mentality to collectivities can be interpreted in two ways. First, we might mean that there are properties of collectivities (in this case, mental states) that are not merely properties of the individual members of that collectivity. These traits will, of course, be multi-level traits.[70] However, for any claim about the mental states of a collectivity to be theoretically interesting, the proponent of collective mental states has to hold that these states are something beyond the states of the individuals that compose the collectivity (Wilson 2001, S265). Second, that ascriptions of collective mentality could be understood as claims about individual psychological states that are exhibited only when individuals are part of a collectivity. Wilson (2004b, 418) calls the latter position the social manifestation hypothesis (SMH).

[70] By multi-level traits, Wilson intends to pick out those traits that can be possessed both by the collectivity and by the individual organism. According to the most promising articulation of the collective mentality hypothesis, collectivities will have mental states like beliefs, desires, intentions, perceptions, and the like which are *of the same sort* as the mental states of individuals.

SMH is, at minimum, the claim that some psychological states are manifested only when individuals are embedded in a social group. SMH also, however, allows for an inference from these conditions for the manifestation of these properties to the claim that individual psychological states are all that exist, psychologically speaking. Wilson (2004b, 418) claims that SMH allows us to both recognizes that social situation and cognition are linked "in more than an instrumental way or as cause to effect", as well as allowing us to posit "cognition itself as irreducibly social, and so not as supervenient on the intrinsic properties of individuals". SMH is, thus, capable of capturing many of the intuitions that have typically underwritten appeals to collective mentality. After all, on this view there actually are cognitive phenomena that "arise themselves as social abilities, as ways of negotiating aspects of the social world" (Wilson 2004b, 418). Wilson, thus, dodges all of the arguments against reductive accounts of collective psychology that I have addressed so far. Although he claims that cognitive phenomena will always be explainable in terms of the psychological states of individuals, Wilson also recognizes that there will be numerous phenomena for which we need to appeal to the particular social structures in which a person is embedded in offering psychological explanations. However, in retaining a commitment to methodological individualism, Wilson also acknowledges that we arrive at rock-bottom explanations of psychological phenomena only when we have an adequate account of the psychological states of individuals—as they are embedded in particular social structures—that give rise to these collective phenomena. There is, then, a new range of phenomena that is left to be explained when we turn to the social world in which individuals find themselves. It's just that this social world is best explained in terms of the psychological states of

individuals, rules for their aggregation, and facts about the social constraints on individual action.

In developing a defense of SMH, Wilson argues that the phenomena to which proponents of collective mentality have typically appealed are best understood as cases of SMH rather than as cases of genuine collective mentality. Wilson persuasively argues (see Wilson 2004a, Chapter 11) that the collective psychology tradition of the late 19th and early 20th century failed to distinguish genuine collective mentality from SMH, and because of this, their claims about collective mentality typically rested on the claim about shifts in individual psychology rather than on appeals to genuinely collective cognitive phenomena. Gustav Le Bon (1895/2002), for example, argued that when individuals constitute a crowd, they become unconscious automata under the sway of a sort of hypnosis—prey to the suggestions of a collective mind over which they have no control.[71] Le Bon claims that in crowds, sentiments escalate, becoming the sole determinants of the behavior of the individuals that constitute the crowd and that the individuals that compose the crowd become unresponsive to reason and evidence, and they will follow any leader charismatic enough to direct the sentiments of the crowd. To put the point briefly:

> "[t]he fact that they have been transformed into a crowd puts them in possession of a sort of collective mind which makes them feel, think, and act in a manner quite different from that in which each individual of them would feel, think, and act were he in a state of isolation" (Le Bon 1895/2002, 4).

However, mere appeals to changes in the psychological states of the individuals that compose a collectivity will not, themselves, be sufficient to ground any claim about collective

[71] As Le Bon puts the point: "An individual in a crowd is a grain of sand amid other grains of sand, which the wind stirs up at will" (Le Bon 1895/2002, 8)

mentality. And, at the end of the day, Le Bon actually argued that the explanation of crowd behavior would be best spelled out in terms of the dormant, savage desires in every human being that were left over from 'primitive ages' (Le Bon 1895/2002, 27).

Turning to a contemporary defense of collective mentality, Wilson (2001 and 2004) also demonstrates that the ascription of collective mentality is once again better understood in terms of SMH rather than as an appeal to genuinely collective mental states. David Sloan Wilson has attempted to argue that the capacity for collective decision-making is evidence for the existence of collective mental states. However, collective decision-making, by voting, for example, looks to be just another case of a change in the sorts of states that individuals can exhibit because of the social situations in which they are embedded. Wilson claims that "even if the decision here is viewed as distinct from those of the individual voters—if there is a group mind here it is nothing over and above the minds of the individuals" (Wilson 2001, S269). Likewise, D.S. Wilson's arguments that religion is a group-level trait only makes sense as a claim about the ways in which individual psychologies are made possible partially determined by the collectivities to which they belong. At the end of the day, the propagation and practice of religious norms is contingent on the psychological states of the individuals who belong to some religious movement or other.

In the end, Wilson claims that the appeal to collective mentality is superfluous because there are no explanations of apparently collective phenomena that are not better understood as appeals to the ways in which individual psychology changes when individuals find themselves in groups. Although this is not a knock-down argument against collective mentality, it does suggest that the explanatory value of collective mentality is null unless there is some reason to suppose that there are states of the collectivity that aren't just states of

the individuals that compose a collectivity. The question is: what would it take for there to be emergent cognitive states of collectivities that *were* interestingly distinct from the cognitive states of the individuals that compose that collectivity.

3.10. Two sorts of emergent cognitive phenomena:

I concede that most defenses of collective mentality fail to distinguish between a defense of collective mentality and SMH. However, there are tools for distinguishing between the two sorts of collective states in a way that makes the defense of collective mentality stronger as well as distinguishes between genuinely cognitive collective phenomena and phenomena that ought to be understood as only figuratively or derivatively cognitive. Let's start with an example.

A relatively common case in which one might ascribe a mental state to a collectivity might take the form "The Dixiecrats believe that the South will rise again". In this case, although it may be true of many people who identify as Dixiecrats that they believe the South will rise again, and although it may be true that there is a platform advanced by this splinter wing of the American Democratic Party that constrains the members of the Dixiecrat Party to make claims about how the South will rise again (on the heels of racist practices), it doesn't seem as though we gain any explanatory advantage by ascribing a belief to the Dixiecrat party rather than ascribing that belief to each of the members of that party. A Martian psychologist who knew of all the psychological states of the individuals that compose the Dixiecrat party, and who knew all of the relevant rules of behavior for remaining a member of the Dixiecrat party, wouldn't be missing out on anything explanatorily interesting if she failed to make a claim about the psychology of the Dixiecrat party *per se*. While there may

be reasons such as relative ease of prediction, or epistemic limitations of human agents, that might lead us to attribute such states to collectivities, this is not enough to compel us to think that there are interesting mental states of the Dixiecrat party—at least not in this case.

At least part of what has gone wrong in this case is precisely what went wrong in the study of the collective psychology tradition of the late 19th century. Briefly, starting with crowd behavior as a paradigmatic case of collective cognition was a mistake! Wilson is right to note that crowd behavior is better understood by way of the SMH. It is clearly true that an individual cannot riot by herself. She can set things on fire, throw bricks through a window, and even attempt to turn a car over. But she will not be rioting unless she finds engages in these behaviors under the right social circumstance. Le Bon correctly note that when people find themselves in a riot they see things as reasonable that they wouldn't otherwise, their emotions change, and they become less responsive to reason. However, all of this can be explained in terms of the psychological states of the individuals as well as rules for the aggregation of their behavior.

This points us to an important sort of lesson. The presence of self-organizing behavior in a collectivity is never sufficient, by itself, to demonstrate the existence of collective mental states. In fact, the presence of self-organizing systems bolsters Mill's intuition that there are laws of aggregation for social phenomena that take away the mystery of collective behavior. Perhaps the most widely known case of emergent collective behavior on the basis of simple facts about the individuals that compose a collectivity is the segregation phenomena studied by Thomas Schelling (1971). Here's one way of telling the story. Suppose that we have an equal number of philosophers and neuroscientists in a room who are distributed randomly. Now suppose that these people have to find comfortable situations for conversation, and

suppose that although each of the philosophers is willing to converse with neuroscientists (and vice versa) no philosopher wants to end up in a conversation where more than 50% of her conversation partners are neuroscientists (and vice versa). Schelling (1971) demonstrates that starting from these individual psychological states, using random movements of individuals, an eventual state of the collectivity exhibiting between 75-90% segregation is likely to occur. Although each of the individuals is willing to integrate, the collectivity will end up segregated. The interesting thing about these phenomena is that although they demonstrate that there are interesting states of the group that emerge out of the individual beliefs and simple algorithms for movement within the constraints of the group, these emergent collective phenomena, that are quite interesting in their own right, *do not* lead us to posit anything interestingly cognitive at the level of the collectivity. If we were to assume that statistical trends within a collectivity are sufficient for collective mentality, we would just be falling prey to another failing of the 19th century tradition of collective psychology.

In attempting to make sense of which phenomena are genuinely cognitive, it helps to recognize that there are *at least* two ways in which higher-level phenomena can emerge out of lower-level phenomena. Following Andy Clark (1997, 73ff), I distinguish between direct emergence and indirect emergence. Direct emergence is grounded on the properties of individual elements of a system coupled with rules for composition. In cases where we find direct emergence, the state of a collectivity, even where it diverges from the states of the individuals that compose that collectivity, is determined by the state of the individuals and rules for aggregating these individual behaviors. Note that this model of emergence allows us to retain the intuition advanced by Wilson in the guise SMH by noting that the psychological states of the individuals that compose the collectivity are quite often contingent on the sort of

collectivity to which that individual belongs. However, in this case there are no interesting feedback relations between the collectivity and the environment. To put the point bluntly: if you want to stop a riot, you attack individuals with riot cans of mace, teargas, and nightsticks—and that's because the state of the riot is just an aggregation of the states of individuals.

Indirect emergence, however, requires that the individual systems involved in the production of collective phenomena use various aspects of the environment in order to coordinate unified, genuinely collective behavior. Clark uses the example of the nest-building behavior of termites, which is causally mediated by the modification of local environments with chemical trace intended to coordinate the goals of the termites as a whole. Now, while I'm not at all sure that we should take the termite mound to count as a single cognitive system (though that is, of course, an open and empirical question at the end of the day), there is a quite important lesson to learn about collective cognition at this point. The point that I want to take from Clark (1997) is that cognitive systems, including collective cognitive systems must be systems that traffic in internal representations of some sort. The stimergetic algorithms used by the termites is one sort of representational structure, however, it's not the only one. Fodor's LOT is another—but this is not the only option for thought. I claim that if we are to take a system to be a genuinely cognitive system, then it must be the case that that we will be able to develop some sort of analogue to a psychosemantics for that system in order to explain how it is that genuinely mental representations are passed between the subsystems within a collectivity.[72] To put this point another way, if we are going to attribute genuinely cognitive states to a collectivity, it will have to be the case that there is

[72] There is, of course, a lot to say about this as a genuine possibility. In fact, the most troubling argument against collective mentality is that it won't be possible to give an adequate psychosemantics for a collectivity. But, that's a topic for Chapter 4.

some story to be told about how the intentionally specified behavior of a collectivity is to be understood in terms of the contribution that each of the parts make—as parts of that system—to the furthering of the goal in question. This leads me to my final remark about what we should learn from Wilson's picture of collective phenomena in terms of SMH.

To begin with, consider an analog to the Laplacean Martians that troubled Dennett (1987a, 1991b). Suppose that there were creatures that didn't have to appeal to anything over and above individuals but that could survey a particular aggregation of individuals and determine the psychological state of every member of that aggregate. My contention is that even if there were such people, they would be missing something perfectly objective in the patterns of collective behavior, behavior that is only describable by way of the attribution of mental states to collectivities. Dennett (1987a, 1991b) claims that if his Martians did not see that there were indefinitely many physical realizers that could be substituted for the ones that give rise to fluctuations in stock prices without perturbing the subsequent operations of the market), they will have failed to see a real pattern in the world. I claim, analogously, that if there were cognitive scientists who could predict what press release would be produced by a corporation just by appeal to individual psychological states, they would be missing a real pattern in the world if they failed to realize that this output could be the result of indefinitely many psychological states of the individuals that compose that corporation.

Predicting that a corporation will be secretive about its plan to release a new product that will revolutionize the field is easy from the standpoint of commonsense psychology. Moreover, to suppose that there must be some particular realizer (e.g., a certain sort of neurological machinery or even a system that is physically bounded) or structural just seems silly. The most promising explanatory project, both at the individual level and at the

collective level, is likely to turn on regularities such as the passing of representations from one system to another in a way that's capable of sustaining commonsense psychological regularities. However, to suppose that we need to go all the way to the individual level to explain the apparently cognitive properties of collectivities seems an unwarranted presupposition. Commonsense psychology is not committed to any theory about realization of mental states. Even if commonsense psychology has it that beliefs are information bearing states that arise from perceptions (or something quasi-perceptual) and that, together with appropriately related desires, lead to intelligent action, there will always be further questions as to which critters have beliefs. Unfortunately, there are further arguments waiting in the wings to explain why it is that collectivities cannot have genuinely cognitive states.

3.11 Superfluity arguments: a second attempt

I close this chapter by looking quickly to the range of phenomena that many people have taken to be the most promising avenue for the ascription of genuinely cognitive states to collectivities: formally organized institutional structures. At first blush, these systems seem to be the most likely place to find the sorts of regularities to which I have just appealed. The reason for this is that such systems are set up in such a way that the parts are capable of working together in a genuinely coordinated way in order to produce some intentionally specified behavior. However, focusing on formally organized institutional systems such as labor unions, courts, and corporations, Robert Rupert (2005) contends that there is nothing in such collectivities that should we should be willing to countenance as genuinely mental states. Rupert argues that:

> It seems explanatorily unnecessary to equate these physical formulations with autonomous cognitive states. After all, every step in the construction of such representations, as well as every step in the causal sequence alleged to involve the effects of those representations, proceeds either by brute physical causation (e.g., photons emitted from the surface of the page stimulate the reader's retinal cells) or by causal processes involving the mental states of individuals (Rupert 2005, 5ms).

Rupert's argument is an attempt to demonstrate that because the semantic properties of such purportedly collective representations diverge from the psychosemantics of mental representations, the positing of such representations as genuinely mental will cut no explanatory ice.

There is, of course, an easy and immediate response to this argument—and, strangely enough, it's a response that Rupert (2005, 7ms) himself considers and quickly dismisses. The response runs as follows. Anyone who adopts some version of physicalism about the mind will say that something analogous holds for individual cognition. Every process involved in the production of an individual representational state proceeds either by brute physical causation or by some other causal processes. Without a story explaining why the realization of collective mental states on physical processes is problematic in a way that doesn't prove to be problematic for the realization of individual mental states on physical processes, this argument seems to have incredibly untoward consequences. Without some story about the difference in the import of the realization relations for collective and individual representations, any attempt to deny collective representation on this basis will also be sufficient reason to deny the possibility of individual representation. Presumably, denying the possibility of individual mental representations is not something that Rupert would be too happy about.[73]

[73] In conversation, Rob Rupert has told me that if the cards do fall this way, he is willing to concede that the elimination of mental states is a live option. Rupert believes that he has an account of the semantic properties of

Rupert (2005, 7ms) sees that someone might offer this response to his argument and he responds in kind by claiming that "the two cases differ greatly with respect to our understanding of what are sometimes called 'inter-level relations'." In defense of this claim, Rupert argues that we have little idea how to explain the reduction relationship between psychological regularities and neurological regularities, however, we have a rather clear understanding of the relationship between the representations that collectivities seem to traffic in (e.g., press release and written opinions of a court) and the mental and physical states that underwrite them. But Rupert needs to say more here.

His argument runs as follows. Naturalistic theories of mental representation typically rely on nomic relations between perceptual (or quasi-perceptual) processes and properties of things in the world. Such relations are supposed to explain how neurological states indicate or carry information about properties of things in the world. However, the states of collectivities (understood as such rather than as states of individuals) don't seem to indicate or carry information about anything (*except* as mediated through person-level representations). While the content of person-level representations is specified in terms of nomic relations between perceptual or otherwise information bearing states of individuals and properties of things in the world, the content of public-language structures are specified in terms of the person-level representations required for their production and interpretation. However, if a collectivity exhibits apparently cognitive activity that is reducible to "the cognitive states of individuals (including their construction of rules for combining individual activity in a principled way)" (Rupert, personal correspondence), positing collective mentality seems superfluous. The content of these public-language structures is reducible to

---

mental states that will apply to individuals and not to collectivities. I'm not sold on his response; however, I leave the discussion of the possibility of collective representation for the next two chapters.

the content of individual representational states (including their construction of rules for combining individual representations in a principled way), so claiming that they are genuinely collective representations seems explanatorily superfluous. Fortunately there are problems with this explanatory superfluity argument.

To begin with, naturalistic theories of mental representation require that every process involved in the production of any mental representation proceed by some causal or otherwise physical processes. So, unless there are unique difficulties in positing collective representations, this argument has the untoward consequence that any denial of collective representation on the basis of explanatory superfluity will double as an argument for the denial of individual representation. But unless the denial of individual representation is on the table as a viable option, something must have gone wrong.

Perhaps there are difficulties raised by the realization of collective representations on causal processes involving the mental states of others that are importantly distinct from worries about the realization of individual representations. As Robert Rupert puts the point, the reduction of individual representations and collective representations differs with regard to our understanding of 'inter-level relations'. While we have little idea how to reduce psychological regularities to neurological regularities, we have a clear understanding of how to reduce collective representations to the mental and physical states that underwrite them—even if this should prove to be a difficult task.

But this suggests that the current status of our scientific knowledge is all that prevents us from eliminating individual representations. In analyzing collective representations, we know how to look for the individual representations involved in the production and interpretation of collective representations. However, even our best neuroscience isn't

developed enough to identify the physical states on which any individual representation is realized. However, it's not clear that this is a difference that makes a difference. As our understanding of neurophysiology and its relation to other physical explanations becomes more refined, there's reason to suppose that a coherent story about the realization of individual representations on physical states could become apparent. However, if all we need is a clear understanding of how to reduce individual representations to the physical states on which they are realized, the explanation of all behavior will eventually be specifiable in purely physical terms. When this happens, individual representations will become just as superfluous as their collective counterparts—and this result seems fairly unpalatable.

But, I've moved too quickly, there's a deeper problem here. While individual representations are realized on physical processes, collective representations are realized on individual representational states. Nothing new *in kind* is introduced in moving from individual representations to collective representations, the relevant states all have semantic content. However, in moving from the physical to the intentional, something new in kind *is* introduced. Mental representations have semantic content, the physical states on which they're realized don't. The crossing of explanatory levels is significant in the case of individual mental representations precisely because intentional states have semantic content. So, individual representations can't be made superfluous by scientific discovery—we need them to explain semantically evaluable states of the world. However, every theory of individual representation allows for an explanation of how to move from individual representations to other sorts of derivative representations—even by way of rules of aggregation—all from within the realm of intentional explanations. It's this possibility that

underwrites the most serious argument for the superfluity of collective mental representation—and it's this possibility to which I turn in the next chapter.

CHAPTER IV:

COLLECTIVE REPRESENTATION?

At the end of the 18th century, Wolfgang von Kempelen constructed a chess-playing automaton that he called The Turk. The Turk consisted of human-sized mannequin (with a black beard, gray eyes, Turkish robes and a turban) and a large wooden cabinet housing an incredibly complex set of gears that operated this mannequin. In order to show that the system was completely mechanical, the cabinet doors could be opened, revealing a complex mechanical structure in such a way that you could look straight through the machine. Now, despite being a completely mechanical structure, The Turk was a pretty good chess player—though it did lose some games. It played and won numerous matches against proficient chess players; it even beat Napoleon Bonaparte and Benjamin Franklin. But The Turk was no Deep Blue. The clockwork mechanisms housed inside the cabinet didn't produce the moves on the chessboard. Instead, a person hidden inside the machine operated these mechanisms. The Turk was a complex hoax designed to appear to all inspection to be an automata, but it was nothing but a tool to be used by a person hiding inside the cabinet.

In The Turk, we find a system that (at least on initial inspection) was behaviorally indistinguishable from a chess player. However, the states of the Turk that appear to be psychological states were, one and all, derived from the psychological states of the individual inside the machine. The machine itself had no mental states. The individual inside the Turk had mental states and that individual's intentions fully determined the behavior of the system.

Now, unless we include both the person who is operating the clockwork and the clockwork itself as part of a single cognitive system, it seems unreasonable to say that The Turk has mental states. The task of this chapter is to figure out what distinguishes apparently cognitive systems from genuinely cognitive systems.

4.1 Intentionality as the mark of the Mental

Perhaps the right way to approach this question is by determining what feature of a system marks it as a cognitive system.[74] According to many philosophers these days, a plausible mark of the mental is intentionality. Mental states are purportedly internal states of a system that are meaningful because of their intrinsic representational or contentful structure. I agree that it's at least a minimal condition on a state counting as a mental state that it is *about* something, that it is able to represent the world as being some way or another, or that it refers to something.[75] I have beliefs that are *about* the taste of my coffee, they *represent* my coffee *as* being delicious, and they *refer to* my cup of coffee. However, my cup of coffee is not about anything. It's just a cup of coffee!

However, as Dennett (1978a, 1987a) puts the point, commonsense ascriptions of intentionality are a motley assortment of genuine intentional attributions, metaphors, façons de parler, and countless varieties of clearly dubious intentional attributions. Commonsense

[74] Some philosophers (e.g., D. Rosenthal 1986; Block 1986, 1995) have argued that there are different marks for different sorts of mental states. You might think, for example, that things like beliefs and desires are mental insofar as they exhibit intentionality, but that qualitative states are mental only insofar as they exhibit phenomenal character. I don't want to get into these debates here. So, while I find representationalism (cf., Lycan 1987, 1996; Tye 1997) about qualitative character compelling, If you're unwilling to accept that qualitative states are intentional, you can feel free to take the subsequent discussion to be only about those states that lack qualitative character.

[75] I agree with Lycan and Tye that all of our mental states are representational—even states that refer to the whole person like elation and depression. With these sorts of states, the representations just represent the whole world as being a particular way rather than representing some feature of the world as being that way. I'll not go into the arguments for this position here, but cf., Lycan 2001.

psychology uses ascribes intentionality in describing systems even where we immediately recognize that the system doesn't have cognitive states. We talk as though a computer could try to prevent me from finishing my thesis even though it doesn't have any states that are directed at *me* at all. We also talk as-if vehicles think that it's too cold outside, or as-if the trees believe that it's spring. Keeping this in mind, it is necessary to distinguish between genuine intentionality and as-if intentionality; that is, we need an account of intentionality that distinguishes between cases where a system actually has representational capacities and cases where we merely speak as-if it has these capabilities. But, even this isn't enough.

Some representations are not representational *in the right way* to count as mental states. Kalitozov's *Ya Kuba* represents Batista Cuba and both Pontecorvo's *La Battaglia di Algeri* and Fanon's *Wretched of the earth* contain representations of the bloodiest revolution in contemporary history, but films and books do not represent things in the same way that mental states do. Thus, a distinction is typically drawn between derived intentional content (i.e., intentional content that is assigned by and is dependent on the contentful states of another system) and the underived intentional content of genuinely cognitive states (i.e., intentional content that arises from conditions that are themselves free of intentional content). Keeping this distinction in mind, a rich tradition in the philosophy of mind and psychology has suggested that mental states can be picked out as the only sorts of states that have underived intentionality.[76]

[76] Dennett (1990) has argued that there is no such thing as underived intentional content. He claims that it's interpretation all the way down. On this view of content, mental states of collectivities come far too easily. For Dennett, so long as we need to adopt the intentional stance in order to interpret and explain the behavior of a collectivity, we would be warranted in taking it to have contentful states. Now, although taking this option is tempting and would make my life much easier, the case for collective mentality shouldn't come so easily—it seems to strange to too many people to rest on so feeble a foundation.

At the foundations of this tradition, Franz Brentano argued that all mental states, and indeed only mental states, exhibit the capacity for being directed upon an object in an unmediated way.[77] Moreover, Brentano argues,

> intentional inexistence is characteristic exclusively of mental phenomena. No physical phenomenon exhibits anything like it. We can, therefore, define mental phenomena by saying that they are those phenomena which contain an object intentionally within themselves. (Brentano 1995, 89)

Following Brentano, Roderick Chisholm (1957) argues that an intentional state neither implies nor denies the existence of the object that it represents.[78] Chisholm argues that the capacity for exhibiting intentional inexistence plays the prominent role in psychological explanation. We explain the behavior of intentional systems in terms of the intentional content of their representations; and because these explanations are intentional, they take on a very different character than run-of-the-mill physical explanations. Psychological explanations need to be spelled out in terms that acknowledge these states opacity—or to put the point another way, psychological generalizations exhibit failures of substitution of identicals.[79]

---

[77] Brentano puts the point as follows: "Every mental phenomenon is characterised by what the Scholastics of the Middle Ages called the intentional (or mental) inexistence of an object, and what we might call, though not wholly unambiguously, reference to a content, direction towards an object (which is not to be understood here as meaning a thing) or immanent objectivity. Every mental phenomenon includes something as object within itself, although they do not all do so in the same way. In presentation, something is presented, in judgement something is affirmed or denied, in love loved, in hate hated, in desire desired and so on. (Brentano 1874/1995, 88)

[78] Thus, Diogenes the Cynic carried around a lantern searching for an honest man—and he was able to do so even thought there were no honest men for him to find. Moreover, Ponce de Leon searched for the Fountain of Youth even though it doesn't exist. The reason that this is possible is that the beliefs and desires that underwrite these intentional states have the same content whether they represent some feature of the world, misrepresent that feature of the world, or represent something that doesn't exist.

[79] In offering a psychological explanation of Susanne's recent purchase of Willie Nelson CDs, we advert to her belief that Willie Nelson is a great songwriter. In offering such a psychological explanation we ascribe to her a representation with the intentional content *Willie Nelson* (i.e., a representation that is directed at, or refers to Willie Nelson). However, the important thing to notice is that she does not thereby possess all possible representations that are directed at, or refer to Willie Nelson. After all, unless she has been obsessing about Willie Nelson lately, or unless she has been doing research on country music for a thesis, she probably doesn't

To put the point succinctly, things other than mental states can be intentional; however, things other than mental states can only have derived intentionality. Granny can have intentional states, particularly beliefs about the three 'Xs' and the 'skull-and-crossbones' that she's printed on the labels for the moonshine she's been distilling this month, but the three 'Xs' and the 'skull-and-crossbones' only get their meaning (perhaps that this is some *potent* moonshine that'll go down rough but be good for your dime) by way of what Granny (or a suitably situated interpreter) takes them to mean. But the question is: what does all of this mean for collective mentality?

4.2 Weber's objection to collective mentality:

At the end of the 19th century, Emile Durkheim argued by way of a methodology of functional analysis that there were irreducibly collective representations that would remain stable across variations in the individuals composing a collectivity.[80] Durkheim claimed that collective representations consisted "of manners of acting, thinking, and feeling external to the individual, which are invested with a coercive power by virtue of which they exercise

---

have the belief that THE PERSON WHO WON THE GRAMMY FOR THE BEST MALE COUNTRY VOCAL PERFORMANCE IN 1975 IS A GREAT SONGWRITER. Whereas facts about her WILLIE NELSON beliefs can explain her purchase of Willie Nelson CDs, facts about her PERSON WHO WON THE GRAMMY FOR THE BEST MALE COUNTRY VOCAL PERFORMANCE IN 1975 beliefs are explanatorily impotent. She doesn't have any such belief; so appealing to such a representation is completely misguided insofar as psychological explanation is concerned. Though such dubious attributions are hard to detect, appealing to a representation that Susanne doesn't possess is just as problematic as attributing to the trees the belief that it is spring on the basis of the appearance of blossoms in early January. Things are, however, quite different with physical explanation. If Susanne gets aggravated and throws a beer bottle at Willie Nelson, she *thereby* throws a beer bottle at the person who won the Grammy for the Best Male Country Vocal Performance in 1975—regardless of what she knows about the person at whom she's throwing the beer bottle.

[80] Consider the analogy with functional biology. As a quick look through the phylogenetic tree shows, there are many ways to build a wing (genetically speaking). But, for many purposes, the underlying genetic properties of each of the particular sorts wings are practically irrelevant. If we want to do functional biology, we look to the generalizations we can make on the basis of the functional properties of being a wing. For example, if a critter has wings it will, ceteris paribus, have the capacity to fly. Moreover, if something is a wing, it will probably act as a cooling system for the critter in question—but that's another point for another thesis.

control over him" (Durkheim 1895/1982, 52). With these representations in mind, he suggested that sociological analysis should focus on the collection of data about the prominent habits, legal and moral rules, popular sayings, and facts about social structure for various groups (Durkheim 1895/1982, 82); however, he realized that there are serious difficulties with collecting such data. After all, if you start with the observation of individuals, you find heterogeneity in the subjective interpretation of rules as well as heterogeneity in behavior. Fortunately, however, the emerging science of statistics provided Durkheim the tools he needed to analyze collective behavior.

Durkheim found historical trends in some collective behaviors that appeared to underwrite statistical regularities over birth rates, marriage trends, and suicide rates (Durkheim 1895/1982, 55). He took these statistical regularities to indicate social facts about the mental health of collectivities: higher suicide rates and lower birth rates in France as compared to England suggested that France was more depressed than England. Durkheim then noted that the statistical regularities he was finding were capable of withstanding numerous changes in the members of these collectivities. Even though people die, emigrate, immigrate, etc., the relevant statistical regularities remain relatively stable, Durkheim claimed that this was, at least in part, a result of the fact that when happy people from England and Germany move to France, they are constrained by the statistical regularities over birth rates, marriage trends, and suicide rates. If they were not, there should be wide fluctuations in statistical trends, and there are not. However, Durkheim himself recognized that the methodology of statistical analysis implied "no metaphysical conception, no speculation about the innermost depths of being" (Durkheim 1895/1982, 37). Statistical analysis merely tracks instrumentally useful claims about collective mental states, and it is on

this point that problems begin to arise. If Durkheim is right, then there may be collective psychological regularities; but these might not be enough to yield genuine collective mentality.

On one understanding, the task of psychology is the collection of descriptive data about the overt and covert behavioral dispositions of psychological systems, the systematization of this data, and finally the predictions and explanations of the behavior of psychological systems on the basis of this data. This view of psychology is, and I don't intend to use the term derisively, a behaviorist project; as a behaviorist project, psychology is not in the business of explaining the *cognitive* mechanisms that produce behavior.[81] The proposal that psychology should not attempt to explain the mechanisms that give rise to behavior seems to modern eyes to be a radically misguided project. Behaviorist psychology is useful for making predictions about a wide range of behaviors in humans and other animals; however, there is more to say about mentality—both the mentality of individuals and the mentality of collectivities.

The failings of the behaviorist project have lead contemporary psychologists and cognitive scientists to adopt a different sort of explanatory project.[82] We've come to realize that even the most mundane behavioral dispositions rest on attributions of intentional states. As Dennett (1978ba) notes, we've got reason to believe that even the mouse in the Skinner box has the desire for the food and the belief that if she pushes the bar she'll get the food.

[81] Unless, of course, you're inclined to think (cf., Skinner's radical behaviorism) that cognitive concepts are short hand for complicated behavioral analyses or something of the sort. One could go this way, but I see little reason to go in for either analyticities or behaviorism.

[82] Perhaps the most damning criticism of behaviorism was offered by Noam Chomsky (1959) in his review of Skinner's *Verbal behavior*. Chomsky argued that young children's verbal behavior is always underdetermined by the stimuli that they encounter prior to the lexical explosion that occurs between the ages of two and four. Chomsky claimed that without positing some sorts of internal representations, linguistic competence that outstrips linguistic performance would be inexplicable.

The mouse's overt behavior seems to be good (in fact, the best) evidence for the presence of such states, but these states don't seem to be identical to the behaviors. Moreover, numerous organisms utilize internal representations of various sorts and of various degrees of complexity in order to model non-present situations. Human organisms, for example, have the capacity to make choices on the basis of internal models of what's likely to happen if they make that choice. Perhaps more importantly, I can engage in revolutionary action on the basis of my representation of my society as founded on corrupt political principles. I can also consider the possibility of a genuinely democratic society. But, neither of these seems to be easily accounted for within the behaviorist project (though some fancy footwork might be suggested in defense of behaviorism). To put the point bluntly, hardly anyone wants to be a behaviorist these days, and I don't either.

I assume that everyone concerned with my project will be willing to accept the claim that individuals can be in at least some representational states that are not captured by the behaviorist project. On this assumption, the explanatory project that must be adopted by the psychological sciences has to be radically different from this a project of the behaviorist sort. On a cognitivist theory of mental states, the goal of psychology is to offer an explanation of the underlying causal mechanisms that produce behavior rather than merely cataloging a system's behavior. This project, however, opens up a series of questions about what sorts of mechanisms could possibly be in place to generate genuine cognitive processes. Durkheim's methods failed to give him any account of the mechanisms that give rise to genuinely intentional actions in collectivities—and it is on this point that Max Weber mounted his attack on Durkheim's collective psychology.

In *Economy and society*, Weber argues that the Durkheimian approach to social

science rests content with descriptive analyses of human behavior when it should be an attempt to provide an "interpretive understanding of social action and thereby with a causal explanation of its course and consequences" (Weber 1914/1968, 4). His arguments are grounded on a *verstehen* model for the methodology of the social sciences.[83] According to this model, we form an *understanding* of an action by ascribing to an agent the internal states that would make her behavior rational.[84] Weber argues that understanding subjective meaning only comes about in two ways. In some cases, we directly observe the agent's subjective meaning by attending to her linguistic behavior (cf., Weber 1914/1968, 8). As Elizabeth Anscombe (1957, 18) puts the point, the mental cause of an action is "what someone would describe if he were asked the particular question: what produced this action, thought, or feeling on your part." Unfortunately, looking to linguistic behavior isn't always an option. In cases where it's not, Weber argues, we gain an understanding of subjective meanings by engaging the beliefs and desires of an agent on her own terms. This requires an empathetic understanding of the most severe sort.[85] We have to be able to understand a

[83] I am grateful to Lindsey King for a social scientists view on the Weberian conception of *verstehen* models of sociological method.

[84] Suppose we come across a person, poised above a piece of paper with a pencil and ask: what is she doing? In one sense, merely describing her motions would be a perfectly adequate answer to the question. However, if we want to make sense of her behavior *as an intentional action*, we'll only have an explanation when we know that she is "engaged in balancing a ledger or in making a scientific demonstration, or is engaged in some other task of which this particular act would be an appropriate part" (Weber 1914/1968, 8). Unfortunately, the ways in which individuals makes sense their own behavior varies wildly—so just looking at physical descriptions of a behavior will always to underdetermine what she's doing. As Elizabeth Anscombe (1957, 37-41) famously pointed out, a single set of behaviors such as moving one's arm up and down while holding the handle of a water pump can variously be understood as the activity of pumping water to a house, poisoning communists, and making the world safe for democracy. On the basis of his concerns about the systematic underdetermination of intentional explanation by descriptions of behavior, Weber argued that we must discern the subjective meaning of an action for an agent if we are to offer an account of the underlying psychological cause of that action.

[85] Empathy here comes to English as a translation of the German *einfühlen*, which literally translated means 'to feel into'; in this case it means feeling your way into another's perspective. I claim that Weber requires a severe sort of empathy because he requires that you abandon your own perspective and adopt a strategy that allows you

person's mental states in terms of what she would take it as rational to believe or desire given the way her other beliefs and desires hang together *for her*. Weber holds that these two options are exhaustive for attributing subjective meaning to an agent, and thus that it is only by way of adopting one of these strategies that we can determine whether someone is acting rationally rather than merely behaving as though she were. On the basis of these considerations, Weber developed an objection to Durkheim's claims about collective representations.

Weber begins by noting that collectivities do not offer linguistic utterances that can justify or explain their behavior—so we can't opt for the first route. Moreover, he claims that it's impossible to empathetically engage the circumstances in which a collectivity finds itself. His worries here were grounded on something like Block's (1980a) worries about the possibility of there being anything that it's like to be a collectivity: there's just nothing it's like to be a collectivity and there's nothing that it's like for a collectivity to understand it's own mental states. So, the second route was closed off as well. Weber thus argues that "action in the sense of subjectively understandable orientation of behavior exists only as the behavior of one or more *individual* human beings" (Weber 1914/1968, 13 emphasis in the original). He conceded that for some purposes it might be convenient to treat collectivities as-if they had cognitive states (Weber 1914/1968, 13); however, such claims cannot be construed as literally true. Here's the rub: the only systems that have the capacity for rational action are those whose behavior can be explained in terms of internal mental states; but the only systems whose behavior can be explained in terms of internal mental states are individuals—hence, there is no such thing as collective mentality.

---

to interpret an action *from the agent's standpoint*. I have my doubts as to whether this is even possible, but this is not the place for a discussion of these reservations.

Although few philosophers these days are committed to the details of Weber's explanatory project, Weber is at least right to argue that any account of collective mentality will have to explain how a collectivity can possess genuinely representational states. And there are responses to Weber's worry that seem immediately appealing. The best option for responding to Weber's worries is to adopt Weber's fundamental commitment to continuity between the sort of explanation that is viable for individual mental states and the sort of explanation that is viable for collective mental states. In order to maintain this sort of continuity, I need to demonstrate that an adequate theory of individual representational states allows for representational states at the collective level as well. However, although this task is easily suggested, it's incredibly difficult to carry out. The adoption of any theory of representational content will prove contentious within the philosophy of mind. However, there are reasons for thinking that none of the comparatively viable theories of individual representational states will allow for representational states at the level of collectivity.

4.3 Rupert's contemporary Weberian argument

Robert Rupert (2005) argues that there is no plausible naturalistic theory of mental content that can serve as a legitimate foundation for collective mentality. I am inclined to believe that Rupert's (2005) arguments are the most troubling objection to collective mentality, and his arguments are all the more troubling given that he and I shares a number of important methodological assumptions about what it takes for something to count as a collective mental state. Rupert agrees that it is unacceptable to "simply assert the existence of genuinely autonomous group mental states—because, for example, that is what our everyday

talk presupposes—while claiming that they are not states of minds" (Rupert 2005, 2ms).[86] With Rupert, I agree that assuming radical discontinuity between individual and group mental states is the point on which many accounts of collective intentionality fail. After all, if you introduce enough discontinuity, the states of collectivities will fail to count as genuinely mental. Rupert (2005, 2ms) also contends that if something is going to count as a collective mental state, it must "instantiate the central features of minds as we know them best"; I also endorse a version of this claim. Rupert's position is an attempt to endorse our best philosophical and psychological theories of what a mind is. Finally, Rupert (2005, 2ms) argues that collective mentality requires distinctive causal-explanatory work for collective mental states; in the absence of such causal-explanatory work, attributions of collective mentality begin to look like mere instrumental shorthand for claims about individual psychological states in aggregation. With these assumptions in hand, Rupert argues that without some account of collective representations, we ought not read attributions of collective mentality as literally true *of the collectivity itself*. On this point, I'm also in agreement with Rupert. But this is where the trouble begins.

Focusing on the case of formally organized institutional systems such as labor unions, courts, and corporations, Rupert contends that there are no viable theories of individual representational states than can be applied "in a natural or convincing way to group states" (Rupert 2005, 8ms). In the remainder of this section, I'll briefly survey the naturalistic theories of mental representation that are on the table and the sorts of worries that Rupert thinks will arise in adopting each of them. The first theory that Rupert considers is indicator semantics. Proponents of indicator semantics (e.g., Dretske 1988) hold that:

[86] I've not yet considered this position as a possibility. However, in my discussion of Ron Giere's work on distributed cognition I'll return to offering some arguments against this claim.

> A mental representation MR represents a property P iff MR has an acquired function of indicating P for a system S, and it acquired this function because MR indicates P to S.

Rupert (2005, 10ms) argues that there are no states of a group that can stand in an *indication relation* to the properties that would have to be tracked by a mental representation. After all, indicator semantics typically rely on a relationship between internal perceptual states (which are later decoupled from immediate perceptual stimuli) and properties of the world; however, Rupert claims that there are no states of collectivities that seem to fill the relevant perceptual roles to allow for this sort of content fixing. If public language structures such as a press release or a court's opinion are supposed to play the role of representations, it's unclear how such representations could indicate anything on their own. In fact, such representations seem to be a paradigmatic case of a representation with non-natural meaning (Grice 1957): the meaning of these representations depends completely on the intentional states of the cognitive system that produces them.

Having argued that indicator semantics is unworkable as a theory of collective representation, Rupert considers the possibility of adopting a pure-informational theory of content in defense of collective mentality.[87] According to the most prominent pure-informational theory of mental representation (cf., Fodor 1990):

> A mental representation MR represents a property P iff 1) it's a law that 'P causes MR', 2) some P actually does cause MR, and 3) if something other than P causes MR doing so is asymmetrically dependent on 'P causing R'.

[87] Note that the term 'pure-informational' here is Rupert's choice and not mine. Using this term seems to suggest a view more like Dretske's (1988) since Dretske specifically spells out his theory in terms of the Shannon-Weaver theory of information.

However, Rupert (2005, 11) argues that "it is difficult to see why the proper causal relations would hold independently (or independently enough) of the asymmetric dependencies into which individuals' mental representations enter". If an informational theory of content is going to be adopted in defending collective mentality, a story will have to be told about how it is that some states of the collectivity *qua* states of the collectivity stand in the right sort of informational relation to some property. However, if public language structures such as a press release or a court's opinion are supposed to play the role of representations, it seems clear that informational relations only obtain between the individuals that produce these public language structures and the property that is supposed to be tracked by the relevant representation. In the absence of an account of how some property of the group *qua* property of the group that stands in the right sort of informational relation, it seems reasonable to think that the "group cognitive systems, *qua* group systems, contain no representations whose content is not derived from the content of representations in some other system—not what we should want in a genuine cognitive system (or mind)" (Rupert 2005, 11-12).

Rupert also argues that teleological theories of mental representation are insufficient to underwrite a theory of collective representation. Though there are a number of teleological theories of mental representation on the table, they all seem to hold that:

> A mental representation MR represents a property P iff some privileged relation between MR and P accounts for the continued reproduction of MR (cf., Rupert 2005, 12ms).[88]

There are, however, a couple of ways of spelling out the relevant relation, and Rupert contends that on any interpretation this theory is problematic as a theory of collective representation. If we adopt the evolutionary interpretation (cf., Millikan 1984), Rupert claims

[88] As Bill Lycan has pointed out in conversation, this understanding of teleological theories of representation rests on a standard misinterpretation of teleological theories of representation. As such, teleological theories of mental representation make no mention of reproduction, etiology, or anything of the sort.

that there is no straightforward way to make sense of the application of evolutionary theory to the most promising cases of collective mentality. After all, it doesn't seem as though the public language structures such as a press release or a court's opinion that are supposed to play the role of representations have much of any selectionist story to be told of them—there is "(1) nothing that varies in such a way that its differences might then be inherited by the group and (2) nothing to encode successful variations in fitness, so that they may be passed on to descendants" (Rupert 2005, 12ms). The proponent of collective representations might adopt a non-evolutionary theory of function. However, on these accounts, Rupert (2005, 13ms) argues that there is no account of the relevant causal mechanism for the sustenance of a function that don't simply appeal to the causal relations between the mental representations of individual and the properties in the world that they are supposed to represent.

Another option is to adopt a causal-historical semantics (cf., Prinz 2002). On this view, the content of a mental representation is specified partly in terms of cause and partly in terms of etiology, such that:

> A mental representation MR represents a property P iff MR is disposed to be reliably activated by encounters with P, and encounters with P played a role in the acquisition of MR.

Rupert (1998, 1999, and 2001) advances a similar view according to which the neural structures that realize a mental representation are shaped developmentally, by interaction with the environment. On his view, "this shaping involves a certain statistical pattern of interaction with the very things that thereby come to be represented" (Rupert 2005, 13ms). Here again, the proponent of collective representations faces the worry that collectivities possess no perceptual faculties to ground this sort of representational capacity. Moreover, on Prinz's view, mental representations are, to a large extent, best understood in terms of their

capacity to facilitate categorization, and although words in public languages have the capacity to represent, they must be interpreted in order to facilitate categorization because they are arbitrary symbols (cf., Prinz 2005) If there is nowhere to locate interactions between the mental representations and properties, and if the most promising sorts of collective representations don't seem to facilitate categorization, it looks like this theory is out as well.

Finally, Rupert (2005, 14ms) turns to what he calls 'teleo-isomorphic' theories that ground mental representation in an isomorphism between the structural properties of the representation and the thing represented (cf., Cummins 1996). On this view:

> A mental representation MR represents a property P iff the structure of MR is isomorphic to the structure of P and it is the function of the portion of the cognitive system in which MR occurs to represent P.

Again, according to Rupert, the most promising way to utilize this theory in defending collective representation would be to look of the relevant sort of isomorphism between public language structures such as a press release or a court's opinion that are supposed to play the role of representations and the things that are meant to be represented. However, although

> The ink marks that constitute a court's decision are, taken as an entire structure, isomorphic to some other structures, but we have no reason to think that the abstract structures the decision is about—abstract conceptual structures, typically—will be among those things structurally mirrored by the arrangement of ink on page. (Rupert 2005, 15), 3)

Moreover, the sorts of worries that arise in the case of teleological theories of mental representation once again arise in this case. After all, even if we find the right sorts of functions in place in a collectivity, it seems as though "our best account of why those parts have those functions adverts in a straightforward way to the mental states of individuals" (Rupert 2005, 15-16ms).

Having canvassed all of the most promising accounts of mental representation, and having argued that none of these are applicable in the case of collective representation, Rupert claims that there is a significant hurdle faced by any proponent of collective mentality. If there is to be a viable account of collective mentality, then there will have to be some account of how a collectivity could have the right sorts of representational states. Rupert, much like Weber, contends that there is no obvious way to apply the tools of individual psychology to collective states; Rupert thus voices his skepticism about collective mentality.

4.4 Rethinking Individual representation

There are at least two ways in which one could respond to this criticism of collective representation. First, it could be argued that one of the standard theories of mental representation can be extended, with little or no modification, to apply to collectivities.[89] Second, it could be argued that parity of reasoning requires that whichever theory of representation one happens to adopt, if Rupert's arguments rule out collective representation they will also rule out individual representations. I adopt the latter strategy, because adopting a particular theory of mental representation would suggest far too strong of a link between the truth of that theory and the possibility of collective mentality—and the plausibility of collective mentality is not contingent on the adoption of any particular theory of psychosemantics.[90] Given that Rupert discusses three main types of relations between

[89] I am inclined to think that adopting this sort of strategy is completely viable. After all, I think that there is room within each of these dominant theories for developing a theory of collective representation. However, I think that the second strategy is a better option.

[90] Some theories do, however, make it much easier. If Dennettian interpretivism is true, then we get group minds almost automatically. I don't want to prejudice the case here in favor of this, or any other, theory of psychosemantics.

representations and the things that they represent—perceptual relations, informational relations, and teleological relations—I'll address each of these considerations in turn.

4.4.1 Perceptually based theories of mental representation:

Rupert claims that some theories of mental representation fail to apply to groups because groups lack of collective perceptual faculties. Indicator based theories typically rely on internal perceptual states which can later be decoupled from immediate perceptual stimuli. Causal-historical views also rely on perceptual capacities in order to sustain the causal relations that give rise to mental representations. However, there are many ways in which this reliance can be spelled out. As Prinz (2005, 686) puts the point:

> To say that concepts are perceptually based is to say that they are made up from representations that are indigenous to the senses. Concepts are not couched in an amodal code. Their features are visual, auditory, olfactory, motoric, and so on. They are multimedia presentations.

In order to make sense of the reliance on perceptual states, it will help to start with a familiar paradigm case of a representation. However, by looking carefully at something that clearly counts as mental representation, a number of interesting things that typically remain unnoticed about this reliance become more obvious. I think that it's fair to say that a perceptual representation of one's mother relies on a number of important perceptual capabilities, and I'm inclined to think that most of our other mental representations rely on perceptual states in the same way.

Suppose Amanda sees her mother standing in the doorway of the bar having a conversation with the bouncer, and turns to the person sitting next to her and says, "That's my mother". In this case, Amanda has a perceptual representation of the person standing in

the doorway, and, supposing that the lighting conditions are fairly decent and that Amanda can is capable of seeing that this is her mother, we'll have good reason to take this verbal behavior as evidence that she is mentally representing this person as her mother. However, merely noting that there is such a representation is insufficient to explain what is going on in Amanda's cognitive architecture.

We need to begin by noting that Amanda has looked across the room and perceived someone that she is capable of categorizing as her mother. That is, when she is presented with some sorts of perceptual input, she categorizes this thing as her mother. Thus, whatever else we want to say about Amanda's representational state, we have to note that it has the function of categorization. Moreover, given that she has this capacity to categorize the thing she perceives as her mother, we find that there are typical behavioral responses that are evoked by this thought. Her verbal behavior, the change in her heart rate, and the tendency to run across the room and hug this person, are all characteristic outputs of her representation of this person as her mother. The thing to notice here is that our explanation of Amanda's behavior in terms of her belief that this is her mother relies on the use of a theoretical posit about her internal state. The question, now, is what sort of internal state is this. This is where things start to get a little bit tricky.

When we think about familiar sorts of representational states like this perceptual representation, it's hard to see them as more than simple, homogenous lexical items—perhaps linguistic representations in a language of thought. For this reason, there's been a tendency among philosophers to take a representation like Amanda's representation of her mother and claim that this is just a state built up from a MOTHER concept, some sort of demonstrative concept (to get the 'that is' into the picture), and some sort of possessive

concept. And, at one level of analysis, this is the right way of thinking about things. However, in cases where things seem so familiar and so obvious, it often helps to think about what a breakdown of such a representational capacity would look like. Consider a person with Capgras syndrome. When a person suffers from this delusion, "they are mentally lucid, their memory is normal, and [most] aspects of their visual perception are completely unaffected" (Ramachandran 1998b, 1856). They seem fairly intact. Aside, that is, from the fact that they have an unshakable commitment to the claim that someone to whom they are quite close (typically a parent, a spouse, or even a pet) has been replaced by an imposter, a robot, or an evil twin.

Suppose that Amanda has stroke or a drug overdose and awakens in the hospital seeming perfectly normal—until her mother walks in. At this point, suppose that she exhibits the behavior characteristic of a person suffering from Capgras delusion and reports that the person standing in the doorway is not her mother, but a cleverly disguised CIA agent who has been sent to monitor her. Now, it is no longer true that Amanda believes that this is her mother, and the fact that her representation has changed suggests that there is something different in her representational capacities. The most widely accepted theory of the mechanism underwriting Capgras delusion suggests a failure in the binding of visual representations and the affective representations that drive a feeling of familiarity.[91] This sort of breakdown suggests that the mental representation "That's my mother" is actually quite a

[91] There is some dispute over the precise nature of the mechanism here. V.S. Ramachandran (1998, 2004) argues that the relevant damage is to the structures linking the amygdalla and the inferotemporal cortex preventing the processing of affective information. Young et al. (1993) propose a similar sort of mechanism, at least in so far as they are concerned to demonstrate that this is a localized breakdown in binding affective information to face perception. However, they claim that the breakdown should be understood as a disconnection between the dorsal and ventral streams in the visual system. Regardless of what the relevant neurological mechanism happens to be, however, what matters for my case is that the representation of someone as a person's mother rests on information that is not exhaustively specified in terms of the mechanisms that represent linguistic structure.

bit more complicated that one might have assumed prior to examining this sort of case. In order to have a perceptual representation of a person *as one's mother*, it needs to be the case not just that the visual system is functioning properly, but it also has to be the case that the affective response is correct.

Alternatively, suppose that Amanda awakens with localized damage to her fusiform gyrus. In this case, Amanda might continue to represent someone as her mother, but she might be incapable of doing so by representing her face. She might continue to track her mother by the sound of her voice despite the fact that she has become prosopagnosic and can no longer perceive faces as such. The interesting thing about this case, however, is that the affective response may continue even though she is failing, on any sort of conscious level, to represent this person as her mother (cf., Bauer 1984). In this case, the feeling of familiarity may persist even though the she might no longer have the visual representation of this person as her mother.

The important thing to notice, however, is that the representation of a person *as one's mother* relies on a number of component processes. Many of the representations we deploy in navigating our world are composed out of the outputs of the various subroutines that are operative in that individual. To put the point another way, many of the representations that we take to be genuinely *mental* representations, representations with primary intentionality, supervene on component structures that are themselves already intentional. This, I take it, is the primary insight of the homuncular functionalism that underwrites my account of both individual and collective mentality.

Many naturalistically plausible theories of mental representation *have* focused on the causal *cum* perceptual relations that obtain between a mental representation and a property of

the world. This makes it appear as though mental representations, such as a MOTHER representation, ought to be understood as simple tokens in the language of thought. Now, although there are likely to be some perceptual states of a cognitive system that will stand in the right sorts of causal relations, it's not clear that all of the relevant states of a system have to stand in these relations unmediated. Rupert (2005) argues that "many of our best explanations of how mental representations get their content assign a privileged role to perceptual or quasi-perceptual processing, thereby requiring a cognitive architecture that group systems typically do not possess." However, this claim relies on too simplistic a view of the inter-level relations in individual cognition. Something like the following is likely true.

> In order to a mental representation MR to represent a property P: 1) encounters with P have to play some causal role in the acquisition of MR, and 2) MR has to have the function of representing P to the system in which it is a mental representation.

However, while some nomic relations certainly obtain between perceptual states and properties of the world (e.g., in edge detection, color detection, phoneme detection), most person level representations *derive* their representational content from lower-level states that are themselves already semantically contentful.

This fact suggests a powerful argument from parity against Rupert. If the explanation of Capgras syndrome discussed above is approximately correct, a visual representation of one's mother can be fully explained in terms of 1) the properties of discrete and static representations in the visual system (construed rather broadly), 2) affective responses to this stimuli that are feelings of similarity, and 3) rules for the association of visual and affective representations. However, Rupert's superfluity argument would suggest that there's no need to posit a person-level representation of MOTHER since every step in the construction of a representation of MOTHER proceeds either by physical causation (e.g., the stimulation of

retinal cells by photons reflected from the stimuli) or by causal processes involving the intentional states of the subcomponents of the visual system and rules for the association thereof. Here the states are all representational, as they are in the reduction of collective to individual representations. So, if superfluity arguments preclude collective representations they preclude individual representations of mothers, and as Fodor would put it, if a theory can't allow for the representation of one's mother, it's the end of the world!

4.4.2 Informational theories of mental representation

What, then, of theories of mental representation that rest on informational relations that do not rely on perceptual states as such, but instead rely on counterfactually stable causal relations between a cognitive system and its environment? Rupert contends that it's hard to see why the relevant sorts of causal relations would obtain in virtue of states of the collectivity *qua* states of the collectivity rather than in virtue of states of the member of the collectivities *qua* members of the collectivities. At least initially it seems as though responding by way of an informational theory like Fodor's should be easy. After all, according to Fodor there would only need to be some state of a group that could stand in the right sort of actual and counterfactual causal relations to some state of the world. As Fodor understands mental representation, in order for some mental representation "X" to be a representation of a property X it must meet three conditions:

1) It has to be a law that X causes "X";
2) Some "X"s have to be caused by X; and
3) If anything other than X causes "X", it's causing "X" must be asymmetrically dependent on Xs causing "X"

So, all that would need to be the case in order for there to be a collective mental

representation is that there would have to be some state of a collectivity that stood in these sorts of causal relations to some property of the world. Fortunately, there are cases of collective representation that meet all of these conditions in the actual world. Consider the case of Naval vessel navigation discussed by Edwin Hutchins (1995).

Hutchins takes as his primary example of a collective representation the navigational fix cycle. The fix cycle is used to establish the location of a ship in relation to various sorts of landmarks in order to facilitate a computation of the trajectory of the ship (Hutchins 1995, 117). The interesting thing about the fix cycle is that it is the implementation of a computation that contains a number of processes, some of which are internal to persons and some of which are external to persons. As Hutchins (1995, 117 *emphasis in the original*) puts the point, "the fix cycle is accomplished by the *propagation of representational state* across a series of *representational* media". Briefly, the representation of the ship's location is produced through the interaction and association of a number of different lower level processes—each of which is already in the business of producing representations. Although the media that are produced by each of the relevant subroutines vary wildly, they are nonetheless capable of being brought into coordination with one another in order to give rise to a representation that can direct the behavior of the ship.

The navigation system of a ship consists of a number of systems that are designed to be sensitive to a variety of one-dimensional constraints in the world (Hutchins 1995, 118). The output of each of these systems is propagated across a number of media until the fix cycle produces a representation of the location of the ship on a chart. None of these various sub-systems (e.g., neither the alidade user, the hoey, the chart, nor the fathometer) is capable of producing an authoritative representation of the location of the ship. Instead, it is only by

bringing these various representations into coordination—often by way of taking repeated measurements—that a representation of a ships location is produced by the navigation system. That is, it is only by way of the coordinated activity of a variety of systems that the location of the ship can be determined and a representation of this location can be produced in such a way that it is usable for setting a course for the ship. Moreover, because of the way in which training occurs in the US Navy, the representations produced by the various individuals in the crew are typically only capable of being understood by those who are trained to take measurements using a particular device. The persons working on a particular task take *as inputs* the information (here we have the production of an analog representation) produced by some technology or the information they receive in a visual representation of the ships location from the bow. They then engage in some sort of computation in order to produce a representation that can be read by someone else. They then output a digital representation that can be read by another system and that will eventually be capable of being coordinated with other sorts of information. Notice, none of the individuals in the navigation crew represents the position of the ship. It's only the navigation crew as a whole that represents the location of the ship.

Here is how this sort of system can be used to demonstrate the possibility of an informational theory of collective representation. The fix cycle is capable of varying lawfully with the location of the ship in the same way that person-level representations are supposed to vary with features of the world.[92] If this were not the case, then there would be a whole lot

[92] Human operators in association with their machines produce the constitutive representations, and it is no law of nature that people don't make mistakes; so, it's quite hard to see how nomic relations between representation and representatum could be established. However, it's not at all clear that you get anything like strong nomic relations between neural states and features of the world. That said, my inclination is to take the relevant relations to hold *ceteris paribus*, and I would probably fill in the *ceteris paribus* clause with some functional claim concerning the proper operating of the system. This might be to abandon such theories of content. However, if that's the case, it's going to be true both for individuals and collectivities. And I say: So be it!

more chaos on the sea, more ships would be lost, and there would be a lot more ships running aground. This allows the fix cycle to meet the first of Fodor's criteria. Moreover, the fix cycle of a given ship must at least some times be a representation that's caused by the actual location of the ship—for exactly the same reasons. This allows the fix cycle to meet the second of Fodor's criteria. Finally, it is very likely that, on at least one occasions, a fix cycle might produce a representation that fails to accurately map the location of a ship. However, its doing so will have to depend asymmetrically on accurate representations of location. Were it not the case that the deliverance of the fix cycle was the sort of system that delivered accurate representations except where there was some failure of the informational channels in the system, it would not be a representational structure that was capable of representing the location of the ship rather than just recording it in a way that accidentally happened to covary with some state of the world. The fix cycle thus meets the third of Fodor's criteria. It looks, then, as though the fix cycle can represent the location of the ship. Why, then, would anyone be worried about the capacity of a collectivity to stand in the right sorts of causal relationships to things in the world?

Perhaps the fact that collective systems are widely distributed systems should itself be seen as a problem. Although collectivities are interconnected in important ways, they are also spatially distributed. Since there's no unified consciousness that controls individual bodies through telepathic mind-control (as the Overmind does in Arthur C. Clarke's (1953/2001) *Childhood's end*), the distributed computations which would have to take place in collective cognition will require information to be passed between distributed component systems. However, if the information from such systems will be useable to form unified intentions in a way that will allow for genuinely intentional action that's responsive to the ever-changing

world in which we find ourselves, it seems that it will have to be the case that there is some person who, in the end, produces the final representation on the basis of her beliefs about the location of the ship. If this is true, it looks like *she* is doing the representation of the location of the ship rather than the navigation crew as a whole. Or so the objection goes.

There are, however, a number of things to be said in response to this objection. First, there are numerous cases in which a thing whose constituent parts are spatially (and even temporally) distributed seems to count as a single entity with stable behavioral dispositions (e.g., flocks of seagulls, the New Zealand All Blacks, Bank of America, the Black Panthers, and the British Navy (cf., Bloom and Kelemen 1995)). As Dennett (1989) puts the point, what is particularly striking about termite colonies, is that they are examples of complex systems that are capable of functioning in a "purposeful and integrated" way simply in virtue of having lots of subsystems doing their own thing without any central supervision. And, as Mitchel Resnick (1997) suggests, many systems that appear to have central controllers (and are usefully described as having them) do not. Moreover, as we've already seen, a person's visual system is spatially distributed throughout her brain and across a number of different systems (e.g., the eye, the optic nerve, the lateral geniculate nucleus, primary visual cortex, the prefrontal cortex, the fusiform gyrus, etc.) yet we are not concerned about calling the visual system a single system. On the whole, individual cognitive systems aren't all that different in this regard from collectivities. While the individual neurons of any particular cognitive system will be interconnected in important ways, they will also be spatially distributed. Moreover, if homuncular functionalism is the right view of the mind, individual mentality will be distributed across different systems because the computations that are required in order to give rise to intentional action are far too complex for any of the

individual systems to execute on their own.

Does this mean that there is no objection to be made on the basis of distribution? Of course not. However, if you are willing to throw out spatially distributed systems merely on the basis of their spatial distribution, many other things are going to have to go as well. I am inclined to think, however, that at the point where an individuals visual system becomes problematic *as a case of a cognitive system*, something has gone radically wrong. If all of human mentality falls away because of spatial distribution, we'll have far bigger worries about the possibility of doing cognitive science than worrying about whether collectivities can have mental states. Unfortunately, however, there is a further objection lurking in the wings.

This more promising objection is grounded on worries about the limitations on the flow of information through a distributed system. The most promising version of this argument is based on an objection to centralized decision making in large-scale economic systems offered by Friedrich Hayek in "The use of knowledge in society". Hayek (1945) recognized that one of the key problems facing any economic order is a worry about how to utilize the highly dispersed, incomplete, and often contradictory 'data' possessed by various individuals in a society in order to produce rational economic activity. In considering answers to this worry, Hayek (1945) argues that decision-making is possible only in cases where there is some single individual that actually makes a decision.

Here is one way in which this objection might be developed. In any society where people are engaging in collaborative activities, planning will rest, at least to a large extent, on information that has not been gathered by the person who will in the end make the decision to execute the plan. Instead information will be collected by a number of individuals, each of

whom will have a unique perspective on the information that she's collecting. The problem is that when information is passed to the planner, it will take a variety of different forms depending on the context in which the information was gathered as well as facts about the psychology of the person who is gathering the information. In some cases, this even leads to contradictory information being gathered by the planner. Furthermore, the more distributed the system is, the more likely it will be that there will be huge differences in the information that is collected since much of the information that will necessary for collective action will be highly dependent on the immediate circumstances at hand and the more distributed the system is the more differences there will be in immediate circumstances. The problem is that the planner will now either have to make a decision about which of this information she will pay attention to, or she will have to make a decision of her own about what to do—in which case the decision will be based exclusively on her preferences and not on the preferences of the people who have collected the information. This seems to suggest that the decision-making that underwrites a collective action is really just individual decision making of a planner embedded in complex social circumstances.

In the case of the economic decisions, the continuous flow of goods and services that is required to maintain a functioning economy requires continuous deliberative adjustment. However, in cases where quick action is required because of changing circumstances, as in the case of response to economic problems of various sorts, the filtering of information by a central planner will be far too slow to effectively respond to changes in circumstance.[93] Note, however, that the problem is not merely a problem with the distribution of the informational

[93] This sort of argument is far and away the most compelling argument against the centralized socialism of the former Soviet Socialist Republics as well as the Eastern European countries that adopted, or were forced to adopt, Soviet economic policies. The argument also, however, cuts against any form of centralized economy including state capitalism. If Hayek is right, the only option is a radically decentralized political apparatus.

content across a large system, but a worry that the sort of informational content that will be relevant to making a particular decision will change as the circumstances with which the system is faced change. So, if there is a centralized decision making system, it will have to be able not only to monitor the diversity of information coming into the central system, but it will also have to send out requests for the right sort of information at the right time—but this won't be possible without interpreting the information coming in from the various sources, which will always be out of date and incomplete. As Hayek puts the key point:

> The problem which we meet here is by no means peculiar to economics but arises in connection with nearly all truly social phenomena, with language and with most of our cultural inheritance, and constitutes really the central theoretical problem of all social science.

For any highly distributed system to which we might want to attribute collective mentality, there will be a difficulty with the diversity of information that will be prohibitive of rapid action in response to changes in circumstances for the system in question. So, if there has to be a central control system for collective decision-making, this will lead in the end either to an individual decision made by one member of the collectivity (in which case, it would not be a collective mental act), or it will lead no deliberative activity at all on the part of the collectivity (since the circumstances in which the system finds itself will change much too quickly for the system to respond). Neither alternative looks to be a good result for the mentality of collectivities.

This objection is quite compelling. However, rather than pushing us away from collective mentality, this argument actually suggests a key point for the defense of collective mentality. To begin to answer the objection, we must note that it presupposes that cognition takes place in some sort of centralized processing system. However, we have very good reasons to think that even human cognition isn't centralized in this way. In order to make

sense of the sort of architecture that we should expect for a collectivity, then, we need to stop and think briefly about the way in which human cognitive architecture happens to be organized. However, Daniel Dennett and Marcel Kinsbourne (1991) have offered compelling arguments for the claim that the human capacity for consciousness requires nothing like a central observer. While it is true that the brain has to be able to bind things together in some way, there is no need to suppose that this must happen in one place. In place of a centralized Cartesian subject that experiences things from inside a Cartesian theater, Dennett and Kinsbourne (1991) propose the multiple drafts model according to which conscious thought is accomplished by using multiple processes of interpretation and elaboration. On this view, each of the subsystems in the brain make localized and specialized observations that fix informational content. Each of these observations, then, reflects the state of the brain at the time of the observation. The question, however, is whether or not there must be a single process that unifies these informational states into a single narrative.

Dennett and Kinsbourne argue that localized discriminations should not be understood as states that are meant to be fed-forward for consideration by a central discriminator. Instead, they argue for an account of consciousness as *content sensitive settling* (Dennett and Kinsbourne 1991). Using the analogy of synching sound tracks to films, Dennett and Kinsbourne argue that temporal inferences, for example, are drawn by comparing the content of several data arrays. Moreover, they argue that once such a temporal ordering is drawn, it need not be drawn again by a higher-level discrimination. Supposing that something like this view of human consciousness is plausible, we see an immediate parallel to the sort of discriminations that give rise to a fix cycle in the navigation of a naval vessel. In a fix cycle, the location of a ship is determined by the synchronizing of a number

of low-level observations and the representation that is produced by the fix cycle does not need to be re-checked by a centralized observer. Instead, the captain is merely the sort of system that is capable of *consuming* this representation in order to fulfill his function of driving the ship. The important thing to notice here is that there is a big difference between taking the representation that is produced as an input for the production of further representations or behaviors, and having to make a set of new discriminations about this information. On the view I am advancing, the captain is merely another subroutine rather than a central system that is making the decisions.

There are, of course, those who would be unwilling to adopt the sort of model of consciousness advanced by Dennett and Kinsbourne. However we need not turn to conscious phenomena in order to make the point that I am suggesting here. Consider the development of distributed representational structures in behavior-based and autonomous robotics. Randall Beer (2000; Beer and Chiel1993) and his colleagues have developed autonomous robots that are capable of locomotion on a hexapodic platform. Instead of producing a robot with a centralized system for the production of particular sorts of behavior, Beer and his colleagues have built robots that consist of a number of localized discriminatory systems that rely on localized sensory feedback in order to determine what the next movement of the system will be. Although the various different systems have the capacity to receive information from one another as well as from each other, they do not have to wait on a decision to move from a centralized system. Instead, motion emerges from the complex interaction and coordination of the representations produced by various low level systems. But, there are further problems here.

It seems as though these sorts of structures, since they are merely detectors of various sorts, are going to be able to produce actions that only appear to be (or are as-if) intentional. We might think, for example that it is only by the introduction of already minded subroutines into a collectivity that we see the emergence of genuinely intentional action in collectivities. Once we realize that the sorts of data that will be pertinent to action of the system will be localized, as well as already intentional, a worry starts to emerge that the representations that happen to get used by a system are going to be under the control of the particular people that happen to be playing the relevant role in the collectivity. However, if this is the case, then we have to ask why we should think that there's any sort of determining role from the collectivity that will do the relevant work to constrain people in the right way.

On this point, however, it is important to note that there are commitments that one adopts when one joins collectivity. While it is true that a person within a collectivity will always have the capacity to reflect on each and every one of her actions, "an initial decision to identify with a collectivity will render it inappropriate, and perhaps even incoherent, thereafter to engage in deliberation over whether to identify on *every* occasion" (Graham 2002, 127). Part of what it means to play a role within a collectivity is to have at least some of your practical reasoning—in particular, your reasoning about whatever it is that the collectivity is meant to be doing—constrained by the commitment to act in accordance with the interests of the collectivity.[94] As Keith Graham puts the point, "collective identification involves *on appropriate occasions attempting to think and act as if for the collectivity itself*" (Graham 2002, 128). Now, this is not to say that each and every decision that a person makes

[94] Here's one way of putting the point. "To act as a member of the team is to act as a *component* of the team. It is to act on a concerted plan, doing one's allotted part in that plan without asking whether, taking other members' actions as given, one's own action is contributing toward the teams objective…It must be sufficient for each member of the team that the plan itself is designed to achieve the team's objective: the objective will be achieved if everyone follows the plan" (Sugden 1993, 86 cited in Graham 2002, 129).

in her role as a member of a collectivity will be perfectly consonant with the ends toward which that collectivity is directed. However, if a person fails to fulfill her role in a collectivity and instead decides to do as she would prefer to do, the intentional capacities of the collectivity will likely founder.[95]

Consider a case in which a member of a navigation crew decides to write things down as she wants to rather than as her training dictates. Or, even more problematically, suppose that she decides to take a nap instead of doing her job of producing a particular representation. In this case, there are a number of things that might happen. Given that there is a bit of redundancy in a navigation crew, and given that it will be possible to coordinate other representations in order to successfully produce a representation of the location of the ship, perhaps things will be fine. Alternatively, perhaps her failure to do the job will result in a misrepresentation of the location of the vessel—which might lead to a nasty run-in with an iceberg. Whichever way things go, the parity with individual mental states remains. When a subsystem begins to produce representations that are not consonant with what is expected, other systems have to compensate in order to successfully continue to represent the world. When this fails to occur, misrepresentations are often produced where we would find

---

[95] There are a number of points that need to be distinguished here. However, exactly how they are distinguished varies between collectivities. The relatively weak claim that an individual's deliberation will be constrained in a collectivity is always true. However, the degree of constraint varies. In small collectivities, deliberation often continues until every individual's decision is consonant with the decision that is adopted by the group: in a small society every decisions can be made while sitting around the fire. However, this is not always the case. Often, whether any particular individual does what she prefers (even where that is not consonant with the goals of the group) will be irrelevant to the practical activity of the collectivity. Provided sufficient redundancy within the functional organization of a group, the practical activity of a collectivity (much like the practical activity of a connectionist network) will exhibit a sort of graceful degradation when an individual fails to play her role. However, in cases where the failure of a collective action would result in extraordinarily bad consequences, and where success is dependent on every member of a collectivity doing her job (as with the case of pilotage discussed by Hutchins 1995) it's likely that structures of reward and punishment will ensure that each and every member of the group does exactly what she is supposed to be doing. After all, in such cases, individual failures could result in disaster, and that's enough to put serious normative constraints on the behavior of the each of the individuals that compose a collectivity.

successful representations in a fully functioning system. This seems to suggest that breakdowns in collective identification will have ramifications that are quite similar to the ramifications of breakdowns in individual cognitive systems.

There is, however, a deeper worry. Each of the representations that are produced by an individual will be produced in accordance with the way that she happens to understand the world. That is to say, her representations won't necessarily be veridical representations of the way that the world is. Instead, they will be intimately linked to whatever esoteric system of beliefs the producer happens to possess. This would be a serious problem, were in not also the case that our own sensory systems are not nearly so concerned with what's good for the whole organism as one might initially think.

Consider our own sensory systems. As Kathleen Akins (1996, 342) has put the point, the traditional view of the senses, in its strongest guise, takes a very solipsistic view of the brain. The brain is like a control center of the body, the place where all of the planning and thinking goes on. It learns about the world through the deliverances of the senses and it sends out motor commands so that the organism in which it is housed can respond to the various sorts of stimuli it encounters. The brain (or some part of it) thus becomes the centralized decision maker that is operating on the more or less veridical deliverances of the senses. However, when we consider the case of thermoception, Akins (1996, 345ff) argues an adequate model of perception is better understood in terms of narcissistic sensory systems, concerned only with how particular sorts of stimuli *affect them*. The question, then, is why is this not a problem for individual level cognition?

Suppose that you have hiked to the bottom of the Grand Canyon. It's 118 degrees Fahrenheit, it's hotter than hell, and all you can think to do is put your head under the lowly

patch of sagebrush that's providing the only shade around. Then you remember *this is a canyon* with a cool river at the bottom. You can jump into that river and cool off! Unfortunately, what you don't know is that the average water temperature of the Colorado River is 42 degrees Fahrenheit. You hurry to the river and jump in, only to find that it is *miserably* cold—every thermoceptor on your body is firing like crazy, especially the ones on your scalp! But after a few minutes, you adjust to the temperature of the river and it becomes quite pleasant.

The first thing to notice about this case is that the function of the thermoceptive system is best understood in terms of its ability to detect changes in temperature that will be relevant to the well-being of a particular organism. In order to carry out this function, cold receptors are integrated with various motor subroutines, and when these start firing they will make you shiver and bring your limbs in closer to your body (among other things) in order to preserve body heat. Moreover, the information that is passed to the motor subroutines needs to be quickly accessible in order to guarantee that the relevant behaviors occur. However, *pace* Hayek, this doesn't require that the information that is being passed along be veridical. In order for an organism to get along in the world, its behaviors don't even need to correspond in any direct way with to the actual state of things in the world. For biological systems like us, satisficing is good enough—and perhaps the best that we can do. All that needs to be the case for a sensory system to be adaptive is that in the majority of the cases where something (e.g., extreme changes in temperature) is a potential threat to the well-being of an organism, the perceptual system will pass on information to the motor subroutines that will protect the critter—and this requires far less than veridical information being passed to motor subroutines. Thermoception at least does this.

I, thus, claim that facts about our own cognitive architecture should lead us to expect that the sorts of problems with the flow of information predicted by Hayek's argument will not necessarily be any more problematic in the collective case than the individual case. There are, of course, real worries about how distributive systems are able to respond quickly and efficiently to rapidly changing stimuli. However, these worries are not insurmountable and they are no different at the level of the individual than they are at the level of the collectivity. These are just worries about how particular systems happen to be put together. I would contend that given this parity between individual level cognition and group level cognition, this objection shouldn't worry us too much.

We started with the datum that human organisms are able to respond to dangerous stimuli in a way that is consonant with quick intentional action. We find that there is good reason to think that the architecture of the human mind is widely distributed but it is still able to maintain the right sorts of lawful covariations with things in world in order to sustain a sort of informational theory on the basis of these distributed systems. We are able to get around in the world, and this give us good reason to think that we are able to utilize intentional representations in order to facilitate action guidance. The fact that collective systems such as the navigation crew of a large naval vessel suggest that the representational state that are being propagated across representational media in order to respond to ever changing, and often dangerous stimuli gives us good reason to think that these sorts of representational states are being used to guide action in relation to these stimuli. On the basis of such considerations, we should think that there are collectivities that can instantiate the right sorts of lawful relations required for mental representations if individuals are. Thus, informational theories are not impugned as theories of collective representation.

4.4.3 Teleological theories of mental representation

Rupert also argues that teleological theories of mental representation cannot be extended to collectivities. If we adopt an evolutionary interpretation of teleosemantics, Rupert claims that public language structures such a court decisions and memos, which seem to him to be the most promising place to look for collective representation, won't have a selectionist story to be told of them. Moreover, if we adopt a non-evolutionary interpretation of language, the relevant causal mechanisms for sustaining a function will rely on the causal relations between the mental representations of individual and the properties in the world that they are supposed to represent. I agree with both of these claims. However, the recognition that many of the collectivities with which we might be concerned have not themselves evolved in a way that will sustain the relevant sorts of representational architecture does not impugn the possibility of collective representation. Moreover, non-evolutionary theories of function allow for collective representation even though they rely on the representational capacities of individuals.

To begin with, note that there's a difference between recording and representing. As John Haugeland (1998, 180) puts the point "recording is a *process* of a certain sort; and to be a record is to be the result of such a process. By contrast, representing is a functional *status* or *role* of a certain sort; and to be a representation is to have that status or role". Were we to find a system that merely mechanically and witlessly translated inputs into a system of internal symbols, we would have no reason to take that system to be a system with mental representations. My keystrokes on the keyboard of my computer are taken as inputs into the word processor and they are then recorded in the document on which I am working.

However, these recordings are not then used in any interesting way by the word processor. The important thing to notice at this point is that it's not just the production of a representation that matters, but it's also the way in which the representation is consumed that gives it a genuine functional role to play. Without playing this role, the recording does not count as representational in the sense that relevant for the intentionally if mental states.

Building on an intuition much like this, Ruth Millikan (1989) argues that although relations of indication are either produced by a system or not (because of the way that the system is put together), the success at representing is always dependent on the proper consumption of that indicative structure. Now, according to Millikan, any phenomenon that counts as intentional does so because the semantic relations obtaining between producers and consumers are sustained by the fact that the information is so produced and interpreted; However, Millikan is also quick to note that

> 'Producers' and 'interpreters' are cooperating devices that produce and use the intentional device and *that sometimes are and sometimes are not* contained within the same individual organism (1984, 90).

Contrary to Rupert's worry, Millikan allows for intentional representations that are not bounded by the skull.

Consider a favorite example of Millikan's. Bee dances indicate the location of nectar by using variation in tempo and angle to indicate different locations of nectar. In this case, the interpreting mechanism of the watching bees serves its function just when these representations correspond to the location of the nectar, where the representation is of a dance-at-a-time-at-a-place-at-a-tempo-with-an-orientation. The thing to notice here is that the representations that are relevant for the behavior of the consuming system are not exhaustively specifiable at the level of the individual producer or the individual consumer.

The relevant representations take place at the interface of the two critters. So, while nothing is going to count as a representation unless there is a consumer for that representation, the sort of evolutionary history that ensures that the production and coordination of representations is sufficient to insure propagation of that representation is often distributed across individuals. And this seems to suggest that at least some representations have been selected for *between* individuals. Why, then, does Rupert think that it is so difficult to tell a teleological story about the production and consumption of representations in a collectivity?

Here is on way in which Rupert's worry can be developed. Jesse Prinz (2002, 4) argues that organisms like us, by which he means organisms that have beliefs and other paradigmatic mental states, "act with flexibility and forethought, choosing between different courses of action and anticipating future consequences. These abilities seem to demand representations that stand in for external objects". Millikan agrees, and claims that merely having representational structure is not sufficient for something to count as a mental state. In order for a system to be a believer in the fullest sense of the term, not only must there be internal, representational states, but there must also be, within that system, some interpreting structures that have the capacity to draw inferences (Millikan 1984, 338n2). Moreover, at least the most interesting cases of human representational capacities are decouplable from immediate stimuli, separate the indicative from the imperative aspects of a representation, and allow for disagreement about the cause of a representation (cf., Millikan 1989). Rupert's concern is to find richly representational structures like these in collectivities—and in fact these will be the most interesting cases for demonstrating the capacity for collective representation. If I can demonstrate that collectivities can possess these richly

representational states in the same way that people do, then I will succeed in demonstrating the possibility of defending a teleological theory of mental representation for collectivities.

To begin with, Millikan argues that human thought does not always take the form of inner sentences. It does, however, require that there be some intentional structures that are capable of coordinating the behavior of person in accordance with her representations of the world. In the case of belief (cf., Millikan 1984, 138), there is likely to be some correspondence between these representations and the physiological features of a human organism. These physiological structures, however, each have their own jobs to do, and the performance of each of these jobs—coupled with all of the other systems doing their jobs—contributes to the proliferation or survival of the organism (Millikan 1984, 138). Take for example the fight-or-flight response that typically occurs when a person is threatened. In this case, a number of subsystems need to be coordinated in order for the initial stimuli to eventuate in the relevant sort of organism-level behavior (cf., Millikan 1984, 117). The immediate fear response needs to be able to trigger the release of adrenaline, and this release of adrenaline needs to be interpreted by a number of different organs in the body as well as ready the organism, cognitively, for action.

However, Millikan also acknowledges that having this particular physiological structure is not type-identical to having a particular mental state, it is merely the way that beliefs happen to be implemented in humans. What is important here is that there be a number of structures that are coordinated in order to give rise to person-level behavior. So, in order for something to count as a representation, it has to be a representation for some system. The system for which something counts as a representation need not, however, be a whole organism. Now, the key insight of Millikan's teleofunctional theory of mental

representation is that in order for the output of the visual system to count as a representation, it has to be interpretable by some other system (or by some subcomponent of the system itself). Because of the way in which the human mind is organized, information that is passed from the early visual system to the affective centers associated with the release of adrenaline these stimuli can be interpreted by the affective system as dangerous stimuli. Now, in order to achieve full-blown belief, there must also be a system in the individual that can make inferences about the information contained in the release of adrenaline and decide whether to fight, flee, or calm down because the stimuli in question is not really dangerous.

This, then, leads once again to an argument from parity against Rupert. First there must be no difference *in kind* between the representations that are passed between the various subsystems in an individual and the public language and iconic structures that we see passed between persons within a collectivity. Second, the organization of a collectivity with genuine mental representations will have to be such that the states of the individuals that compose the collectivity, as well as the linguistic and iconic representations that they produce, facilitate intentional action at the collective level. Third, it will have to be the case that the collectivity is capable of reflecting on the information contained in these representations in a way that's similar to a human believer's reflective capabilities.

Suppose that the research wing of the public relations department of Wide Awake Coffee runs a series of phone surveys and collects the information that, overwhelmingly, people view Wide Awake Coffee as overpriced and cluttered with hipsters (i.e., as customers). There will, of course, be a wide degree of variation in the actual responses to the surveys. However, in order to package this information in a form that is usable by the planning wing of the PR department, the research wing will produce graphs and memos that

can be consumed by the planning wing in order to facilitate the development of a new marketing plan to both increase business (perhaps by lowering prices) and decrease the hipster quotient. Suppose further that a number of possible marketing plans are produced and that there must be some sort of collective decision made about which marketing plan would best suit the needs of Wide Awake Coffee. A plan will have to be decided on and then passed on to the board of directors, who will then have the option of accepting the marketing plan, or rejecting it (perhaps on the grounds that Wide Awake Coffee has recently signed a contract with Bright Eyes that is projected to increase the revenue from hipsters by 74 percent). In this case, we will have precisely the sorts of structures that are necessary to underwrite representational states that will allow a corporation to count as having mental representations. This is just a case where the producing system is one person (or perhaps a group of persons) and the interpreting system is also a person (or perhaps another group of persons). Once we see two person systems that are capable of using sentences as representations, it's not at all clear that these systems can't aggregate to form more complicated collective cognitive systems. So Millikan's semantic theory does not seem to be ruled out by Rupert's argument.

There are, however, still two lingering worries. First, I have yet to establish that public language structures should count as representations. Second, there probably isn't much of an interesting evolutionary history for Wide Awake Coffee—and without such a story, evolutionary history, teleofunctional theories will rule out the possibility of representation. As Millikan (1984, 91) puts it, even though the right sorts of structures are at play in Swampman, he has no contentful states because he has no evolutionary history.

Fortunately, teleofunctional theories do have the recourses for answering both of these worries.

First, evolution *has* designed humans to be incredibly flexible in their use of representations. As Millikan rightly notes, if a system is designed to be a learning system, then changing its structures through learning is precisely a part of its proper functioning! Unlike bee dances, human language is not evolved but learned (Millikan 1984, 98). Thus, the structures at play in a person are flexible and adaptive structures that allow for producing systems to be coupled to various different interpreting systems that happen to speak the same language. As Millikan puts the point, "each individual human must develop his or her own programs by a process probably involving trial and error. But these programs must govern the production of inner terms at least many or most of which match terms in the public language of the community in which the individual lives" (Millikan 1984, 140). However, once these programs are developed, the same representational state can be used for a variety of purposes (e.g., in an individual agent it might be used in both theoretical and practical reasoning). Why not think that these representational capacities can be extended to facilitate the coordination of collective action. After all, if there is a story to be told about the evolution of language, it's likely to rely, at least in part, on its capacity to facilitate coordination.

However, as Rupert will be quick to note, all of the relations that would underwrite the propagation of representations within a collectivity will already be in place outside of the collectivity, and for this reason it seems superfluous to posit the representations as interestingly collective representations. But this objection cannot serve the required function. In order for the exaptation of a representational capacity to be possible, the only thing that

needs to stay the same is that the representation must correspond to environmental configurations in accordance with the same correspondence rules for each of these activities. That is to say, the evolutionary considerations that confer meaning on a particular sort of representational structure must remain the same so long as the representation is to be usable as the sort of representation that it is. Now, given that human language has developed as a method of representing that sits half way between two people, humans are precisely the sorts of organisms that can learn to redeploy these public language structures in order to produce behaviors that will regulate the behavior of collectivities. The content of the representations is indeed fixed by the evolutionary history of the organisms that make up the collectivity. However, these representations are being used for a different end—even though the content remains stable. To put the point a different way, the structures that regulate the continued production of a particular sort of representation are outside of the collectivity in question; however, these representations are redeployed in order to direct the behavior of a collectivity in a way that allows the collectivity to respond to changing features of *its* environment. Corporate entities, for example, use already existent representational capacities in humans in order to facilitate and coordinate social actions and in a way that allows for the sort of reflection that is required for collective mentality.

4.5 Representation and action guidance

There are, however, further worries about the capacity of the subsystems in a collectivity to actually give rise to collective behavior; and, at this point we can return to the claim that collective representations are superfluous in explaining the behavior of a collectivity. In order to be successful, proponents of superfluity arguments must change their

attack. Even though there are no differences in the inter-level relations as regards *the production* of representations, there seems to be a difference in the causal sequences of effects produced by the visual representation of motion and the visual representation of discrete and static representations of objects. If someone throws a beer bottle at my head, and I perceive it moving rapidly toward me, I'll try to get out of the way. But, my avoidance behavior is a response to the perceived motion of the bottle, not a response to a sequence of static representations that underlie the state of the collectivity. In order to explain my behavior, person level representations are required. Put briefly, commonsense psychology is an effective tool for prediction and explanation for *a wide range of behavior*, and this predictive and explanatory advantage give us good reason to retain individual level representations. I agree; however, if we take cognitive science to be directed at the explanation of behavior, this also provides good reason for retaining collective representations as explanatorily valuable structures.

I have suggested that a theory of individual representation must allow for causal mediation by lower-level representational states; but it would be nice to know why we should expect such representational mediation. Taking representation to be a relation of indication or bearing information obtaining between neurophysiological states and properties of the world has led to theories of mental representation focusing on static and unchanging states of the world. However, if cognitive science is really an attempt to explain behavior, these static representations are the wrong place to focus. Presumably, any theory of representation that's sensitive to the evolutionary and developmental origins of our capacity to represent must note that the reason for having representational capacities is that they allow us to cope with a rapidly changing and dangerous world.

The capacity for sensory representation is the most primitive representational capacity in biological organisms, and attending to the operation of sensory systems, rather than high-level conceptual structures, suggests that representational capacities often have the function of informing action. A visual system can alert me to the fact that something is moving rapidly towards my head, and the production of a visual representation is, in many cases, sufficient for me to engage in avoidance behavior. This behavior, however, is produced by a variety of simple systems working in parallel in order to produce this behavior. Perceiving a beer bottle that is being thrown at my head requires that my retina be irradiated and that the information about the stimulation of retinal cells be propagated toward the LGN (the visual systems relay center) as a digital representation (in Dretske's sense) of the stimulation the retina. Upon arrival at LGN, the information is dispersed to various regions of the visual system where some information is processed by systems dedicated to capacities such as detecting edges and color while other information is processed by systems dedicated to capacities such as spatial awareness and the guidance of action. However, as becomes painfully obvious in the case of blindsight, the visual representation of a beer bottle being thrown at your head is dependent on the proper functioning of all of these areas working in coordination and passing relevant information to each other. The blindsight patient might, when prompted to make a guess, be able to determine whether a beer bottle is being held sideways or upright with a relatively high probability of being right. She might be able to tell you what color the bottle is. However, she's not going to move out of the way if you throw it at her because she's not going to represent the dangerous stimuli. The only way for her to represent an oncoming beer bottle is by having a number of representational systems working together for the production of such a representation.

Now, if we attend to the practice of cognitive science, we see that it rests on the assumption that representation occurs much earlier than the level of individual representation. In order to make sense of the behavior of individuals, cognitive science assumes cognitive specialization and differential processing occurring throughout the brain—and it is on this point that the methods for collecting data (e.g., fMRI, EEG, PET, etc) in cognitive neuroscience turn. However, it's also clear that at no point during the production of the visual state prior to the final output in conscious monitoring do we have an adequate account all visual representations. Some sorts of visual representations can only be specified at the level of a whole person, but these visual representations themselves must be seen as having a rich representational structure.

In much the same way that accounts of individual representation have typically gone astray, the defense of collective representation has typically taken static representations (e.g., court decisions and press releases) to be the only collective representations. However, these states aren't the most promising avenue for developing an account of collective representation either. These states are the result of computations over lower-level representations. Taking *these* public language structures to exhaust the representational states of collectivities is analogous to taking an individual's utterances to exhaust her mental representations; and from there it's a short step to behaviorism! So, while it is clearly right to acknowledge that public language structures facilitate the propagation of many collective representations, a far more promising strategy for establishing collective representational capacities is to begin by considering the ways in which various representations within a collectivity are "propagated from one representational medium to another by bringing the states of the media into *coordination*" (Hutchins 1995, 117) in order to guide behavior.

Consider the case of crime scene investigation (CSI). In CSI, "evidence is likely to be collected by one group of people, analyzed by another, and interpreted and presented to Court by another group" (Barber et al. 2006, 358). The collection of data may begin with the data collected at an emergency call center where a call-handler codes the caller's analog representation of the crime scene, in real time, as a digital representation what the caller says. This representation is sent to a dispatch operator who interprets it, gating off information that's irrelevant to dispatching officers. The dispatch operator thus converts this information into a digital representation that can be consumed by the investigating officers.

On the basis of this representation, investigators proceed to the scene and collect data. They dust for fingerprints, examine footprints, and collect stray hair follicles and discarded clothing. Investigators take the entire scene and distil it into evidential representations such as photographs, clothing, and fingerprint dustings; however, these representations must be made digital in order to be consumed by those not trained in CSI—noise must be distinguished from data in a way that's consonant with what investigators take to be relevant to prosecuting someone in this case. Once these data are collected, it must be analyzed (to determine whether there's sufficient evidence to prosecute) and converted into a narrative structure in order to facilitate prosecution. This narrative structure, however, is just the end result of a complex interaction of various low-level representations produced during data acquisition. At this point we could appeal to the representational states of the individual who pens the narrative and the representational states of the investigators who collect the data as the cause of this narrative representation. However, this leaves too much out.

The propagation of information through these sorts of collectivities does not depend exclusively on the architecture of the system, nor does it depend exclusively on the

intentional states of the individuals that compose the collectivity. Which representations are passed between individuals also depends on shared background assumptions, which features of the environment happen to be salient, global considerations about what sorts of information will be useful in achieving the goal of the collectivity, and facts about how data was interpreted in the past (Heylighen et al. 2004, 8). Each investigator "only needs to know what to do when certain conditions are produced in the environment" (Hutchins 1995, 199), but through their interaction, the narrative emerges, and it's only through the production of this narrative that goal of prosecuting becomes a possibility.

We could, of course, abandon the level of analysis at which the narrative is produced, or we could merely attribute the narrative structure to the last person implicated in its production. However, if we do this, we must make a parallel move in the case of the individual, for the account of specialization at play in the collective case parallels the sort of specialization that we find in human psychology. But, if this argument succeeds, retaining individual representations warrants retaining collective representations.

4.6 Conclusions and looks forward

If all has gone well up to this point, I have shown that the most promising philosophical objections to the possibility of collective mentality miss their mark. In Chapter 2, I suggested that arguments against collective mentality on the basis of considerations about consciousness were unlikely to impugn the possibility of collective mentality. In Chapter 3, I suggested that the same sorts of considerations that allow us to understand individual psychological states as autonomous from neurological states could also be used to defend the autonomy of collective psychological states. And in this chapter, I have suggested that there

is no reason to rule out the possibility of collective representation on philosophical and conceptual grounds. At this point, the *possibility* of collective mentality seems to be well established. However, there is more work to be done.

My claim is not merely that collective mentality is possible. Rather, my claim is that there are cases of collective mentality in our world—cases that ought to be studied by the cognitive sciences. Having responded to the objections that have occupied me for the last three chapters, I now have the tools to develop a theoretical framework in which to ground the study of collective mentality. In the remaining chapter, I take my task to be as follows. I address a number of research projects that have attempted to establish the existence of genuinely cognitive collective systems. The dominant strains in this sort of research tradition have focused on collective decision making, collective memory and distributed cognition. In Chapter 5, I, thus, begin by rehearsing a series of desiderata for what it will take to suppose that a collectivity has genuine cognitive states. Having laid out these desiderata, I then address a number of overtly cognitive phenomena that appear in collectivities.

CHAPTER V:

COLLECTIVE MENTALITY REVIVED!

In the previous chapters, I have argued for the *possibility* of collective mentality. Within philosophy I have a few allies;[96] however, the dominant trend in defending collective mentality attempts construct collective mentality out of individual mental states. If my theory of collective mentality is correct, we should start by looking at the behavior of collectivities, and then look at the computational architecture of these systems to see if they are sufficient to produce genuinely cognitive states. Given that my project is a piece of theoretical cognitive science, I now must attempt to establish the *actuality* of collective mentality in accordance with the model that I have thus far developed. In this final chapter, I demonstrate that a number of collectivities in the actual world ought to be counted as genuinely cognitive systems. However, before moving to an analysis of these collectivities, I'll make some brief preliminary remarks about what it will take to establish my case.

5.1 A few preliminary remarks on collective representation:

In the previous chapter, I argued that intentional states are typically layered in individuals—that is, person-level representations are typically constructed out of lower-level

[96] Bill Lycan (1981) gestures towards such a theory and D.H.M. Brooks (1986) makes a similar move in "Group minds". However, although these theories are sketches of the sort of project that I have here been developing, neither is fully elaborated in a way that suggests itself as plausible theory to be extended to research programs within the cognitive sciences. At this point, I can now demonstrate the ways in which my account of collective mentality can be developed into a research strategy for a non-individualist cognitive science.

representational states. On the basis of this consideration, I argued that the fact that collective mental states are composed out of lower-level intentional states does not preclude the possibility of collective mentality. I then claimed that wherever we find a purportedly cognitive system that possesses a system of identifiable internal states or processes that are 1) capable of bearing information about some state of affairs in the world, and 2) that are capable of directing the immediate behavior of that system, we have good reason to suppose that we've discovered a sort of intentional state. However, more needs to be said on this point.

First, the fact that a system can be in a state that is capable of bearing information about the world, and is capable of directing the immediate behavior of that system, even where this state is composed out of lower-level intentional states, is *by itself* insufficient to insure the presence of genuinely cognitive states.[97] In order to count as a genuinely cognitive state, these internal states need to be representations rather than mere reportings (cf., Haugeland 1998). As John Haugeland (1998, 180) puts the point, while something needs only be the result of a particular processes in order to count as a recording, "representing is a

[97] Consider the way in which the fuel injection system of a modern automobile works. The fuel injection system is constituted by a number of functionally organized components, each of which is designed in such a way that it detects changes in its environment in order to facilitate acceleration. In a modern automobile, when the gas pedal is depressed the throttle valve is opened in order to increase the amount of air in the system. When the air increases, the engine control unit detects the open throttle valve and increases the rate at which fuel is flowing into the engine in order to ensure that the fuel-air ratio remains constant. This is achieved by using a magnet to force open the fuel injector—causing a highly pressurized stream of fuel to be released into the engine manifold. However, this isn't the whole story. In order to ensure that the right amount of fuel is being released into the engine manifold, a series of sensors, including the mass airflow sensor (which monitors the amount of air entering the engine), oxygen sensors (which monitor the amount of oxygen in the exhaust system), and the throttle sensor (which monitors the position of the throttle valve) have to produce representations that can be coordinated in the engine control unit in order to determine the amount of fuel that must be released into the engine manifold. The state of the fuel injection system bears information about the state of the fuel-air ratio in an engine and it is capable of directing the behavior of the engine on the basis of such information. Moreover, each of the subcomponents of the fuel injection system are capable of bearing some sort of information about some state of the engine and in virtue of this information, they are capable of directing the immediate behavior of some component of that system. Unless my account of collective mentality is capable of distinguishing between a genuinely cognitive system and the fuel injection system of a modern automobile, something has clearly gone wrong.

functional *status* or *role* of a certain sort, and to be a representation is to have that status or role." The question, then, is: what sort of functional role has to be filled if something is to count as a representation. In what follows, I'll follow Haugland (1998, 172; see also Clark 1998) in his rough and ready desiderata on the sort of functional organization required for something to count as a representation. I suggest that if we're going to appeal to representational states of a collectivity in offering genuinely *psychological explanations* of a collectivities behavior,

1) The system must possess internal states that have the function of adjusting the system's behavior in ways that allow it to cope with features of its environment in ways that are not fully determined by the design of the system;

2) Such states must be capable of standing-in for various features of the environment that are important to the system, even in the absence of immediate environmental stimuli;[98]

3) Such stand-ins will have to be part of a larger representational scheme that allows a variety of possible contents to be represented (in a systematic way) by a corresponding variety of possible representations, and

4) There must be "proper (and improper) ways of producing, maintaining, modifying, and/or using the various representations under various environmental and other conditions" (Clark 1998, 147).[99]

Since functionalism is a topic-neutral theory, I remain noncommittal about the occupants of this functional role. However, because the representational states to which I appeal take the

[98] There are, of course, deep questions about the diachronic question here. It seems as though we can do pretty well at representing some things that we've never come into contact with (e.g., unicorns, griffins, Harry Potter, and radical democracy). There is much to say about these sorts of representations, but I'll not get into those disputes here.

[99] Note that this does not require genuine misrepresentation. Scorpion can be tricked into responding as though there were prey in front of it by introducing something into its environment that creates the same sorts of vibrations that would be created by something that the scorpion would eat in an untainted environment. More familiarly, you can fill a frog with buckshot by shooting BBs past it in the lab. In each of these cases there is a representational failure. However, these systems are not genuinely misrepresenting their environment—they just have detection systems that are too impoverished to discriminate food from near-non-food. To put the point briefly, representational failures, like the disjunctive kind Flies-or-BBs, come in different flavors.

form of images, icons, maps, graphs, and honeybee dances, some philosophers might worry that such states can't be anything more than mere recordings. The important thing to notice, however, is that the intuition that some *vehicles* are incapable of representing rests on the though that the analog information conveyed by an imagistic representation (for example) can be consumed by a computational system *only after* it is converted into a digital representation such a 'word' in a language of though. Within the philosophy of psychology it is often taken as given that images record information but only concepts represent.

In contrast to this trend in philosophy, however, a substantial tradition has developed around the claim that there's not much that one can do with an arbitrary symbol. Provided a system is built in such a way that it can interpret some sorts of symbols as standing-in for some feature of the environment, there's a lot that a system can do with such structures.[100] In his analysis of the popular media, Marshal McLuhan claimed that 'the medium is the message' and oftentimes there is much truth in this claim. Some systems are designed to interpret only a particular sort of media as a representation. These systems have translation rules built into them for converting one representation into another by appeal to a system of background rules for interpreting that are designed right into the system. For such systems, the medium is all-important! If, for example, the human visual system is organized in such a way that the detection of stable geometric properties are immediately converted into representations that can be consume by higher-cognitive systems anterior to the visual cortex, then there will likely be structures in the visual cortex that immediately and witlessly convert representations of geometric properties into whatever sort of representation is capable of being consumed by higher-cognitive systems. However, even in this case *it's not the medium*

[100] I am thinking here of philosophers as diverse as Martin Heidegger (1927/1996) Ludwig Wittgenstein (1953/2001), Wilfrid Sellars (1963/1991), Ruth Millikan (1984), Andy Clark (1997), and Jesse Prinz (2002).

but the message that makes these outputs of the visual system representations. The important point is that these visual outputs have a representational structure that facilitates coping with the environment. It doesn't matter whether these representations are logical symbols, iconic representations, cognitive maps or anything else for that matter—provided that they can be consumed by a system *as a representation* that allows the system to engage in intentional action of some sort. What does the work of helping a system to navigate its ever-changing environment is not necessarily a fact about the *vehicle* of the representing, but the *content* of the representation.

Keeping these theoretical commitments about collective representation in mind, I am faced with one further problem in defending the actuality of collective mentality. Because the cognitive and social sciences have developed primarily (though not completely) in North America and Western Europe, there is a strong bias in favor of research programs focusing on individual's mental states (cf., Huebner et al, under review). This makes it difficult to find evidence that bears directly on the existence of collective cognitive system. However, this commitment to individualism has started to falter in recent years. Dynamical systems theory and theories of self-organizing systems in biology provides some promising evidence for collective mentality. Moreover, biologists such as David Sloan Wilson, who are committed to multi-level selection, have also provided some evidence for the existence of collective cognitive properties of groups. From another perspective, social scientists (e.g., David Sloan Wilson and Daniel Wegner) have self-consciously attempted to resuscitate the idea of a "group mind" on the basis of experimental evidence about the behavior of small groups on cognitive tasks. Finally, on the basis of arguments initially offered by Rumelhardt, McClelland and the PDP research group (1987), an anti-individualist cognitive science

collecting data from academic fields as diverse as informatics, robotics and cognitive anthropology has attempted to locate numerous cognitive states outside of the skin and has recently began to produce interesting results about a range of phenomena called 'Distributed Cognition'. In the remainder of this chapter, I review each of these traditions to see if they succeed in providing evidence for the existence of collective mentality.

5.2 Cognition and self-organizing systems

In their groundbreaking book*, Self-organization in non-equilibrium systems*, Gregoire Nicolis and Illya Prigogine (1977) utilized the tools of nonlinear fluid dynamics to model the behavior of large and transient populations of animals. Though the mathematics is incredibly complex, the basic idea is simple. The behavior of macroscopic biological systems (e.g., the movement of water buffalo in search of new pastures, the construction of termite mounds, and the milling of fish) can be tracked as regular and observably stable patterns. By treating the individuals in these collectivities as 'black boxes' with relatively simple desires and computational capabilities, it is possible to mathematically model the behavior of animal collectives in a way that demonstrates a formal correspondence between these systems and the simple chemical reactions of non-linear fluid dynamics (cf., Sumpter 2006, 6). The key insight of this non-linear modeling is that the aggregation of relatively simple states of the individuals that compose a collectivity is sufficient to produce emergent collective behavior that is not exhibit by the component systems. The desires and mechanisms required to explain these collective phenomena is often so simple that the models can be constructed by children in elementary school armed with some simple yet cleverly designed computer software (Cf. Resnick 1994). However these models are also explanatorily powerful. The

question, however, is whether these models are capable of explaining anything that should count as genuinely cognitive.

At the end of chapter three, I argued that an appeal to self-organizing systems is never *by itself* sufficient to justify claims to collective mentality. Although the segregation phenomena modeled by Thomas Schelling (1971) is incredibly interesting *as an emergent phenomena*, there is nothing in Schelling's model to suggest that the segregation behavior rests on a cognitive state of the collectivity. However, the fact that *these phenomena* are not cognitive does not give us reason to dismiss self-organization as capable of producing cognitive states. After all, brains are self-organizing systems whose constituent parts, neurons, obey relatively simple rules (cf., Kelso 1995); and it had better be the case that brains are capable of producing genuinely cognitive states if anything is. The question thus arises: are there collectivities in which genuine cognition will emerge from the interaction of relatively simple algorithms?

There are three plausible places to look for cases of self-organization that produce cognition in collectivities: 1) aggregative phenomena in large groups of humans such as crowds, 2) flocking, schooling, and herding behavior, and 3) the behavior of social insects. I'll take each of these phenomena in turn to see whether there is any interesting sense in which they give rise to collective cognitive phenomena.

5.2.1 Tipping points and rioting mobs: Adam Smith (1776/1996) famously argue that that economic trends emerge from the aggregation of individual desires, and John Stuart Mill (1843/1988) claims that on the basis of a few psychological platitudes we can explain the presence of market trends. A radically individualistic industry of game theory and rational choice theory has been constructed on the basis of these assumptions, and this industry has

had *some* success. While I don't find such explanation appealing or promising, there is a very important insight to be had in attending to these models. As those philosophers and economists that defend methodological individualism have always noted, there are numerous phenomena for which we will not have reached a 'rock-bottom' cognitive explanation until we have explained the cognitive states of the individuals that compose a collectivity (Watkins 1952). The reason for this is that the phenomena that occur at the level of economies, for example, are emergent phenomena but not emergent *cognitive* phenomena. The behavior of markets emerges from the interaction of individuals, and they are even directed at states of affairs in the world (often, with bringing about states of the world that are completely hostile to the interests of the individuals that compose these markets); however, these markets are not, themselves, thinking things.

Similar cases abound with what one might call, following the sociologist Morton Grodzins, 'tipping point' phenomena. Grodzins used the term 'tipping point' to refer to the point at which 'white flight' occurs in inner city neighborhoods. Grodzins (1958) collected data on the emigration from metropolitan areas and argued that the flight of white working and middle-class people to the suburbs was not explicable on the basis of a linear model, but required recognizing that *neighborhoods* had 'tolerance' for the number of minorities that they could allow before the white members of the community left. This phenomenon of white flight, as with the segregation phenomena modeled by Schelling (1971), is produced by the complex interaction of the psychological states of the individuals in a particular neighborhood. But as Schelling notes, by positing a relatively small in-group bias, we can explain the emergence of a high degree of racial segregation. Schelling also noted that once movement started within a particular area, it would be self-sustaining, because of the way in

which in-group biases force individuals to resist living in a neighborhood where they are becoming outnumbered by members of an out-group.

Phenomena such as 'white flight' should not lead us to suppose that there is collective mentality underwriting racial segregation. After all, the explanation of the phenomena is straightforwardly explicable in terms of the psychological states of the individuals that compose the collectivities in question, provided, that is, that we attend to the way in which these psychological states are embedded in their social environment. However, this does not mean that there are not emergent phenomena here that are worth studying. It is, after all, only by recognizing that there are significant effects of living in a social world that we can even begin to study the way in which individual psychological states are modulated by their external environment.

Malcolm Gladwell (2000) has recently adopted the term 'tipping point' in an attempt to explain a wide range of social issues. Gladwell (2000) argues that phenomena such as the popularity of a shoe, the decrease of crime in Manhattan, and the prevalence of smoking amongst teenagers are best explained according to sets of simple, aggregative rules. He argues that a few people who act in a particular way are capable of causing large-scale social changes. One or two people behaving in a particular way are not likely to produce large-scale social changes. However, there is a propensity toward imitation in our species that is capable of producing large-scale phenomena as people 'catch-on' to behaving in a particular way. The interesting thing about such phenomena is that the best way to intervene in them is not by modifying the collective as a whole, but by modifying the psychological states of a few individuals within a collectivity in such a way that an idea will spread throughout the rest of the collectivity.

These mechanisms appear to be variants on those suggested by R.A. Wilson in his discussion of the social manifestation hypothesis (SMH). Segregation is a collective property; however, our best explanation of the mechanisms that give rise to this phenomena are not collective cognitive mechanisms but individual psychological mechanisms embedded in a social world. Analogously, there is much to be learned from the study of situationist psychology about the ways in which an individual's world modulates her social states. However, merely recognizing that these forces are at play in the social world is not sufficient to generate collective mentality. Merely causal relations between the mental states of individuals are not enough to produce collective mental states. What we need is a story about how it is that the states of the components of a collectivity are organized in such a way that they constitute a single unified cognitive system—and these tipping point phenomena are unlikely to be sufficient to achieve constitution rather than mere causation. However, there is one more incredibly important lesson to be learned from attending to the sorts of phenomena that I think ought to count as genuinely collective phenomena but not as collective cognitive phenomena.

Numerous social psychologists in the late nineteenth and early twentieth century were concerned to account for the behavior of crowds in terms of collective psychological states. For example, Gustav Le Bon (1895/2002) argues that individuals in a crowd become unconscious automata controlled by the suggestions of a collective mind over which they have no control. He also argues that crowds possess a sort of mental unity, producing a sort of *sui generis* entity distinct from the individuals that compose a crowd (Le Bon 1895/2002, 4). In developing his account of the group mind, Le Bon claims that individualistic psychology ignores the lawful generalizations occurring at the level of crowds; although

there are numerous significant differences in individual intellectual capabilities and desires, such facts are irrelevant to the behavior of many crowds and there are numerous things that we can know about crowds without reflecting on these beliefs and desires of the individuals that happen to compose those crowds. My analysis of collective mentality would be conspicuously lacking were I not to address this issue of crowd mentality.

To begin with, I must note that there are many things that can be accomplished by crowds that cannot be accomplished by isolated individuals. Flipping over a car, tearing down a statue of a deposed leader, succeeding in a coup d'etat, or shutting down an intersection are things that groups can do but that individuals cannot do by themselves. However, the mere fact that there are activities that require collective action does not require collective mentality, and this is the important point in considering the mentality of crowds. Although it is indeed possible that some crowds might have mental states, it is not always the case that a crowd has sufficient functional organization to produce anything like a unified cognitive state.

Consider the case of the race riots that followed in the wake of the Rodney King decision. On April 29, 1992, three of the white officers who had brutally attacked King a little over a year earlier were found innocent of all charges. That evening, following the verdict, riots erupted throughout Los Angeles and continued for several days until the National Guard and the Marines intervened. In understanding this riot, there are significant psychological phenomena that we must understand.

The collective action that became a riot began as a peaceful protest outside the courthouse where the verdict had been passed down. A number of people gathered to peacefully register their disagreement with the decision. But the crowd grew rapidly, and as

the police withdrew, fearing their own safety, things made a rapid turn for the worse. To put the point succinctly, the crowd became *angrier* throughout the day. A significant contributing factor in the change of the valence of the emotional state of the crowd was the widely shared belief among members of the African American community that the court decision was the result of *a racially biased system* that had targeted African American residents through racial profiling, police harassment, and unfair treatment in the courtroom. This belief produced a legitimate feeling of anger in a number of people in the crowd. This anger spread rapidly, and it is at this point that a number of the insights suggested by Le Bon (1895/2002) become quite important. The question is: Do we need to explain the anger of *the crowd* or can everything be explained by appeal to the states of the individuals within the crowd?

There was a broad consensus within the collective psychology tradition of the late 19th Century that the emotional states of individuals can be brought under the control of the *sui generis* mental states of a continually maddening crowd (cf., Le Bon 1895/2002, McDougall 1920, and Freud 1921/1975). Le Bon (1895/2002) even went so far as to argue that understanding the behavior of crowds requires us to recognize that "the sentiments and ideas of all of the persons in the gathering take one and the same direction, and their conscious personality vanishes" (Le Bon 1895/2002, 2); to put it another way, crowds are completely irrational and driven by exuberant emotion (cf., Le Bon 1895/2002, 101). Now, although this turns out not to be true of all crowds (cf., McPhail 1991), there is good reason to think that the emotional states of people in a crowd will, at least in some cases, tend to converge.

As I've just noted, in the riots following the Rodney King decision, a legitimate feeling of anger among a number of people in the crowd at the courthouse spread rapidly and

was enlivened. Both Le Bon and Freud were unaware of the mechanisms at play in giving rise to the mental unity of crowds; however, when this thought is coupled with an insight from Charles Darwin (1872/1965) that there is a tendency toward imitation, that acts independently of the conscious will, we can begin to construct an account of the mechanism that gives rise to the mental unity of crowds.

To begin with, it is a familiar phenomenon that all of the babies in a nursery will begin to cry when any other baby does—and this happens within hours of birth. Moreover, we are all quite familiar with the fact that when we see another person smiling or laughing, we are likely to do the same. More to the current point, it is quite hard to remain happy or emotionally neutral when someone who is a real downer walks into the room. These facts about human psychology suggest that humans possess some mechanisms for producing a sort of emotional contagion. As David Hume (1739/2000, 365) puts the point, "the minds of men are mirrors to one another, not only because they reflect each others emotions, but also because those rays of passions, sentiments and opinions may often be reverberated".

Now, this capacity is not merely an artifact of commonsense psychology. The claim that emotional contagion is a psychologically robust fact about the world is supported by a wide variety of results from the cognitive sciences. Preston and DeWaal (2002), for example, review data that suggests that animals that see a conspecific in a threatening situation are also likely to experience a behaviorally and biophysically noticeable fear response. This imitative response also seems to be underwritten by neurological mechanisms dedicated to motor mimicry. In 1996, Giacomo Rizzolatti and his colleagues (1996) found a population of neurons in F5 of the macaque premotor cortex that were active both when the monkey performed an action and when the monkey observed the same action being preformed by a

conspecific or an experimenter. These 'mirror-neurons' respond to actions that are purposive or goal oriented such as grasping movements and others that respond selectively to various sorts of gestures (cf., Gallese 2001, 2003a, 2003b; Rizzolatti et al. 1996).[101] It seems reasonable to think that emotional states are likely to be perceived as goal oriented, and as Ralph Adophs (2002) has shown, the perception of an emotion activates the neural mechanisms responsible for the production of emotions in much the same way that these 'mirror neurons' are supposed to operate.

Now, to return to the race riots following the Rodney King verdict, we can now see how the distributed activity of the members of a crowd can appear to be unified, can even appear to possess a sort of mental unity even in the absence of any collective cognition. There is good reason to think that crowd behavior is much more like the behavior of a herd of antelope fleeing from a predator (which I discuss below) than it is like genuine cognitive activity. We should begin by recognizing that some of the people who showed up at the courthouse were angry about the decision and about the racially biased practices of the LAPD. Because of the way that emotional contagion works in human individuals, this anger rapidly spread through the rest of the crowd. As this emotional state spread to each of the members of the crowd, the crowd began to exhibit a unified state of anger. At this point, we must consider the truth of the claim that the rioting mob was becoming angrier was the day wore on.

As the mental states of the members of the collectivity become more and more unified, the tendency to speak as-if the collectivity has a mental life of its own will become more pronounced. As Bloom and Veres (1999) demonstrate, commonsense psychology

[101] Jean Decety and her colleagues (2002) have argued for the existence of a similar system in humans based on PET data on imitation. Converging evidence has been offered by Kevin Pelphrey (2003) with fMRI data on the perception of biological motion.

allows for the description of the behavior of a collectivity using intentional idioms in cases where the behavior of a collectivity appears to be unified. In the case at hand, the emotional mirroring that occurs in a rioting crowd produces an apparent mental unity (as was aptly noticed by Le Bon and Freud). However, although this behavior is apparently unified, it is underwritten not by anything that produces genuinely collective mental states but by the aggregation of increasingly similar *individual* mental states. Furthermore, in this case we begin to see behaviors that cannot be accomplished by individuals on their own; and in a number of cases, the aggressive and violent behaviors of the members of the crowd will only be exhibited from within a crowd. For these reasons, it *is thus true* to say of the mob that it was becoming angrier as the day went on. However, this claim has to be read distributively. As the day wore on, more and more *individuals* became angry.

Although it may prove useful to describe the behavior of the collectivity in terms of the increasing aggravation, this is only because the behavior produced by the aggregation of individual mental states is a self-organizing and emergent behavior. Presumably, although the modeling of such phenomena is likely to prove incredibly difficult (especially given that there will be a wide variety of initial starting states, variations in susceptibility to emotional contagion, and variations in the willingness to commit violent acts when angry) modeling the behavior of a rioting crowd will be possible without positing mental states for that crowd. This is not, of course to say that there is no such thing as mob mentality. However, unless there is good reason to assume that there is sufficient functional specialization to insure that collective representations are being produced in the way that is required for genuine mentality. Although I am unaware of any cases such as this, I am unwilling to foreclose the possibility of mob mentality in the actual world—it is just likely to be incredibly rare.

If the picture of collective mentality that I've been developing is correct, focusing on the psychology of the crowd turns out to be a radically misguided project. Contrary to the views advanced by Freud and Le Bon, I argue that an adequate theory of collective mentality must focus on the way in which the specialization of function in a collectivity facilitates the propagation of representational states across a variety of representational media. I also argue that this functional specialization must be integrated enough that the collectivity ought to be seen as possessing goals and projects that are not specifiable as states of the individuals that compose that collectivity. This sort of functional specialization is analogous to the way in which the various subroutines in human cognitive architecture are organized and that the way in which representations are propagated in some collectivities is analogous to the way in which this occurs in subroutines in individuals. And, I claim that once we realize that a homunctional theory is the best explanation of individual mentality, there is no reason to bar the possibility of collective mentality grounded on functionally similar homunctional architectures.

5.2.2 Herd mentality: Building on the research on self-organizing systems, the next place that it seems plausible to look for collective mentality is in flocks, herds and shoals. These collectivities are specialized for dealing with a particular range of phenomena in the world, and they are capable of acting in ways that the individuals that compose those collectivities are not capable of acting. Consider the way in which flocks of birds, herds of land animals, and schools of fish form in response to looming predators. Although a flock of birds is made up of multiple unconnected birds, flocks seem to move fluidly as unified systems. The question, however, is whether these complex systems themselves possess distributed mental states that produce in these behaviors.

As Craig Reynolds (1987) demonstrates (using a computational simulation of flock behavior), the behavior of the flock can be modeled by providing each of the birds in the simulation with just a few simple rules for in-flock behavior. First, given the benefits of being inside rather than outside of a flock,[102] each of the birds must have the desire to match the velocity of other birds. Second, given that the benefits of being in a flock are optimized at the center of the flock rather than on the periphery, the birds must each have a desire to stay as close to the center of the collective as possible. Finally, to keep the flock airborne, the individual birds that compose the flock must also have the desire not to crash into one another. Similarly, Iain Couzin et al (2002) have provided a model that adequately simulates a wide variety of schooling behaviors in fish on the basis of three simple rules. Taking the center of the school as the origin, they posit a zone of repulsion, a zone of alignment, and a zone of attraction. The zone of repulsion (ZOR) excerpts a great deal of force over a small area (such that fish in ZOR always move away from one another. The zone of attraction (ZOA) excerpts less force on an individual fish, but excerpts a force over a much larger area. Couzin et al found that by modulating the size of the zone of alignment (ZOO) while maintaining the ratio between ZOR and ZOA they could create a variety of different schooling phenomena that are perceived in nature. When ZOO was small, the 'fish' had the character of a loosely packed stationary form. As the size of ZOO increased they began to circle around the center of mass of the school, and then began to move together as a unified school in a single direction. Thus, by positing a few simple rules for behaving within a

[102] Research on the reasons for joining a flock, herd, or school suggests a number of reasons why doing so can be beneficial to an individual animal. However, as Reynolds (1987, 28), notes, "The basic urge to join a flock seems to be the result of evolutionary pressure from several factors: protection from predators, statistically improving survival of the (shared) gene pool from attacks from predators, profiting from a larger effective search pattern in the quest for food, and advantages for social and mating activities".

school, Couzin et al were thus able to adequately model many common behaviors of schools of fish.

Each of these cases of self-organizing behavior is grounded on the 'selfish herd' model that is often posited in order to account for emergent behavior in herds of land animals (Hamilton 1971). The key posit of such models is that animals will always struggle to gain access to the center of the herd because at the center of the herd, animals are more cable of avoiding predation without standing guard. After all, if there are other animals around you who have to watch for predators, all you need is a simple rule that says: if the animal next to me is running, I should run too. While some animals (the unlucky ones who end up at the edge of the herd) are forced to watch for predators, those near the middle are capable of grazing and sleeping without having to attend to the highly dangerous world in which they live. In this way herd gain an important benefit from being in the herd that they would not have outside of the herd. Here again we have an important sort of emergent behavior—the question, however, is how well do these systems fair as cognitive systems according to the desiderata that I've laid out for collective mentality.

The important question, for my purposes, however, is whether the specialization of function in a herd is capable of propagating representational states in a way that will yield genuinely collective mentality. Consider the case of a herd of antelope grazing on the savannah. The animals at the edge of a herd have to be able to detect a looming predator, and I think it is fair to say that an adequate theory of representation will attribute to this antelope the representation of a predator. This representation of something as a predator will then trigger an immediate flight response. Now, the antelope who have not seen the predator don't need to represent much of anything other than that there are a number of conspecifics

running. More stupid evolved mechanisms for retaining herd integrity will then kick in and cause the antelope at the center of he herd to flee as well. We can look at the behavior of the herd and (by adopting the intentional stance) we can explain the behavior of the herd as the intentional phenomena of fleeing from the predator or perhaps even trying to confuse the predator. However, here's the important point. The things that appear to be cognitive states of a herd of antelope are nothing more than simple aggregations of individual cognitive states of individual antelope.

If we want to explain why the herd turns left at a particular point, this behavior will be explained in terms of the behavior of the antelope at the edge of the herd turning in response to an obstacle. This will, of course have a rippling effect, causing many of the members of the herd to respond in such a way that the herd appears to move as a unified system. However, there are no states of the collectivity here that are doing any interesting cognitive work. This is not, of course, to say that there are not any interesting emergent phenomena here. In fact, the modeling of complex self-organizing systems has given us a great deal of insight into the sorts of mechanisms at play in producing collective behaviors in animals. However, there is just insufficient functional specialization within a herd to produce anything that looks interestingly cognitive. Sure enough, the states of the individuals within the herd are determined (if not fully, then) to a large extent by the state of their local environment. However, we gain neither predictive nor explanatory advantage by appealing to the cognitive states of the collectivity above and beyond the cognitive states of the individuals that compose that collectivity.

In fact, this is built right into the model of explanation here. By positing a set of beliefs and desires that are sensitive to the local environment, these self-organizing systems

models are capable of explaining the emergence of large-scale herd phenomena without having to posit any cognitive state of the collectivity itself. Clearly there are emergent social phenomena to be explained here. However, they are best understood in terms of Wilson's (2004) SMH. Herding behaviors are likely to emerge in groups because there are dumb mechanisms that are triggered only when these organisms find themselves in certain sorts of situations. 'Herd mentality' fails to be cognitive in precisely the same way that the behavior of crowds fails to be cognitive. What, then, are we to say of even simpler systems such as eusocial insects?

5.2.3 Hive mentality: Philosophers have often seen eusocial insects, such as bees, as an interesting test case for the viability of collective mentality. Perhaps the most famous use of bees comes from the scathing satire 18th Century English politics in Bernard de Mandeville's (1728/1962) *The grumbling hive, or Knaves turn'd honest*. Mandeville takes political society to be analogous to a beehive in which the selfish interests of the individual bees are aggregated in such a way that they tend towards the 'common good' of the hive. While I'm not concerned with the details of Mandeville's argument, I am inclined to think that there is much to be learned by thinking about bees. In a series of recent papers (many of which are reviewed in Seeley 1995), Thomas Seeley has argued that colonies of honeybees should be seen as a unified system with a rich functional organization. He claims that this functional organization allows for the propagation of representations between bees in a way that allows the hive to respond to changing environmental stimuli. There is thus, some reason to think that colonies of bees might be genuinely cognitive systems.

Seeley begins by noting that there is a growing consensus among biologists that colonies of eusocial insects can legitimately be treated as systems for the purposes of

biological research. Even Richard Dawkins (1989), who is a rabid 'smallist' about explanation, concedes that honeybees possess sufficient functional organization to qualify as vehicles for natural selection. Building on this suggestion, Seeley and his colleagues argue that colonies of honeybees have a cognitive life much richer than that of the individual bees that constitute a colony. They focus on three sorts collective states in defending the view that colonies of bees have mental states: foraging behavior (as a sort of collective sensation), the coordination of pollen processing and foraging, and the process of nest selection. I'll briefly canvas Seeley's data regarding each of these phenomena before turning to an analysis of whether Seeley (1995) is right to claim that honeybees possess a 'hive mind'.

Seeley (1993) first turns his attention to the functional specialization of a sub-group of foragers that seem to function as a sensory apparatus for the colony. Seeley (1986, 1992, 1997) monitors the behavior of foraging bees, and finds a wealth of data suggesting that colonies of bees are able to monitor their environment in order to track the location and richness of food sources. Briefly, the process works as follows. A colony of honeybees sends out foragers that act as a diffuse sensory extension of the hive into the environment, extending in numerous directions simultaneously in order to locate food sources.[103] Each of

[103] One might attempt to object at this point that sensation is something that has to be localized to some area internal to the system that is sensing. However, while it may seem initially strange to think of a sensory system that extends into the environment, this strangeness is likely to be an artifact of a misplaced generalization from the senses with which we are most familiar (and even on this point, there is reason to think that we might just be wrong about how our own sensations happen to work—but that's another project). Consider, for example, the echolocation systems used by some species of bats, dolphins, and whales. These organisms are capable of using sounds that extend into the environment in order to assist with navigation and foraging. Perhaps more intriguingly, some species of fish posses the capacity for electroception. Weakly electric fish and duck-billed platypi, for example, actively generate an electric field that extends into the water and they detect distortions in these fields using elecroreceptor organs. This ability allows them to navigate murky waters in which sight and smell are relatively ineffective. To put the point briefly, although human senses are typically understood as passive receivers, there are cases of active sensory apparatus in other species that extended beyond the bounds of the organism into the environment. Given that this is the case, sensation through forager bees to detect the location of food, water, and nesting cites should be no more shocking than the use of electroception to navigate murky waters.

these bees is sent out in a random direction. However, once this first group of 'sensory' foragers maps the surrounding environment, further foragers are allocated to various nectar sources in such a way that the collection of pollen is optimized within the hive. Using this capacity, the colony is able to search for patches of food as far as 10 km away (Seeley 1997, S23), and it is able to accurately find the richest foraging sites within 2 km.

The interesting thing about this process is that the allocation of bees to a particular resource is not determined by any centralized decision making system, but instead, is the result of limited information being consumed by unemployed foragers within the hive (cf., Seeley 1983, 1986, 1997, and Thom et al 2000). As employed foragers return to the hive, they advertise the distance, direction, and quality of a foraging site by way of 'waggle dances'. Rather than attending to all of the bees that happen to be on the dance floor at a particular time, each unemployed forager typically follows only one bees dance (cf., Seeley et al 1991, Seeley and Towne 1992); whether or not an unemployed bee will be recruited to a foraging site is determined by the duration and vivacity of a foragers waggle dance. Those foragers that have visited desirable worksite dance for a longer period of time as well as with more vivacity than those bees that have visited less desirable foraging sites—and in some cases, bees who have visited less desirable sites will fail to dance at all or will stop working all together. This organization, then, allows for a huge amount of sensory information to be distributed across the employed foragers in a way that does not require a centralized decision making structure to allocate unemployed foragers to new foraging sites.

The second sort of phenomena to which Seeley and his colleagues turn is the capacity of a colony to modify its foraging behavior to suit the quantity and quality of food sources

---

Thanks to Rob Wilson for alerting me to the phenomena of electroception in weakly electric fish, as well as for an intriguing discussion about electroception as sensation.

within the range of foraging bees. By modulating the quality and quantity of a pair of artificial food sources, Seeley (1997, S28ff) has shown that honeybee colonies become more selective when food sources are abundant, but in times of scarcity it will allocate foragers to lower profitability nectar sources in order ensure that they will continue to acquire nectar even when resources are scarce. The mechanism by which selectivity occurs is best understood in terms of the threshold level of food quality at which waggle dances occur. When foragers return to the hive, they have to search for a receiver bee that will accept their nectar for storage. When the food sources are sparse and the colony's nectar influx is relatively low, returning foragers find receiver bees quickly and thus have a low dance threshold. In this case, food sources that are less profitable will be exploited. However, when food is abundant in the hive, the search time required to find a receiver bee is longer, and in this case the dance threshold rises. In this case, only highly profitable food sources are exploited. This allows the hive as a whole to allocate resources for the collection of food on the basis of changing facts about the environment, even in the absence of a central processing system dedicated to monitoring the abundance or scarcity of food. To put the point briefly, the colony as a whole is capable of monitoring the relative prevalence of food sources even though none of the individual foragers or receivers is capable of representing this.[104]

Further complications occur, however, when nectar collection and nectar processing are out of synch in a colony. In such cases, foragers returning to a hive will engage in a behavior known as the tremble dance. When a forager finds an incredibly rich food supply, a

[104] Künholz and Seeley (1998) tell a similar story about the control of the collection of water in honeybee colonies. In this case, however, the amount of water than needs to be taken into a hive is at least partially a function of the relative temperature within the hive as well as the number of infant larvae that have recently hatched. When the temperature of the hive increases, or when there are more young to care for, those foragers returning with water have an easier time finding a receiver so will continue to collect water. In times of dangerously high temperatures, foragers dedicated to searching for water will recruit other bees to deal with the dangerous situation.

colony has to be able to boost its nectar collection rate in response to a high quality foraging site, but it also has to increase the rate at which nectar is being processed in order to allow those bees returning from a high quality foraging site to find receivers for their pollen. Thus, when a forager returns from a high quality foraging site and finds that it has an extensive search time for finding a receiver, it engages in a 'tremble dance'. In the presence of tremble dances, unemployed bees working inside the hive, the tremble dance carries the information that they should begin processing nectar; for bees who have been foraging, it carries the information that they should refrain from recruiting additional foragers (hence inhibiting the waggle dances of other bees). The tremble dance is thus used in order to insure that the rate at which pollen is being processed is adequate to the quantity and quality of pollen within range of the colony.

The final range of phenomena to which Seeley and his colleagues (Seeley 2003, Seeley and Buhrman 2001, Seeley and Visscher 2003, Passino and Seeley 2006, and Beekman et al 2006) turn is to the process by which the selection of new nest sites occurs. When a colony of honeybees outgrows its hive (typically in the spring or early summer), the colony will split and half of the bees will swarm and begin to search for a new nest. Initially, the new colony will swarm around a number of tree branches and then send out scouts (approximately five percent of the swarm) to look for a new nesting site (Beekman et al 2006, 162). In the initial phases of searching for a nest, the scouts will typically find a dozen or so potential nest sites—each of which will be evaluated by the scout according to six desiderata: cavity volume; entrance size, height, direction, and proximity to the cavity floor; and presence of combs in the cavity (Seeley and Burkham 2001). As the scout bees return to the hive, they begin to dance in a way that indicates to the swarm these features of the

potential site. The interesting thing to note, however, is that each scout will only dance for one site, and they almost never dance for another site once they have made their initial choice. Here's where things get really interesting.

Although there is no shifting in choices, there will eventually emerge a consensus on one site, and it the swarm reliably chooses the site that best satisfies the six desiderata listed above (rather than the first adequate site, for example). The swarm will only move however, when there is complete consensus on a single site. The question, then, is how is a decision reached if no bees are switching their allegiances. Seeley and his colleagues (Seeley 2003, Seeley and Buhrman 1999, Seeley and Visscher 2003, and Passino and Seeley 2006) have shown that this consensus occurs as follows. First, the initial scouting bees return to the swarm and dance for the site that they have found. Those bees that have found a merely mediocre or passable nest site will dance less vigorously than those bees that have found a high-quality site. This then leads to heavier recruitment for higher-quality nest sites and, eventually, to a cessation of dancing for lower-quality nest sites. To put the point briefly, lower-quality sites lose support until only the highest quality site is being danced for. This then leads to the reliable selection of the highest quality nest site without requiring any of the individual bees to have a broad knowledge of all of the alternative possible nest sites that are under consideration by the swarm.

These phenomena are quite interesting. However, the question is: should these behaviors of honeybee colonies be understood as genuinely cognitive phenomena? Seeley's data gives us very good reason to think that the specialization of function in a honeybee colony facilitates the propagation of representational states (e.g., states that represent the location of nectar, the quality of a foraging site, and the location of a nest site) between bees

with very different functionally specified tasks. Moreover, the ways in which these representations are propagated through a honeybee colony produces a range of emergent states that are more complex than the cognitive states of the individual bees that compose the collectivity. Although none of the bees are capable of comparing the quality of foraging or nesting sites, there are mechanisms at place in the collectivity as a whole that allow for these comparisons. Thus, it is reasonable to say that comparative judgments are realized in the hive as a whole and not just in the aggregation of the individual states.[105]

At this point, it also seems fair to say that honeybee colonies are capable of representing a variety of facts about their *umwelten* in a way that allows them to deal with the pressing problems of a hostile world. More importantly, by positing cognitive states of honeybee colonies, Seeley has been able to explain such diverse phenomena as the decision to build a nest in one site rather than another and the decision to allocate more resources to collecting or storing nectar. These predictions are only possible when the cognitive states are ascribed to the collectivity rather than the individual bees. The choice of a nest site is a striking demonstration of this fact. After all, the colony chooses the best nest site possible even though none of the individuals has the capacity to chose or even represent any of the nest sites as better or worse than any other. It is only through the coordinated activity of a number of bees, and only through the representation of particular facts about particular nest sites across various bees that this capacity can emerge. This gives us good reason to think

[105] One might object that there is nothing more than the aggregation of representations by individual bees plus a mechanism for settling on one option or another. To make this move, however, is problematic when we consider the states of an individual's brain. At one level of explanation, neurons are designed in such a way that they mechanically fire when they are presented with a certain sort of stimulation, and there are mechanisms at play in the human brain for taking the activity of particular neurons and aggregating them so that they produce a certain sort of behavior. However, we need to be careful to note that neurons and populations are the vehicles for particular sorts of representation, and it is only by continually checking outside of the system to see whether these patterns of activity have the function of representing that we can take them to be genuine representations rather than mere recordings. I am inclined to think that if patterns of activity across neurons can mean something, then so can patterns of activity in honeybees.

that there is a sufficient amount of emergent phenomena here to give the collectivity a rich life of its own. However, there are still a couple of lingering questions about whether these emergent phenomena should be seen as genuinely cognitive.

I have claimed that the appeal to cognitive representations in offering a *psychological explanation* of collective behavior require four things. First, I claimed that a genuinely cognitive system must possess internal states that have the function of adjusting the system's behavior in ways that allows it to cope with features of its environment in ways that are not fully determined by the design of the system. Consider the representations that require a decrease in foraging when too much food is coming into a hive too quickly. None of the individual bees represents a need for a decrease in foraging. However, the system is designed to be sensitive to the relation between incoming nectar and nectar storage. When the rate at which nectar is being returned to the hive exceeds the rate at which it is stored, the system is designed to decrease the amount of nectar that is coming into the system. The important thing to notice, however, is that it is not a matter of the absolute quantity of the input or output that is relevant to the decrease in foraging, but a relation between the current state of a honeybee colony and the current state of the foraging sites in the area that modulates the collection of nectar. It is only by way of these internal states that behavior that is sensitive to changes in the environment is produced. However, the way in which the honeybee colony behaves is fully a function of the evolutionary design of a honeybee colony.

Recall the distinction that I borrowed from Dennett (1996) between four sorts of cognitive systems. Although I don't think that we have enough data to decide whether the systems at play in a honeybee colonies will classify them as Darwinian systems (whose behavior is unreflective and static) or Skinnerian systems (that modify their behavior in

response to stimuli by way of some sort of dumb feedback mechanism), what is clear is that the states of the honeybee colonies do not possess the sort of representations that could be used in order to preselect behaviors on the basis of internal models—that is, honeybee colonies are neither Poperian nor Gregorian systems..

Consider the mechanisms by which a honeybee colony represents its environment. Waggle dances record the location and quality of food sources that can and ought to be consumed by unemployed foragers; search time for a receiver indicates the rate at which nectar is being stored in relation to the rate at which it is being collected; and the tremble dances of foragers indicate the presence of a high-quality food source that is not being adequately foraged. While these dances and search times are capable of standing-in for features of the environment (specifically the location of a pollen source and the rate or consumption by the system), they do so only when the system is immediately presented with raw data about the natural environment. The dance times as well as the vigorousness of an individual bees dance are fully determined by features of the world, and the behavior of unemployed bees and collectors are fully determined by the dances of the returning forager bees. It thus seems reasonable to claim that honeybee colonies are incapable of engaging in behavior that is as rich as our own cognitive behavior. But this just raises a question about whether or not decouplability should be seen as a necessary for genuine mentality.

Andy Clark (1997, 144ff) argues that we should not rule out states that are not decouplable from their immediate causes as genuinely representational. He argues that if we do rule such states out, we will also be forced to rule out the population of neurons in the rat's parietal cortex that indicates the direction in which the rat's head is facing. After all, this population of neurons is active just when it is actually detecting the direction in which the

rat's head is facing (which is, incidentally, most of the time). Now, while Clark (1997, 145) is surely right to note that we gain a great deal of explanatory power by treating this population of neurons as representing the position of the rat's head, and that it helps us to understand the flow of information through the rat's cognitive system as a whole, it is not clear to me that we gain the right sort of explanatory power in order to count this as a genuinely cognitive states. We should clearly count these states of the rat's parietal lobe as the sorts of states that are to be studied by cognitive science. However, they don't give us much of anything like the central cognitive systems that are at play in things like belief and desire.

I suggest that although there is no immediate reason to rule out collective mentality in the case of honeybee colonies, it is important to note that the mental life of a honeybee colony is far more impoverished than the mental life of a human. Honeybees engage in cognitive activities directed towards strategic interactions with their environment, however, these states are at best minimally cognitive states.[106] The emergent phenomena in honeybee colonies suggest an interesting range of phenomena to be studied at the level of honeybee

[106] Similarly cognitive phenomena is suggested by the mound building behavior of termites (Bonabeau et al 1998) and in the familiar case of the use of pheromones by ants to mark the trail used in returning from a food source. Here's the way the story goes. When ants forage for food, they typically utilize a random search pattern. However, the collection of food produces behavior that is organized in such a way to make food collection efficient. This efficient behavior, however, emerges from the random search patterns and a pair of simple rules: 1) if you food, deposit a pheromone trail on your way back to the nest, and 2) if you encounter a pheromone trail, follow it to the food source. As should be immediately obvious, these two rules are mutually reinforcing—the ants that are recruited to a food source will also lay down a pheromone trail, this then makes the trail stronger and recruits more ants to the food source. Sumpter (2006, 7) reviews evidence that suggests that ants can use these simple rules to solve the problem of find the shortest route from a food source to the nest. In an experiment (Beckers et al 1992) where ants were provided with a food source and two bridges of differing lengths that provided paths for returning to the nest, a majority of ants chose the shorter bridge in the majority of cases. Here's the reason: the length of the bridge modulates the strength of the pheromone trail. The pheromones are continually evaporating on both trails; however, the longer trip time means that the it will take more time to get to the food source and so the reinforcement of the trail will be a slower process and "when trail following ants make the choice between two bridges they detect a higher concentration of pheromone on" the shorter bridge, thus reinforcing the shorter trail even more strongly (Sumpter 2006, 7). Similar results explain the production of trail networks by army ants engaged in a raid (cf., Deneubourg et al. 1989).

colonies. Moreover, it's an interesting range of phenomena to be explained from within the cognitive sciences. However, these states are cognitive states only to the extent that the states of the neurons in the rat's parietal cortex are cognitive states. That is, there is a range of explanatory projects within the cognitive sciences that should focus on such states; however, psychology is better served by dissociating such states from the core cases of cognition. They are clearly states that are important for explaining behavior, but we need to be sure not to get carried away in ascribing such things as beliefs and desires to honeybee colonies. I am thus inclined to count such states as cognitive states but only with the recognition that they are more like the states of a rat's parietal lobe of the states of my visual cortex than they are like core cases of intentional phenomena like beliefs, desires, hopes, wishes, and dreams.

5.3 Multi-level selection theory and cognition in human groups

As I have argued, genuinely cognitive systems possess homuncular decompositions into functionally specialized subroutines that propagate representational states across a variety of representational media. Thus far I have focused on self-organizing systems. However, I now turn to another plausible model for establishing the actuality of collective mentality: multi-level selection.

In the previous section, I noted that many people think of eusocial insects as vehicles of natural selection. However, even biologists who are willing to allow for multi-level selection with eusocial insects are typically *unwilling* to apply such a model to more complex organisms. As Eliot Sober and David Sloan Wilson (1998, 338) put the point, whatever it might be convenient to say about bees, the existence of superorganisms is typically seen as a dead issue and its death is seen as one of the greatest achievements of mid-twentieth century

biology. There are, however, holdouts against this claim. D.S. Wilson has been the most visible of these defenders of multi-level selection within biology. Wilson has spent most of his career arguing for a version of multi-level selection theory and in a series of recent papers, he argues that once we have an adequate understanding of multi-level selection, and once we adequately understand the role of evolutionary mechanisms in explaining cognitive capabilities, it will become clear that these mechanisms can underwrite the sort of functional specialization required for collective mentality.

Wilson's argument runs as follows. Any traits for which an evolutionary story is adequate will be best understood as an adaptation to a biological system's environment, and a trait is adaptive just in case it has the function, within the overall structure of the biological system, of increasing the relative fitness of this system (i.e., of increasing its capacity to cope with some significant feature of its environment relative to other systems with which it is competing). What this means for multi-level selection is that if there are group traits that are best understood as adaptations, they will have to have the function of increasing the fitness of that collectivity relative to the other collectivities with which it is competing. Now, biologists have typically averaged relative fitness across groups, focusing exclusively on the genotypic traits that are common across populations. However, Wilson claims that in order to give a complete and adequate explanation of many biological traits, it is necessary to explain not only *relative fitness within a group* but also *the relative fitness of groups relative to one another*. Although some traits are disadvantageous to particular members of a group, they are adaptive when considered in terms of the functional role that they play within a group.

The clearest example of a trait that is adaptive for groups but not for all of the individuals that compose that group is a trait that results in biological altruism. For example,

when a vervet monkey offers an alarm call after seeing an eagle or a leopard, it is putting itself in danger in order to warn its troop of the impending danger. If individual selection offered an exhaustive explanation of vervet monkey behavior, then such warning behavior would be anomalous. Because warning behavior is dangerous for the monkey uttering the call, and because hiding from the eagle or leopard would provide more selective advantage than making a call, we should expect alarm calls to be weeded out of a population relatively rapidly. In accounting for behavior such as the alarm call of vervet monkeys, D.S. Wilson introduces the idea of a trait group. D.S. Wilson claims that for the purposes of evolution by natural selection, it is not the boundaries of a body that proves significant but sharing a common fate that determines the unit of selection (D.S. Wilson and Sober 1989). In the case of warning behavior, a troop of monkeys shares a common fate for purposes of avoiding predators, and a troop of monkeys that has some monkeys that are fill the functional role of sentinels will fare better at avoiding predators than will a troop of monkeys that lacks such monkeys.

To make sense of this idea of sharing a common fate, D.S. Wilson (1975) introduces the idea of a trait-group. Trait-groups are defined in terms of individuals that interact in order to achieve a particular goal (D.S. Wilson 2002, 15). Rather than categorically rejection group selection by starting with the capacities of single individuals and trying to explain why a particular collective behavior would emerge on the basis of the interactions of individuals, D.S. Wilson claims that evolutionary theory requires the units of selection to be evaluated on a case-by-case basis. In some cases, a particular behavior is advantageous at the level of between-group selection even though it is disadvantageous at the level of within-group selection. The primary claim behind multi-level selection is that we should be willing to posit

group-selection in those cases where the selective advantage of a behavior is best explained by appeal to between-group selection, as is the case in the vervet monkey alarm calls. This model, thus, shares a number of important methodological characteristics with the model of collective mentality that I have been suggesting.

Building upon this account of multi-level selection, D.S. Wilson argues that because psychological traits are biological phenomena, we should expect that some psychological traits would be explained in terms of group-selection rather than individual selection. In short, the claim is that "groups can also evolve in adaptive units with respect to cognitive activities such as decision making, memory, and learning" (D.S. Wilson 1997a, S128). Building on insights derived from evolutionary psychology, D.S. Wilson claims that evolutionary pressures would be significant for determining how people will evaluate information. He claims that the sorts of decision processes in which people typically engage are likely to be significant not just at the individual-level but also at the level of social groups. Now, since there will likely be a number of important consequences for both individuals and the groups of which they are a part, "it is likely that the psychology of decision making has been strongly shaped by natural selection at both the individual and group levels" (D.S. Wilson 1997, 346). Perhaps more importantly, D.S. Wilson claims that there may even be cases in which groups are so integrated and the contributions to particular goals so partial that "the group could literally be said to have a mind in a way that the individuals do not, just as brains have a mind in a way that neurons do not.

Having developed a theoretical apparatus for the possibility of collective mentality, D.S. Wilson then argues that his model is satisfied in a number of cases. D.S. Wilson (1997, 358) begins his analysis of collective cognition within the multi-level selection framework by

distinguishing between two hypotheses about cognitive cooperation. The first hypothesis is that there are cases in which individuals act as individual decision makers with the goal of making a decision that is good for the group. The second hypothesis is that there are cases in which an individual's cognitive activity is important precisely *because of* the functional role that it fills in contributing to the cognitive activity of a collectivity. However, as R.A. Wilson (2001, S268) notes, only the second hypothesis will provide evidence for the existence of collective cognition. After all, the mere fact that individuals are capable of working together to solve problems does not provide evidence that they are thinking as a group. As I have argued throughout this thesis, only a system that consists of functionally organized components dedicated to collective computation will count as a genuinely cognitive collective system. The question, then, is whether there is any data on cognitive cooperation that lends credence to the latter hypothesis.

D.S. Wilson et al (2000) suggest that one promising place to look for cognitive cooperation is in the role of gossip in groups. By engaging in gossip, groups police their members and insure that people do not defect from their assigned role. The hypothesis is that a person who fears being gossiped about if she defects from her social role is less likely to defect than a person who does not share this fear. In order to test this hypothesis, Wilson et al (2000) used a survey style experiment to test people's intuitions about the normative status of gossip. Wilson et al (2000) found that although people were highly critical of self-serving gossip, they thought that gossip was acceptable in cases where it is directed towards a norm violation (provided that the gossip occurs in a responsible manner). Building upon this data, Kevin Kniffin and D.S. Wilson (2004) decided to study the use of gossip in a more ecologically valid situation. They studied the use of gossip among the members of a

University rowing team, using a voice-activated tape recorder, carried by the first author when he was a member of the team, to track the content of conversations between team members. Kniffin and Wilson (2004) found that gossip served the function *within the rowing team* of enhancing conformity with the norms of the group. They then claim that because the rowing team shares a common fate with regard to the task of rowing, and because the use of gossip helps the group to satisfy its goals, this gossip should be seen as a sort of collective cognition.

However, it is not clear that this data demonstrates anything more than Wilson's (1997) first hypothesis: that there are cases in which individuals act as individual decision makers with the goal of making a decision that is good for the group. Clearly, the use of gossip to ensure cooperation in accordance with the goals of a group are significant for the success of a group such as a rowing team, and it is quite likely that gossip can play an important role in many other sorts of social groups as well. Moreover, Wilson might even be right that the cohesiveness exhibited by groups that engage in gossip will be likely to allow these groups to outperform other groups that do not engage in gossip (or something else that is functionally equivalent for norm enforcement). In fact, it is a near truism of social group-selection (cf., Boyd and Richerson 2005) that some such structures for ensuring social cohesion, and likely even more complicated structures of punishment and meta-punishment, are a necessary condition on cooperative activity. However, the mere fact that people are working together does not tell us anything about whether or not they are thinking as a group. That is, the mere fact that a group that is constrained to act in accordance with a system of norms does not yet give us reason to think that we have found a case of genuinely collective

cognition. However, D.S. Wilson does have more data that he takes to lend credence to the hypothesis of collective cognition.

Wilson et al (2003) ran two experiments on cognitive cooperation, based on the game '20 questions'. In the first experiment, they set up a variety of more and less difficult games about job titles by modulating the obscurity of the word to be guessed. In a first condition, they assigned volunteers either to play the game alone or to play as part of a same-sex group consisting of 3 members. In a second condition, those volunteers who had been in groups played alone and those who had played alone played as members of same-sex groups consisting of 3 members. Wilson et al (2003) found that in games that were solved quickly, there was little difference between individuals and groups. However, as more questions needed to be asked in order to solve a game, groups begin to solve more games than even the best individuals. Wilson et al (2003) hypothesize that the increased memory load required to remember which questions have been asked, as well as the increased computations required to recognize which options had been closed off, allow groups to outperform individuals in more difficult games.

In order to demonstrate that the mechanism at play in improving performance for groups was not merely an artifact of the number of people engaged in a particular game, Wilson et al (2003) ran a second experiment in which volunteers were asked to think of as many job titles as they could. In a second condition, volunteers were then provided with a partially completed game of 20 questions in which 7 questions had been asked and were asked to think of as many jobs as possible that had not been ruled out; volunteers were also assigned to either 1) a nominal group, in which two people worked alone but their answers were combined, or 2) a real group, in which two people worked together. Wilson et al (2003)

found that there was no statistically significant difference in performance between nominal and real groups in the first condition. However, in the second, more difficult, condition Wilson et al (2003, 237) found that real groups had a performance advantage of approximately 50% over nominal groups (94.8 vs. 60.8 items recalled, p = 0. 003). Based on this data, Wilson et al (2003) argue that the value of cognitive cooperation is most pronounced in those cases where task difficulty exceeds the cognitive abilities of single individuals. Importantly, this is precisely what is predicted by the model offered by D.S. Wilson (1997) in his defense of the multi-level selection of cognitive adaptations. The question, however, is whether this sort of data demonstrates that there actually are cases in which an individual's cognitive activity is important precisely *because of* the functional role that it fills in contributing to the cognitive activity of a collectivity.

Unfortunately, D.S. Wilson provides no reason to think that the sort of functional integration required for something to be a genuinely cognitive system are present in the cases that he addresses. Although it is clear that *the individuals* are using representational states, it is not clear that these representational states are being used in the production of genuinely collective representations. A plausible reading of the results collected by D.S. Wilson and his colleagues suggests that there are interesting facts about the way in which individuals behave when they are members of groups, and more importantly, the differences exhibited by individuals within collectivities could indeed provide selective advantage within the context of between-group selection. However, as R.A. Wilson's SMH suggests, the mere fact that there are numerous cognitive phenomena that emerge only within the context of groups is not itself sufficient to demonstrate the existence of collective mentality. The collectivities with which D.S. Wilson is concerned may indeed possess internal states with the function of

adjusting the system's behavior in ways that allow it to cope with features of its environment, and these states may also be capable of standing-in for features of the environment that are important to the system. However, until we have a better understanding of what sorts of mechanisms are sufficient to ground the functional specialization required for genuinely collective mentality, I am, at least at this point, unwilling to attribute collective mentality to the systems with which D.S. Wilson is concerned. Fortunately, there is a promising attempt to explain these mechanisms, and it is to this attempt that I now turn.

5.4 Transactive memories and the group mind

The most promising attempt to resuscitate the idea of a group mind from within the discipline of psychology has been developed by Daniel Wegner and colleagues (Wegner 1986 and 1995; Wegner and Wegner 1995; Wegner et al 1985) in a series of papers on 'transactive memory'. Although memory is typically understood individualistically, Wegner argues that some groups form memory systems in which each person in the group possesses only a subset of the information relative to the to the activity of the group, but through the coordinated activity of the person's that compose a collectivity, the collectivity as a whole can remember things that the individuals alone cannot (Wegner et al 1985, 256). Each individual has "in internal storage many items, labels, and locations, and knowing that the locations are in the other's memory" (1986, 189-190). Wegner begins by thinking about how this system could be similar to a memory system in an individual.

Individual memory is often divided into three stages. First, perceptual inputs must be *encoded* as discrete representations that are 'understood' by a system as having a particular content; they are then *stored* in such a way that they can later be *retrieved*. Making this

distinction, however, leads to a worry about how we are able to store information in such a way that retrieval will be fast and accurate enough to facilitate practical activity in an ever-changing world. One promising answer to this question is to note that memories are not stored individually and separately from one another, but are organized in various associative networks. The thought is that by encoding information in such a way that it is already sorted topically, the mechanisms dedicated to recall won't have to search the entire memory system but will only have to search within a particular topic. This, however, raises a question about how the recall mechanisms could know where to search. Wegner (1986) claims that the most promising theory of retrieval is grounded on the idea that retrieval mechanisms contain a system of metamemories, which are to be understood as directories of memories indicating the location of a particular sort of information. This is where things get interesting.

While it might be true that memory processes depend crucially on neurological structures, there's no reason to suppose that the information that's been encoded and stored in memory couldn't be located externally to the system possessing a particular metamemory. On the basis of the claim that memories and metamemories can be located in different systems, Wegner takes transactive memory to be the logical development of adopting a computational theory of memory. Wegner's key claim is that on the assumption that our best understanding of individual cognition is computational, we ought to understand at least some sorts of social groups as computational networks (Wegner 1995, 319). Wegner (1995) elaborates on this suggestion by noting that networking a set of computers is often achieved by duplicating directories on all of the machines while physically locating the information specified by the directory on only one of the machines. This allows each computer in the network to make use of a virtual memory that spans across all of the machines, thus allowing

for increased speed in processing and decreased load on the memory for every particular machine, without a decrease in the number of tasks that can be executed by the network. Such a system requires that the various machines in the network be able to 1) update their directories without accessing all of the memory items on the various computers, 2) ensure that information is allocated to the various machines in such a way that the information does not become excessively redundant across the machines, and 3) ensure that the information that is spatially distributed is accessible to any machine that might happen to need that information (Wegner 1995, 324-326). The key question for this analogy, then, is whether there are analogous systems in place for purported cases of human transactive memory.

To begin with, there are a number of different ways in which such metamemories can refer to memories located in other systems. Default assumptions based on morphology and surface characteristics of another person (e.g., stereotypes formed on the basis of the perceived gender, race, or class) are often used as a starting point for determining which persons in a collectivity ought to be responsible for a particular range of information (1995, 327). However, such assumptions are often unwarranted, and as a collectivity develops the allocation of information begins to develop as well. Assumptions made on the basis of things we come to know about a particular person often play a key role in deciding who will be likely to store a particular range of information. In many cases, collectivities end up with ranges of specialization that are explicitly negotiated (Wegner 1995, 327), producing metamemories through explicit planning about what each person should focus on (e.g., you remember the first four digits of the pass code, I'll remember the next four). Other times, the metamemories are classified merely on the basis of a perceived expertise grounded in the interaction of individuals within a group (e.g., Dylan always remembers Fodor's arguments,

Felipe remembers Searle's, and Jacek remembers Jackson's). Such implicit judgments of expertise can take place either on the basis of functional specialization required by the structure of a particular organization or on the basis of quick judgments on the basis of paradigmatic cases of recall (Wegner 1995, 327).

In all three cases, methods of information allocation produce a differentiated transactive structure that contain a lot of overlap in general information about who is likely to do what, but that reserves the particular details of a particular category for one persons memory alone (Wegner et al 1985, 264-65). These metamemories can then be updated on the basis of assumptions about which person is likely to be the specialist on some particular topic. Although this is not always a successful way to engage in the updating of memories (especially when you've made an unwarranted assumption about who is likely to specialize in remembering a particular sort of information on the basis of an ungrounded prejudice), it does, by and large, allow for the successful navigation of our social world. More importantly, assumptions about where a particular sort of information is likely to be located that are built up on the basis of explicit or implicit negotiations about who is to take care of a particular sort of information are likely to be far more successful.

In order to demonstrate how this computational model of transactive memory works, Wegner must demonstrate that a person can retrieve the information in another person's memory, and this has to be possible *because* she knows that this other person is the location of a piece of information with a certain label. Wegner (1995, 334) claims that the first step of retrieving a memory occurs when one checks to see if this is the sort of information that they remember. If it fails to be the case that this is something that you are supposed to remember, you check to see if there is someone else in the group who is supposed to

remember this sort of information. When you look elsewhere you ask the person who is supposed to have that memory and then you deploy that information to solve the practical problem at hand. This system is, however, occasionally undercut by an immediate attempt to look elsewhere. In these cases, there are triggers (e.g., if I am asked about numerosities I'll won't feel that I know about such things, however, I'll be immediately drawn towards asking Mr. Numberbaum) that indicate that this is the sort of information that is stored in some other location.

This provides a theoretical foundation for a theory of collective memory. We begin by recognizing that memory has to be divided into three discrete sorts of processes. We then see that there are many things that can fulfill the functional role of storage. One way in which memories can be stored is in other people's heads—they form a sort of external hard drive over which we have limited access. Given that there are ways of accessing the information that is stored in another person's head, and given that there are metamemory systems in place that can recognize the information that is located in another person, we have good reason to think that other people can act as an external storage devise for one another. Now, if there are memory tasks on which the groups can outperform the individuals that compose these groups, we'll have reason to think that these systems are acting as a single cognitive system. And fortunately, there are such data in the offering.

Wegner et al (1991) report the results of a study of 118 individuals in close dating relationships. Pairs of subjects were asked to remember a list of 34 items from seven different categories. These pairs were either the dating pair (natural couple condition) or they were randomly assigned opposite-sex couples (impromptu couple condition). These couples were then randomly assigned to a condition in which areas of expertise were assigned (i.e.,

each person in the assigned expertise condition was told to remember items from a specific category) or not assigned (i.e., each person in the unassigned expertise condition was allowed to focus on whichever information she or he could remember more easily). Wegner et al (1991, 926) found that in the unassigned expertise conditions natural couples (M=31.40) remembered more items than did impromptu couples (M=27.64). Moreover, the items that were remembered were, for the most part not overlapping.[107] However, in the assigned expertise condition impromptu couples (M=30.14) *remembered more items than* natural couples (M=23.75)! This result initially looks somewhat surprising. However, this data lends credence to the value of functional specialization.

Natural couples seem to develop a transactive memory system without being directed to do so. This system then facilitates improved performance on this sort of memory task. Here's what appears to be happening. Very early in relationships, individuals within couples begin to realize that their partners specialize in retaining certain sorts of memories and that they specialize in retaining others. On the assumption that the relationship will last a long time,[108] and that they will be able to act *as a couple*, people in relationships begin to rely on their partner to retain some or their memories. However, the interference produced by the introduction of a new functional architecture that's produced by assigning expertise prevents the natural memory structure that has emerged in the natural couples from being used. More importantly, focusing on remembering the areas that you are supposed to focus on in the experiment takes an added toll on the subjects because they have to remember not to remember the things that they would usually remember in this couple. This provides us with

[107] For overlapping memories (M=5.28) for non-overlapping memories (M=22.8)

[108] Wegner et al (1991, 925) report that 52.5% of the subjects believed that their relationship would last forever. An additional 31.4% of the subjects believed that the relationship would continue for some time.

very good reason for thinking that transactive memories tend to emerge in close couples. However, there are serious questions about how much further such transactive memory systems can be extended.

In an interesting extension of the plausibility of transactive memory, Liang, Moreland, and Argote (1995) investigated the practical implications of Wegner's claims about transactive memories for group performance. In the first phase of their experiment, they trained subjects in small groups, consisting of three people, to assemble transistor radios. In the second phase of their experiment, subjects were later asked to build the same sorts of radios, either in the same group where they were initially working or in a new three-person group. Liang, Moreland, and Argote (1995) found that groups who trained together were better able to recall the assembly procedure and were thus able to build better radios. Moreover, when they coded videotapes of each of the work groups from the second phase of the experiment (in a third phase of the experiment), they found that the improvement in recall and performance occurred primarily by way of the functional specialization that is predicted by Wegner's model of transactive memory.

There are a number of ways in which these results could be explained by appeal to the individual states rather than the states of the collectivities. For example, recall that Robert Wilson's (2004) SMH claims that there are likely to be state of the individuals that compose these work groups that will only be manifested when these individuals are part of the relevant collectivity. The question, then, is what sorts of states are these likely to be and why are they best understood as exclusively states of the individuals rather than states of the group that they compose? At this point, the most plausible response is to say that the structures of communication that obtain between the members of these groups are more highly developed

in groups that have trained together. People who have trained together know what sorts of questions to ask of one another in order to more effectively attack the problem at hand. If this were they case, then the increased capacity for communication, rather than the functional specialization *within the group* would be a much more plausible explanation of the phenomena—and the theoretical virtues of appealing to transactive memories as collective mental states would dissipate.

Fortunately, however, this sort of explanation turns out not to be as plausible as it first seems. Moreland and Myaskovsky (2000) demonstrate that functional specialization within a group, in which each member of a group is responsible for retaining a particular range of memories, is responsible for the improved performance of groups in such tasks. They find no significant difference between the performance of groups that were trained together and those who were given handouts specifying the tasks that would be performed by their group mates—but each of these sorts of groups continued to have an edge on those groups in which such specialization of memory structures was established.

The question, then, is how well transactive memory meets the desiderata that I have laid out for cases of collective mentality. The project of transactive memory is grounded on the claim that within collectivities, various individuals specialize in the sorts of information that they will remember. This is the sort of cognitive specialization that can facilitate the propagation of representational states across individual memories in order to achieve various sorts of collective cognitive projects. This means that at least the structure of the theory of transactive memory is of the right sort to count as underwriting a case of collective cognition. Moreover, Wegner's data demonstrates that there are states of the transactive memory systems that are not exhaustively described by an appeal to the state of the individuals.

Because of the way in which informational specialization occurs, appealing to the cognitive states of collectivity yields both predictive and explanatory advantage beyond what can be achieved by an appeal to merely individualistic cognitive science.

The collectivities studied by Wegner and his colleagues also appear to have the capacity to be in states that are specified internal to the system that facilitate coping with at least some memory tasks in ways that the individuals who compose the collectivity could not. Moreover, these strategies are not exclusively a function of the design of the system but emerge in the interaction between the individuals who compose the couples. Given that these states of the collectivities are memories, they are (almost by definition) capable of standing in for various features of the collectivities' environments even in the absence of immediate environmental stimuli, and these transactive memories form larger representational schemes that allows a variety of possible memory contents to be represented. Finally, there are indeed proper and improper ways of producing, maintaining, modifying, and using the various memories under various conditions. However, there is a bit more to say on this point.

In claiming that transactive memories should be considered genuinely cognitive states of collectivities, I am committed to the claim that there will be ways in which the transactive memory system can fail to function properly for the systems in question. Fortunately, there are a number of places were we start to see evidence of the improper functioning of transactive memory systems. To begin with, there are some cases in which we find transactive memory systems that contain incomplete specifications of the relevant pathway information about who's responsible for what. Such incomplete pathway information can easily lead to new sources of error within the group (1986, 198).[109] Moreover, unwarranted

[109] Suppose Mark makes a mean martini; however, when Margaret, the mistress of margaritas, mojitos, and martinis was hired at the bar, he began to focus on Manhattans—he thought that any time he needed to make a

'feelings of knowing' can occur when an individual overestimates what other people in the group are likely to remember.[110] Finally, when one of the members of a small group leaves, this can leave metamemories without access to the locations where the relevant memories are stored. This can produce all sorts of failures of practical activity by the collectivity. Things that the collectivity used to be able to do with ease will now be either much harder or maybe even impossible to do.

Given that transactive memory systems meet the criteria for collective mentality that I have discussed in this thesis, Wegner's claim that transactive memory provides a new foundation for claims about the group mind are well placed. However, there are still significant questions about the value of such structures. Wegner's data focuses primarily on two-person heterosexual couples in close romantic relationships. Liang, Moreland, and Argote (1995) and Moreland and Myaskovsky (2000) focus only on very small work groups. I contented that the success of transactive memory for these sorts of small groups provides us with good reason to pursue research on various other sorts of small groups, as well as on larger groups such as laboratories, corporations, and philosophy faculties.

5.5 Distributed Cognition:

While the experiments developed by Wegner and his colleagues establish the functional specialization required for collective mentality in small groups, extending these

---

martini, he could get Margaret's advice. However, if Mark's metamemory fails to specify Margaret as the one who possesses memories relevant to making masterful martinis, there is a real chance that Mark's martinis will be merely mediocre.

[110] Suppose Tracy thinks that Theodore knows how to change a tire. When her mother asks if they'll be safe on their drive to Tuscaloosa, she might believe that they know how to deal with any difficulties that they might encounter. However, if Theodore thinks that Tracy is trained in all things automotive, when the tire blows in Twin Falls, the two of them will be in real trouble.

methods to larger groups proves incredibly difficult. As we turn to larger groups, the methods required become far more theoretical and far more ethnographic. Building on the insights concerning the use of external representations developed within the sociology of science (e.g., Latour 1999 and Latour and Woolgar 1979) the method of cognitive anthropology has developed in order to lend credence to the existence of distributed cognition. According to proponents of distributed cognition, distributed and collective systems are capable of possessing cognitive properties that differ from the cognitive properties of the units out of which they are composed, and no matter how much we know about the properties of components, we cannot infer the cognitive properties of the collectivity (Hutchins, 1995b).

When researchers study such distributed systems, however, there are a number of things that they must keep in mind. First (following Kirsch 2006), the study of distributed cognition must focus on the variety of ways in which coordination is possible within groups of humans. As should be familiar to anyone who has engaged in collective deliberation, the members of a group bring a wide variety of beliefs, beliefs, and goals to the deliberation—and it is rarely the case that all of these mental states are consistent with one another. For this reason, the "key question which the theory of distributed cognition endeavors to answer is how the elements and components in a distributed system—people, tools, forms, equipment, maps and less obvious resources—can be coordinated well enough to allow the system to accomplish its tasks" (Kirsch 2006, 258). Although it is not likely that all groups will be organized in such a way that they allow for such coordination, there are a number of promising cases of such interdependence, and cognitive scientists have recently shown growing interest in these systems.

I have already discussed two interesting cases from the literature on distributed cognition. First, I considered Hutchins' account of the "fix cycle" used by contemporary navigation crews to establish location and compute the trajectory of a naval vessel (Hutchins, 1995). As I noted in the previous chapter, some of the representations used in this computation are internal to the individual crewmembers and others are external representations conveyed from individual to individual in the service of the collective cognitive task of constructing a representation that directs the behavior of the ship. Because of the training that these crewmembers receive, the representations produced by various different subsystems are capable of being understood only by those who are trained to take measurements using a particular device. Thus, none of the crewmembers working on their own particular tasks is capable of representing the position of the ship by herself. It's only the output of the navigation crew as a whole that is capable of representing the location of the ship.

The second case is crime scene investigation (CSI). As I noted in the previous chapter, in CSI, "evidence is likely to be collected by one group of people, analyzed by another, and interpreted and presented to Court by another group" (Barber et al, 2006, p. 358). Evidence must be analyzed to determine whether there's sufficient evidence to prosecute and must then be converted into a narrative structure in order to facilitate prosecution. This narrative structure, however, is just the end result of a complex. In this case too, the various investigators and interpreters only need to know what to do when they are presented with a range of conditions in their immediate environment. However, through the interaction of systems concerned only with local information, a narrative emerges that facilitates the achievement of the goal of prosecuting a particular person.

In the remainder of this section I turn to some other ways in which distributed cognition has been developed in the service of establishing the actuality of collective mentality.

5.5.1 Distributed assessment systems: In a recent paper, Christophe Heintz (2006) has argued for the existence of two sorts of collective systems dedicated to the assessment of the quality of a piece of work. First he considers the system dedicated to assessing the quality of a particular Web page. This system consists of the community of Web-users who link documents and a search engine, such as Google. Heintz argues that these subcomponents of the assessment system are functionally specialized for particular tasks that contribute to the efficient use of cognitive resources for Web users. "In these cognitive systems, the cognitive task of Web users (as authors of Web pages) is to assess Web documents, and the cognitive task of the search engine is to compile these assessments an produce a usable representation of the result" (Heintz 2006, 387). Consider the way in which Google operates.

In ranking the results of a search, Google takes as input the linking behavior of Web-users. The more links that exist for a particular page, the more highly ranked that page will be. Heintz (2006, 388) claims that it is a consequence of this design "that search engines together with web-users constitute a distributed cognitive system for the attribution of reputation, visibility, and, eventually, credibility." Although it could easily be the case that none of the individual authors of Web pages would rank pages in the way that Google ranks them, through the interaction of these various individuals and the algorithm utilized by Google, a ranking emerges that many Web users are willing to take as an reliable source of credibility. Because of the way in which the results of such searches are presented by Google, the people who use search engines are able to adopt as sort of "bounded rationality,

which relies on a simple heuristics with quick halting procedures rather than complying with the demoniac rationality" that we might think would be required for evaluating the reputation of Web pages (Heintz 2006, 398). Web-users recognize that it would be nearly impossible to search through all possible results for a particular search term and they thus offload this sort of evaluation onto the system that consists of other Web-users and Google.

Building on such distributed structures of assessment, Heintz draws an analogy between search engines and considers a second case of the distributed machinery used by scientists to determine the credibility of one another. Scientists evaluate the reputation, visibility, and credibility of one another by appeal to publication record, academic home, collaborators, and mentors. Moreover, individual research often depends to a significant extent on collective judgments about which articles are important, which articles are of a high quality, and which articles are relevant to a particular project. In order to determine which articles ought to be read, scientists often appeal to the community's judgments about the reputation of particular authors and particular journals. In most cases, the evaluative judgments of the scientific community are more determinative of what a person will read than are her own interests and judgments. Heintz thus argues that the "evaluation of a scientist could not be brought about by a single person: every scientist goes through a very large number of assessments, which stretch over a whole career, and may require different kinds of specific expertise. Thus, the evaluative process in science is fundamentally distributed" (Heintz 2006, 402). The question, now, is whether this sort of distribution and functional specialization is sufficient to qualify such structures of assessment as cases of collective mentality.

To begin with the specialization of function in both the case of the assessment of the

quality of Web pages and the assessment of scientific data does facilitate the propagation of representational states across representational media. However, it is unclear how such states are more than a mere aggregation of the individual cognitive states. In fact, the algorithm designed by employees of Google is designed to do nothing more than to look for the most frequently cited Web pages. This is precisely what allows for the possibility of the 'Googlebomb'.[111] The Google algorithm is merely an aggregative tool that looks for the statistically most common link on public Web pages. Thus, if we had full knowledge of the psychological states of the individuals who were building Web pages and knowledge of the algorithm used by Google, we could produce a ranking exactly like Google's. Admittedly, this would be a hard task. However, because appealing to the states of this system yields no explanatory advantage beyond the advantage of a fully informed cognitive science, such appeals are unwarranted. Something similar is also true of the judgments of reputation used by scientists in evaluating one another. Although the psychological states that are fed into this system are much more complex, they are surely explicable in terms of the aggregated judgments of people within a particular discipline. Although it may be useful to treat the community of scientists as a cognitive system because of our current pragmatic situation, this is likely to be an artifact of our own epistemic imperfections.

5.5.2 Science as distributed cognition: The most promising attempt to extend distributed cognition to larger systems is developed by Karin Knorr Cetina and Ronald Giere to show that some scientific labs should be treated as cognitive systems.[112] In a pair of

[111] The most famous case of a Googlebomb occurred when a savvy Web-user worked out the algorithm used by the Google search engine and built a Web pages in which the words 'miserable failure' were repeatedly linked to the official government Website of George W. Bush. Eventually, there were enough links in the Web page that the first hit on a Google search for the words 'miserable failure' pulled up the Bush Website.

[112] Giere, however, is unwilling to concede to these systems any sort of mental states. His objection is bewildering. It runs as follows. Suppose we attribute knowledge to a collectivity. Some of the computational

ethnographic studies, one focusing on the production of knowledge in a molecular biology lab (MB), the other focusing on a high energy physics (HEP) lab, Knorr Cetina (1999) argues that because of differences in the standards for collaboration in the two fields, HEP labs should count as a single cognitive system but MB labs should be understood as consisting of a number of individual scientists working together as discrete cognitive systems. Building on the insights developed by Hutchins (1995), Giere argues that we should understand the Hubble space telescope system as a single cognitive system. I'll address each of these systems in turn. I'll then argue that both the HEP lab and the Hubble system should count as genuinely cognitive systems in the sense required by my theory of collective mentality.

5.5.2.1 *Epistemic Cultures:* Knorr Cetina (1999) offers ethnographic data, drawing from an extensive study of experiments conducted between 1987 and 1996 at the European Center for Nuclear Research (CERN). The CERN is a massive HEP lab, employing as many as three thousand scientists at a time, with collaboration on particular experiments often occurring between as many as a thousand scientists.[113] Knorr Cetina argues that because of the size, and complexity of the detectors that are used in experiments, the duration of the experiments (some lasting as long 20 years), and because of the degree of collaboration on a

---

systems to which we will want to attribute mental states that are distributed across huge distances both physically and temporally. However, our ordinary concept of a mental state is intimately bound up with our concept of a mind. Our commonsense notion of a mind holds that minds are localized rather than distributed. So, our commonsense understanding of mental states takes these states to be localized rather than distributed. So the states of distributed computational systems cant count as mental states. I have two responses. First, it is not clear to me that our commonsense understanding of minds actually precludes the distribution of mental states across a group (cf., Knobe and Prinz, forthcoming, Arico et al., in prep, and Huebner et al., under review). Secondly, even if our commonsense notion of mental states holds that they must be local rather than distributed, I'm not sure why we should believe commonsense on this point. After all, our best scientifically and philosophically informed theories hold that minds can be distributed—or so I've argued—so, so much the worse for commonsense!

[113] Knorr Cetina notes that when experimental results are published by CERN they list the authors in alphabetical order, without regard to seniority. These lists of authors typically run up to 5 pages of a journal article!

particular experiment, there is good reason to think that the importance of the particular individuals becomes less important than the role that they are playing in the production of a particular piece of scientific knowledge.

In establishing this point, Knorr Cetina (1999, 127) argues that the physiological and psychological differences between individuals are less salient in a HEP lab than they are in most other situations. Her data suggests that members of this lab are not concerned with what their colleagues look like (much of the information that is passed between people is passed in the form of emails and memos) or what they do when they are away from the lab.[114] Knorr Cetina (1999, 128ff) builds her argument for this claim by noting that if there are strong enough structures of social coercion to force a person to see herself, at least for the purposes of her lab, as filling a particular functional role, then there will be good reason to think that, at least for the purposes of working in the lab, these people will begin to occupy those roles in a way that will allow them to adopt particular computational roles in much the same way that the parts of the parts of a computer function together.

Knorr Cetina notes that the structure of the lab is conducive to such a view of the individuals who work there. CERN is divided into a variety of groups, each of which is focused on the collection or evaluation of a particular sort of information. Each group consists of a number of people and a number of devises for measuring or evaluating various sorts of data. The persons who are employed in a group are the only ones that have access to the data that is studied by that group, and they constantly have to appeal to people from other groups in order to obtain information collected by that group. The various groups are functionally specialized to focus only on the collection or evaluation of a particular sort of

[114] Knorr Cetina (1999, 328) reports an interaction with one physicist in which her informant reported never having been asked anything about her personal life in the three years that she had been employed at CERN

data, and it is only through the transmission of information from one group to another that the results obtained by one group can be coordinated in order to produce anything worthy of being called an experimental result (Knorr Cetina 1999, 129).

This structure emerges through the regimes of trust that develop at CERN, and this causes CERN to have a democratic structure in which authority is necessarily distributed. In the experiments that are carried out at CERN, data that is passed from one group to another will only be taken seriously if it viewed as being passed along by an expert. However, because of the size of CERN and the diversity of the data that is being collected and interpreted, this expertise simply cannot be centralized. There is no one at CERN who knows everything that needs to be known in order to carry out any of the experiments that are done at CERN. This enforces a 'management by content' in which the most important and experienced experimenters coordinate the information produced by their group rather than determining what ought to be done within that group. "What gets done, and when, depends mostly on the technical problems that need to be solved to achieve the goal of a meaningful and reliable result" (Giere 2002c, 2-3). More importantly, the structures of trust that underlie the transmission of information from one group to another are kept in place by a sort of professional gossip. As D.S. Wilson aptly notes, the use of gossip can play an incredibly important role in stabilizing functional roles within a group. If a group contains members that are not fulfilling the roles that they are supposed to fulfill in collecting or interpreting data, for example if they are ore concerned with their own research than the collective research in which they are engaged, members of other groups will gossip in such a way that suggests that these people and their groups are not to be trusted in producing collaborative data (Knorr Cetina 1999, 201ff). This sort of criticism plays a very strong role in insuring that various

members of a the HEP lab fill the roles that they are meant to fill rather than worrying about their own personal interests. To put it bluntly, as Knorr Cetina (1999, 25) does, in a HEP laboratory such as CERN, "the subjectivity of participants is put on the line—and quite successfully replaced by something like distributed cognition"

Having laid out her account of a HEP lab as a sort of collective cognitive system, Knorr Cetina turns to the analysis of a MB lab. In analyzing a MB lab, Knorr Cetina (1999, 217) finds that "the person remains the epistemic subject" such that "laboratory, experimentation, procedures, and objects obtain their identity through individuals. The individual scientist is their intermediary—their organizing principle in the flesh, to whom all things revert." Knorr Cetina argues that because of the way in which publishing in MB is conceived, individuals are forced to focus on their own research projects rather than focusing on anything collective. In MB, as with many of the sciences but as opposed to HEP, an individual is credited with discovering an experimental result only if she is first author on a paper. Rather than developing a community of trust in which individuals rely on one another for the acquisition of various sorts of information, MB produces collaborations that are often tenuous. Each scientist has her own project, and although there may be some overarching goal toward which the lab as a whole is dedicated, collaboration takes a back seat to individual achievement. Thus, Knorr Cetina argues that there is no room for genuinely distributed cognition to emerge in a MB lab.

5.5.2.2 The Hubble space telescope: Building upon a long-standing interest in the cognitive science of science, Giere (2002a, 2002b, 2002c, 2003, and 2004) has focused, in a series of recent papers, on the way in which the team of scientist interpreting data from the Hubble space telescope, when coupled with their technological apparatuses, ought to be

understood as a unified cognitive system from the standpoint of a cognitive science of distributed systems. The problem with which Giere is concerned is that when we learn that the Hubble team has come to some interesting conclusion, we want to know how it is that we can "understand the process leading to the conclusion about 13 billion old galaxies as a *cognitive* process." The standard answer that would be offered by a cognitive scientist of science would be that there are a number of individual cognitive agents, each of whom has a series of symbolic representations (presumably in a language of thought) over which she can run inferences in order to come to some conclusion about the 13 billion year old galaxies (Giere 2003, 2pdf). However, as Giere correctly recognizes:

> There are a number of difficulties that arise when one attempts to apply this paradigm to the Hubble System. One is locating the cognitive agent that acquires the representations and does the computations. The difficulty is not that there are no agents to be found. Rather, there seem to be too many agents. There is a whole team of people who control the movements of the telescope in space. Then there are whole teams of people at the Data Operations Control Center, the Data Capture Facility, and the Space Telescope Science Institute. And of course there are computers all over the place. One thing is clear. There is no one person that can be identified as *the* cognitive agent acquiring the representations and doing the computations. (2003, Pdf2)

Giere suggests that in order to have an adequate understanding any claim about the knowledge of a 13 billion year old galaxy, we must appeal to a system that contains a number of people an a number of technological apparatuses that are distributed widely over both time and space.[115]

The relevant system for understanding knowledge about the 13 billion year old galaxy consists first of a very complicated telescope (which includes the infamous mirror, a series of electronic detectors that are sensitive to electromagnetic radiation, and an onboard computer that organizes and synthesizes the information from these detectors). The information from

[115] The following description of the system is derived from the account in (Giere 2003)

this computer is then sent to a Tracking and Data Relay Satellite and is then retransmitted to the White Sands Complex near Las Cruces, New Mexico. At this point, the data are interpreted and then retransmitted to the Data Operations Control Center at the Goddard Space Flight Center in Greenbelt, Maryland, routed to the Data Capture Facility, and then it is finally sent to the Space Telescope Science Institute where a team of astronomers and space scientists interpret the data. (2003, Pdf2)

In making sense of this system as a cognitive and computational system, Giere divides the system into three sorts of computational apparatuses. First he suggests that there is a set of input systems that have the function of taking analog information about the world and converting it to digital information that can be interpreted by later computational structures. The second system then takes the digital output of these systems and converts it into the sorts of images that can be interpreted by scientists. Finally the third system consists of the team of scientists that interpret the images and converts the data into a form that can be reported in scientific journals and in the popular press. The important thing to keep in mind about all of these systems, however, is that "Each of these components is itself a distributed cognitive system including the hardware, software, and the many people who operate it" (Giere 2003, 3pdf). Each of these systems is dedicated to the acquisition and interpretation of a particular sort of information. Moreover, this information must be interpreted sequentially. Later systems always take as input the information that has been processed by earlier systems. To put the point briefly, throughout the process that propagates information forward throughout the system, "the representation is transformed in many ways thought to make it most informative to the astronomers who will eventually judge its scientific significance" (Giere 2003, 3pdf).

Finally, and most importantly for the thesis of distributed cognition, Giere demonstrates that at the third stage of interpretation, there are a number of scientists who are looking at images, comparing them to previous images, and interacting with one another in order to interpret the data that is presented in the final image that results from the sequential processing addressed above. Giere then argues that because of the way in which the processing of information occurs in this system, cognitive scientists should be less concerned with what is going on in the heads of individual scientists and should instead focus on the way in which the *external* representations are "evaluated for their implications regarding 13 billion year old galaxies" (Giere 2003, 4pdf).

5.5.2.3 Are these cases of genuinely collective cognition? The first thing to note about each of these cases is that the specialization of function in both CERN and in the Hubble system facilitates the propagation of representational states across representational media. Each of the subsystems within CERN and the Hubble system is dedicated to processing a particular sort of information, and unless it does so, none of the other systems will be capable of doing their job. In the case of CERN there appears to be a more densely interconnected system that consists of a variety of mutually interdependent systems. In the case of the Hubble system it appears to be the case that the system is more clearly a feed forward system in which information is propagated from one system to another in a linear fashion.[116] However, regardless of this structure, it seems reasonable to say that there are a number of dependant systems that are functionally specialized for processing a particular sort of information.

[116] Given that the description offered by Giere relies much less heavily on a thorough ethnographic study of the system in question, it is hard to say exactly how the system is organized. However, because it does not matter for my purposes, I'll take the Hubble system to be a linear feed-forward system until we have further data on its actual organization.

Second, the cognitive states of both these collectivities look to be more than the mere aggregation of individual cognitive states. Because of the way in which the information from these various systems is coordinated and because of the way in which each of the systems is dependent on the local states of the systems to which it is connected, we will need a complete story about the state of the entire system and the way in which information is being processed by each of the systems in coordination in order to have a complete story about the cognitive state of the system. Moreover, because of the way in which each of the functionally specialized systems in both the case of CERN and the Hubble system operates as a functionally specialized system, the particular scientists that happen to be working on a particular issue are far less important than the functional roles that they happen to play. For this reason, numerous people can come and go from the collectivity throughout a variety of projects without this having significant implications for the functioning of the collectivity as a whole. For this reason, there will be a number of important facts about the cognitive state of the collectivities that will not be captured by appeal to the cognitive states of individuals in aggregation.

Clearly there are collective states here, and clearly they produce collective representations. However, there remain a number of questions as to whether theses systems are genuinely *psychological* systems. In establishing the case for the existence of genuinely cognitive states in the case of both CERN and the Hubble system, we must note that each of these systems does in fact possess a number of internal states that have the function of adjusting the system's behavior in ways that allow it to cope with features of its environment in ways that are not fully determined by the design of the system. Moreover, these states are capable of standing-in for various salient features of the world. A number of the components

of CERN operate over data that has been collected over a twenty-year period; In a number of cases, no one is actually looking at what is going on in a bubble chamber or looking at the readouts from a detector—they are instead operating on the physical representations of the outputs of detectors in an attempt to make sense of what happened in the world at another point. In the case of the Hubble system, people are looking at the information on computer screens and comparing them to other representations that have been produced in the past. Even in the absence of immediate environmental stimuli, each of these systems is dedicated to interpreting and running computations over a variety of representations.

The representations that are produced at any given time by any of the subsystems within either CERN or the Hubble systems are also best understood as part of larger representational schemes that allow the people within these groups to represent a variety of possible contents in a systematic way by manipulating the representations and producing other representations for consumption by other systems. Finally, there are indeed proper and improper ways of producing, maintaining, modifying, and using the various representations. This is clearest in the case of CERN. The use of regimes of trust and the use of gossip to ensure that each of the individuals in a particular group are producing representations in the way that they are supposed to is meant to ensure that the representations are produced in such a way that they adequately represent facts about the physical substrates of the world. More importantly, if the systems fail to operate as they are supposed to, then they will misrepresent the world, and they will produce publications that will be refuted, shown to be somehow mistaken, or mistakenly adopted by other collectivities. To put the point briefly, CERN is capable of misrepresenting the physical facts.

Finally, in order to prevent the production of misrepresentations, CERN runs

numerous experiments over numerous hypotheses in order to produce the most accurate representation of the world that it is capable of producing. This allows CERN to produce hypotheses that allow their hypotheses to die in their stead by deciding which papers will be published on the basis of internal models and a series of internal checks and monitors. CERN can even bee seen as a sort of Gregorian collectivity (Dennett 1996, 99-101) that engages in meta-representation to the extent that they can genuinely ask if they are correctly modeling the world in a way that produces the optimal publications.

I see no reason, then to deny the status of a cognitive system to at least some scientific labs. How far this can be extended is of course an empirical question, and as Knorr Cetina (1999) aptly demonstrates, there are some sorts of labs with structures in place that actively militate against the possibility of collective representations. However, it is only by taking the thesis of distributed cognition seriously that it is possible to engage in the ethnographic research that can underwrite any claim to collective mentality in scientific labs. More in depth studies such as the ones offered by Knorr Cetina would be quite useful in studying the cognitive science of science.

5.3 Intuitively plausible cases of distributed cognition:

I want to close with a couple of cases about which we do not yet have data, but that have gained a prominent place as intuitively plausible cases of distributed cognition. To begin with, I've considered the existence of collective memory from the standpoint of the literature on transactive memory; however, in a recent article on the theoretical underpinnings of distributed cognition, John Sutton (2006) argues that autobiographical memory should, at least in some cases, be recognized as a sort of distributed cognition.

Although Sutton concedes that there will be a number of cases in which we remember significant facts about our lives on our own, there are also many cases in which autobiographical memories can be reconstructed only by depending on the memories of others. In agreement with Wegner, Sutton (2006, 238) argues that the sharing of memory is "an ordinary human activity with great psychological and social significance" and it is often the case that sharing "memories brings into being new emergent form and content through the transactive nature of collaborative recall" (Sutton 2006, 238). The question, then, is: what distinguishes the cases of genuinely collective memory from merely aggregative memories.

To begin with, we should take a cue from Wegner. Although we might expect our own lives to be something about which we are more likely to be specialists than are our friends, families, and coworkers. The important thing to keep in mind, however, is that to a large degree we are indeed strangers to ourselves. In a significant number of cases, our actions are more significant to others than they are to us. I might not remember the biting criticism that I made of a friends thesis project, or the flirtatious comment I made to a member of the wait-staff at a restaurant that I frequent. However, these things might be significant to my friend or to my romantic partner. Given the intimate links between significance, attention, and the strength of a memory, it is safe to assume that there will be a number of cases in which facts about my autobiography will be more likely to be internally stored by the people around me than they will be to be stored by me. This provides for a sort of specialization that distributes the facts about my life across various individuals. The interesting thing about cases of collective reconstruction in autobiographical memory is that the distribution here is likely to have a very different structure from the one posited by

Wegner. We are, for example, in a number of cases not going to have meta-memories that assign another person as the location of my memories. Although there will no doubt be such cases (e.g., you might know that your mom remembers all of your important—and unimportant—achievements from kindergarten through graduating high school), the majority of cases in which autobiographical memories are distributed will lack any sort of formal structure. Collective autobiographical memories are often produced by a process in which one person's memory causes another person to remember something else, and this continues until the group produces a narrative, and at some point, all of the members agree that the narrative is probably what happened. This process of reconstruction is far less like recall than it is like telling a story, and just as we've seen in the case of CSI, the distribution and integration of information in the production of a narrative can often have an importantly collective structure.

One additional intuitively plausible case of distributed cognition is the process of coauthoring a paper.[117] I have recently done collaborative work with a pair of friends on the attribution of phenomenal states to groups. In writing up our results, the three of us had rather clear and importantly unique specializations as regards the sort of paper that we were writing. In writing the paper, we each wrote up the parts of the project about which we had the most

[117] This case is suggested by Pierre Poirier and Guillaume Chicoisne (2006, 229). Unfortunately, Poirier and Chicoisne are no friends to this case as a case of genuinely collective cognition. Although they agree that there are some cases in which collective mentality might emerge from the interactions of individuals, they suggest incredibly stringent conditions on anything counting as genuinely collective mentality. Poirier and Chicoisne argue that the only cases in which we are warranted in attributing mental states to a collectivity are those cases in which it is necessarily the case that if one person were removed and replaced with someone who is functionally equivalent for the purposes of the collectivity, the project could never be completed. The strange thing about this claim is that it would also seem to preclude the possibility of individual cognition. After all, if one neuron dies, or if a new neuron is born in the hippocampus, this change in the overall structure of the brain does not eliminate the possibility of individual cognition. More importantly, the human brain is incredibly plastic, and as we know bits of neural architecture can be recruited to perform different tasks when there is damage to the area that is typically implicated in a particular sort of processing. I will thus ignore this stringent condition on distributed cognition and stick to the functional specialization on which I have relied in the rest of this thesis.

expertise, we then read what each author had wrote, commented on their work, rewrote things that we were unhappy with an then met to discuss, debate and argue about the specific ways in which the project should be developed. The resulting paper was nothing that any of us could have produced individually, it was not a paper that any of us would be willing to endorse as individuals, but it was the result of a collective effort to produce a paper on collective mentality. This is not to say that every case of collaboration will be a case of collective cognition. However, in cases where each of the authors has specific expertise such that the content of the paper is the result of distributed intelligence across the authors, it is reasonable to count this as the result of collective cognition.

At this point, it is important to note that these cases currently have the function of being nothing more than intuition pumps. They are cases where it seems reasonable to attribute collective mental states; however, in order to conclusively demonstrate that there truly is a specialization of function within these collectivities that facilitates the propagation of representational states throughout the group, and in order to demonstrate that there really are cognitive states of the groups in question, it would be quite useful to collect ethnographic data on both autobiographical recall and on collaborative authorship. I cannot engage in the collection of this data now. So, I leave this open as a future research project to be developed by myself or in conjunction with others who are willing to collaborate with me on these cases of collective cognition.

5.5.4. A concluding note on distributed cognition: In 2005, Francis Heylighen and Frank Van Overwalle submitted a research proposal to the Free University of Brussels entitled "The self-organization of distributed cognition: a connectionist approach". Heylighen, Overwall, and a team of 15 other researchers proposed to study the possibility of

distributed cognition by way of computer simulation, experiments on the dynamics of within-group communication, and computer mediated games. Specifically, they propose to test the hypothesis that human groups are best understood as self-organizing systems in which individuals learn to cooperate with and trust one another in a way that facilitates the coordination and distribution of information and labor. They further propose that the mechanisms at play in the distribution of computational capacities within a collectivity is organized as a connectionist network in much the same way that computational capacities are organized in single human agents.

The Project knows as the Evolution, Complexity and Cognition group (ECCO) directed by Frances Heylighen will likely produce a number of intriguing results. However, at this point, this is merely a research program. I can only hope that the research in this lab will lend more credence to the defense of collective mentality.

Works Cited:

Adolphs, R (1999). Social cognition and the human brain, Trends in Cognitive Science, 3: 469-479.

Akins, K (1996). Of sensory systems and the 'aboutness' of mental states, The Journal of Philosophy, 93: 337-72.

Allport, F (1924). Social Psychology, Boston: Houghton Mifflin

Ames, RT (1994). Reflections on the Confucian self, in M Bockover (ed), Rules, Rituals, and Responsibility, LaSalle: Open Court.

Anscombe, GEM (1957/2000). Intention. Cambridge: Harvard University Press.

Arico, A, B Fialla, and S Nichols (in prep). The folk psychology of consciousness.

Armstrong, D (1980). The causal theory of mind, The Nature of Mind and Other Essays, Ithaca: Cornell University Press: 16-31.

Baber, C, et al (2006). Crime scene investigation as distributed cognition, Pragmatics and Cognition, 1: 357-385.

Barsalou, LW (1987). The instability of graded structure: Implications for the nature of concepts, in U Neisser (ed), Concepts and Conceptual Development (101–140). Cambridge: Cambridge University Press.

Bauer, R (1984). Autonomic recognition of names and faces in prosopagnosia: A neuropsychological application of the guilty knowledge test, Neuropsychologia, 22: 457-469.

Beekman, M, RL Fathke, and TD Seeley (2006). How does an informed minority of scouts guide a honeybee swarm as it flies to its new home? Animal behaviour, 71: 161-171.

Beer, R (2000). Dynamical approaches to cognitive science, Trends in Cognitive Science, 4: 91-99.

Beer, R and Chiel, H (1993). Simulations of cockroach locomotion and escape, in R Beer, R Ritzmann and T McKenna (eds), Biological Neural Networks in Invertebrate Neuroethology and Robotics (267-285). London: Academic Press.

Block, N (1978). Troubles with functionalism, in CW Savage (ed.), Minnesota Studies in the Philosophy of Science, 9 (261–325). Minneapolis: University of Minnesota Press.

_____(1980a). Are absent qualia impossible? Philsophical Review, 89: 257-274.

_____(1980b). What intuitions about homunculi don't show. Behaviroral and Brain Sciences, 3:425-426.

_____(1986). Advertisement for a semantics for psychology, in P French, et al (eds), Midwest Studies in Philosophy, 10 (615-678). Minneapolis: University of Minnesota Press.

_____ (1990). Inverted earth, Philosophical Perspectives 4: 53-79

_____ (1990b). Consciousness and accessibility. Behavioral and Brain Sciences, 13: 596-98.

_____ (1995). The mind as the software of the brain, in D Osherson, et al (eds), An Invitation to Cognitive Science (377-425). Cambridge: MIT Press.

_____ (2003). Mental paint, in M Hahn and B Ramberg (eds), Reflections and Replies: Essays on the Philosophy of Tyler Burge (125-51). Cambridge: MIT Press.

Bloom, P (2004). Descartes' Baby: How the science of child development explains what makes us human. New York: Basic Books.

Bloom, P and D Kelemen (1995). Syntactic cues in the acquisition of collective nouns. Cognition 56 (1):1-30.

Bloom, P and C Veres (1999). The perceived intentionality of groups. Cognition 71: b1-b9.

Bonabeau et al (1998). Fixed response thresholds and the regulation of division of labor in insect societies, Bulletin of Mathematical Biology, 60: 753–807.

Boyd, R and P Richerson (2005). The Origin and Evolution of Culture. Oxford: Oxford University Press.

Bratman, M (1987). Intentions, Plans, and Practical Reason. Cambridge: Harvard University Press.

_____ (1993). Shared intention, Ethics, 104: 97-113.

Brentano, F (1874/1995). Psychology from an empirical standpoint, AC Rancurello et al (trans). London: Routledge Press.

Brooks, DHM (1986). Group minds, Australasian Journal of Philosophy, 64: 456-470.

Burge, T (1979). Individualism and the mental. In P French, T Uehling, and H Wettstein (eds), Midwest Studies in Philosophy 4 (73-121). Minneapolis: University of Minnesota Press.

_____ (1986). Individualism and psychology, Philosophical Review, 95: 3-45.

Burros, M (1990). Fast food chains try to slim down, New York Times (April 11).

Cabeza, R, and L Nyberg (2000). Imaging cognition II: An empirical review of 275 PET and fMRI studies, Journal of Cognitive Neuroscience, 12: 1-47.

Chalmers, D (1996). The Conscious Mind. Oxford: Oxford University Press

Chisholm, RM (1957) Perceiving: A philosophical study. Ithaca: Cornell University Press

Chomsky, N (1959). A review of BF Skinner's Verbal behavior, Language, 35, (1): 26-58.

Churchland, PM (1979). Scientific Realism and the Plasticity of Mind. Cambridge: Cambridge University Press.

Clark, A (1989). Microcognition: Philosophy, cognitive science and parallel distributed processing. Cambridge: Bradford Books.

_____ (1990). Aspects and algorithms. Behavioral and Brain Sciences, 13: 601-602.

_____ (1997). Being There: Putting brain, body and world together again. Cambridge: MIT Press.

_____ (2002). That special something: Dennett on the making of minds and selves, in A Brook and D Ross, (eds.), Daniel Dennett (187-205). Cambridge University Press..

Clark, A and D Chalmers (1998). The extended mind. Analysis, 58: 10-23

Clarke, AC (1953/2001). Childhoods end. New York: Del Rey.

Couzin, I, et al (2002) Collective memory and spatial sorting in animal groups, Journal of Theoretical Biology, 218: 1-11.

_____ (1975). Functional analysis, Journal of Philosophy, 72: 741-765.

_____ (1985). The Nature of Psycholgoical Explanations. Cambridge: MIT Press.

_____ (1996). Representations, Targets, and Attitudes. Cambridge: MIT Press.

Darwin, C (1872/1965). The Expression of the Emotions in Man and Animals. Chicago: University of Chicago Press.

Davidson, D (1982). Rational animals, Dialectica, 36: 318-27

Dawkins, R (1989). The Selfish Gene. Oxford: Oxford University Press.

Decety, J et al (2002). Rapid communication: a PET exploration of the neural mechanisms involved in reciprocal imitation, Neuroimage, 15: 265-272.

Dennett, D (1978a). Intentional systems. Brainstorms (3-22). Cambridge: Bradford Books.

_____ (1978b). Why the law of effect won't go away. Brainstorms (71-89). Cambridge: Bradford Books.

_____ (1987a). True believers. The Intentional Stance (13-36). Cambridge: MIT Press.

_____ (1987b). Three kinds of intentional psychology (43-68). The intentional stance. Cambridge: MIT press.

_____ (1989). The origin of selves. Cogito, 3: 163-73.

_____ (1990). The myth of original intentionality, in M Said, et al (eds), Modeling the mind (43-62). Oxford: Oxford University Press.

_____ (1991). Consciousness Explained. London: Penguin Press.

_____ (1991b). Real patterns. Journal of Philosophy, 88: 27—51.

_____ (1992). The self as a center of narrative gravity, in F Kessel, et al. (eds), Self and Consciousness (275-78). Hilsdale, NJ: Lawrence Erlbaum.

_____ (1996). Kinds of Minds. New York: Basic Books.

Dennett, D and M Kinsbourne (1992). Time and the observer: the where and when of consciousness in the brain. Behavioral and Brain Sciences, 15: 183-247.

Deneubourg, J, et al (1989). The blind leading the blind: Modeling chemically mediated army ant raid patterns, Journal of Insect Behavior, 2: 719-725.

Dick, PK (1956/2002). The Minority Report. London: Gollancz Press.

Dretske, F (1988). Explaining Behavior. Cambridge: MIT Press.

Durkheim, E (1895/1982). The Rules of Sociological Method, Steven Lukes (ed) and WD Halls (trans). London: Free Press.

Edelman, S (1999) Representation and Recognition in Vision. Cambridge: MIT press.

Farrell, BA (1950). Experience. Mind, 49: 170-198.

Fattah, H (2006). Militia rebuked by some Arab countries. New York Times (July 17).

Fodor, J (1968). Psychological Explanation. New York: Random House.

_____ (1975). The Language of Thought. Cambridge: Harvard University Press.

_____ (1980). Methodological solipsism considered as a research strategy in cognitive psychology, Behavioral and Brain Sciences, 3: 63-73.

_____ (1987). Psychosemantics: The problem of meaning in the philosophy of mind. Cambridge: MIT Press

_____ (1990). A Theory of Content and Other Essays. Cambridge: MIT Press.

_____ (1991). A modal argument for narrow content, Journal of Philosophy, 87: 5-26.

Freud, S (1921/1975). Group Psychology and the Analysis of the Ego. New York: WW Norton & Co.

Gallese, V (2001). The 'shared manifold' hypothesis: from mirror neurons to empathy, Journal of Consciousness Studies, 8: 33-50.

_____ (2003a). The Manifold nature of interpersonal relations: the quest for a common mechanism, Philosophical Transactions of the Royal Society of London, 358: 517-528.

_____ (2003b). The roots of empathy: the shared manifold hypothesis and the neural basis of intersubjectivity, Psychopathology, 36: 171-180.

Geire, R (2002a). Distributed cognition in epistemic cultures, philosophy of science, 69: 637-644.

_____ (2002b). Models as parts of distributed cognitive systems, in L Magnani and N Nersessian (eds), Model based reasoning: Science, technology, values (227-41), Amsterdam: Kluwer.

_____ (2002c). Scientific cognition as distributed cognition, in P Carruthers, S Stitch and M Siegal (eds), Cognitive Bases of Science. Cambridge: Cambridge University Press.

_____ (2004). The problem of agency in scientific distributed cognitive systems, Journal of Cognition and Culture, 4 (3-4): 759-74.

Giere, R and B Moffatt (2003). Distributed cognition: Where the cognitive and the social merge, Social studies of science, 33: 301-310.

Gilbert, M (1987). Modeling Collective Belief. Synthese (73): 185-204

_____ (1989). On Social Facts. New York: Routledge.

Gladwell, M (2000). The Tipping Point: How little things can make a big difference. Boston: Back Bay Books.

Gould, SJ (1996). The Mismeasure of Man. New York: WW Norton & Co.

Graham, K (2002). Practical reasoning in a social world. Cambridge: Cambridge University Press.

Grice, HP (1957). Meaning, The philosophical review, 66: 377-88

Grodzins, M (1958). The metropolitan area as a racial problem. Pittsburgh: University of Pittsburgh Press.

Habermas, J (1984). The Theory of Communicative Action (2 vol), T McCarthy (trans). Boston: Beacon Press.

Hamilton, WD (1971). Geometry for the selfish herd, Journal of Theoretical Biology, 31: 295 -311.

Haugeland, J (1998). Having Thought: Essays in the metaphysics of mind. Cambridge: Harvard University Press.

Hayek, FA (1945). The use of knowledge in society, American Economic Review, 35: 519-530.

Heidegger, M (1927/1996). Being and Time, J Stambaugh (trans). Albany: State University of New York Press.

Heider, F and M Simmel (1944). An experimental study of apparent behavior, American Journal of Psychology, 57: 243–249.

Heinlein, R (1941/1958). Methuselah's Children. New York: Gnome Press.

_____ (1959). Starship Troopers. New York: Putnam Press.

Heintz, C (2006). Web search engines and distributed assessment Ssystems, *Pragmatics and cognition*, 14 (2): 387-409.

Heylighen, F, M Heath, and F Van Overwalle (2004). The emergence of distributed cognition: A conceptual framework, Proceedings of Collective Intentionality IV. Siena, Italy.

Hirshfeld, L (2001). On a folk theory of society: children, evolution, and mental representations of social groups, Personality and Social Psychology Review, 5 (2): 107–117.

Huebner, B, M Bruno, and H Sarkissian (under review). What does the Nation of China think about phenomenal states?

Hume, D (1739/2000). A treatise of human nature. Oxford: Oxford University Press.

Hutchins, E (1995). Cognition in the wild. Cambridge: MIT Press.

_____ (1995b) How a cockpit remembers its speeds, Cognitive Science, 19: 265-288.

Jackson, F (1998). From Metaphysics to Ethics: A defense of conceptual analysis. Oxford: Oxford University Press.

Jackson, F and D Chalmers (2001). Conceptual analysis and reductive explanation, Philosophical Review, 110 (3): 315-60

James, W (1890). The principles of psychology (2 vol), Cambridge: Harvard University Press.

Kashima Y, et al (1995). Culture, gender, and self: A perspective from individualism-collectivism research, Journal of Personality and Social Psychology, 69 (5): 925-937

Kashima, Y, et al (2005). Culture, essentialism, and agency: Are individuals concept of intentional action: A case study in the uses of folk psychology universally believed to be more real entities than groups? European Journal of Social Psychology, 35: 147–169.

Keil, F (1989). Concepts, Kinds, and Cognitive Development. Cambridge: MIT Press.

Kelso, JAS (1995). Dynamic Patterns: The self-organization of brain and behavior. Cambridge: MIT Press.

Kirsch, D (2006). Distributed cognition: A methodological note, Pragmatics and Communication, 14: 249-262.

Kniffin, KM and DS Wilson (2005). Utilities of gossip across organizational levels: multilevel selection, free-riders, and teams, Human Nature, 16: 278-292.

Knobe, J (2006). The Concept of Intentional Action: A Case Study in the Uses of Folk Psychology, Philosophical Studies, 130: 203-231.

Knobe J and J Prinz (2008). Intuitions about consciousness: Experimental studies, Phenomenology and Cognitive Science, 7 (1): 67-83.

Knorr Cetina, Karin (1999), Epistemic Cultures: How the sciences make knowledge. Cambridge: Harvard University Press.

Kreigel, U (2003). Is intentionality dependent upon consciousness? Philosophical Studies, 116: 271-307.

Kühnholz, S and TD Seeley (1998). The control of water collection in honey bee colonies, Behavioral ecology and sociobiology, 41:407-422.

Latour, B (1999). Pandora's Hope: Essays on the reality of science studies. Cambridge: Harvard University Press.

Latour, B and S Woolgar (1979). Laboratory Life: The social construction of scientific facts. Los Angeles: Sage Publishing.

Le Bon, G (1895/2002). The Crowd: a study of the popular mind. New York: Dover publications.

Lewis, D (1972). Psychophysical and theoretical identifications. Australasian Journal of Philosophy, 50:249-58.

Liang, D, R Moreland, and L Argote (1995). Group versus individual training and group performance: The mediating role of transactive memory, Personality and Social Psychology Bulletin, 21 (4): 384-393

Lohr, S and S Hansell (2006). Microsoft and Google set to wage arms race. New York Times (May 2).

Lycan, W (1981). Form, function, and feel, Journal of Philosophy, 78: 24-50.

_____ (1987). Consciousness. Cambridge: Bradford Books.

_____ (1996). Consciousness and Experience. Cambridge: MIT Press.

Mandeville, B (1728/1962). The fable of the bees. New York: Capricorn Books.

Marino, L, D McShea and M Uhen (2004). The origin and evolution of large brains in toothed whales, The Anatomical Record, 281A: 1247–1255.

Marr, D (1982). Vision: A computational investigation into the human representation and processing of visual information. San Francisco: WH Freeman and co.

McCullers, C (1943). The ballad of the sad café. Harper's Bazaar, 77: 140-61.

McDougall, W (1920). The Group Mind. Cambridge: Cambridge University press.

McPhaill, C (1991). The myth of the maddening crowd. New York: Aldine De Gruyter.

Menon, T et al. (1999). Culture and the construal of agency: Attribution to individual versus group dispositions, Journal of Personality and Social Psychology, 76: 701-717

Mill, JS (1843/1988). The Logic of the Moral Sciences. Chicago: Open Court Publishing.

_____ (1866/1961). Comte and Positivism. Ann Arbor: University of Michigan Press

Millikan, R (1984). Language, thought and other biological categories. Cambridge: MIT press.

_____ (1989). Biosemantics, Journal of Philosophy, 86 (6): 281-297

_____ (1996). Pushmi-pullyu representations, in L May, M Friedman, and A Clark (eds), Mind and Morals (145-161). Cambridge: MIT Press.

Minsky, M. (1988). Society of the mind. New York: Simon and Schuster.

Moreland RL, and L Myaskovsky (2000). Exploring the performance benefits of group training: Transactive memory or improved communication? Organizational Behavior and Human Decision Processes, 82 (1): 117-133.

Morris, M, T Menon, and D Ames (2001). Culturally conferred conceptions of agency: A key to social perception of persons, groups, and other actors, Personality and Social Psychology Review, 5 (2): 169–182

Mouawad, J and S Erlanger (2006). Israel and Hezbollah trade barrages. New York Times (July 17).

Nagel, T (1974). What is it like to be a bat? Philosophical Review, 83(4): 435-50.

Neitzsche, F (1887/1998). On the Genealogy of Morality, M Clark and A Swensen (trans). Indianapolis: Hackett Publishing.

Nisbett, R et al (2001). Culture and systems of thought: Holistic vs. analytic cognition, Psychological Review, 108: 291-310.

Nicolis, G and I Prigogine (1977). Self-Organization in Non-Equilibrium Systems. Hoboken, NJ: John Wiley and Sons.

Passino, K and TD Seeley (2006). Modeling and analysis of nest-site selection by honey bee swarms: the speed and accuracy trade-off, Behavioral Ecology and Sociobiology, 59:427-442.

Pelphrey, K. et al. (2003a) Brain activity evoked by the perception of human walking: controlling for meaningful coherent motion, The Journal of Neuroscience, 17: 6819-6825

Poirier, P and G Chicoisne (2006). A framework for thinking about distributed cognition, Pragmatics and Cognition, 14 (2): 215-234.

Preston, S and F de Waal (2002). Empathy: its ultimate and proximate bases, Behavioral and Brain Sciences, 25: 1-72.

Prinz, J (2002). Furnishing the Mind. Cambridge: MIT Press

_____(2005). The return of concept empiricism, in H Cohen and C. Leferbvre (eds) Categorization and Cognitive Science (679-94). Amsterdam: Elsevier.

Purves and Lotto (2003). Why We See What We Do: An empirical theory of vision. Sunderland, MA: Sinauer Associates.

Ramachandran VS (1993). Behavioral and magnetoencephalographic correlates of plasticity in the adult human brain, Proceedings of the National Academy of Sciences, 90: 10413-10420.

_____ (1998a). Phantoms in the Brain. New York: Harper Collins Publishers.

_____ (1998b). Consciousness and Body Image: Lessons from Phantom Limbs, Capgras Syndrome, and Pain Asymbolia, Proceedings of the Royal Society of London B, 353, 1851-1859.

_____ (2004). A Brief Tour of Human Consciousness. New York: Pi Press.

Reichenbach, H (1938). Experience and Prediction. Chicago: The University of Chicago Press.

Resnick, M (1997). Turtles, Termites and Traffic Jams. Cambridge: MIT Press.

Reuters (2006). North Korea is defiant over U.N. Council nuclear resolution. New York Times (July 17).

Reynolds, C (1987). Flocks, herds, and schools: A distributed behavioral model, Computer Graphics, 21(4): 25-34.

Rizzolatti, G et al (1996). Premotor cortex and the recognition of motor actions, Cognitive Brain Research, 3: 131-141.

Rosenberg, A (1988). Philosophy of Social Science. Oxford: Clarendon Press.

Rosenberg, J (1986) The Thinking Self. Philadelphia: Temple University Press.

Rosenthal, D (1986). Two concepts of consciousness, Philosophical Studies, 49: 329-359.

Rozin, P and C Nemeroff (1990). The laws of sympathetic magic, in J Stigler, et al, Cultural Psychology (205-232). Cambridge: Cambridge University Press.

Rumelhart, D, J McClelland, and the PDP research group (1987). Parallel Distributed Processing (2 vol). Cambridge: MIT Press.

Rupert, R (1998). On the relationship between naturalistic semantics and individuation: Criteria for terms in a language of thought, Synthese, 117: 95-131.

_____ (1999). The Best Test Theory of Extension: First Principle(s), Mind and Language, 14: 321-355.

_____ (2001). Coining terms in the language of thought: Innateness, emergence, and the lot of Cummins's argument against the causal theory of mental content, The Journal of Philosophy, 98: 499-530.

_____ (2005). Minding one's own cognitive system: When is a group of minds a single cognitive unit, Episteme: A Journal of Social Epistemology, 1 (3): 177-88.

Schelling, T. (1971). Dynamic models of segregation, Journal of Mathematical Sociology, 1:143-186.

Searle, J. (1969). Speach Acts. Cambridge University Press.

_____ (1980). Minds, brains and programs. Behavioral and Brain Sciences, 3: 417-457.

_____ (1990a). Collective intentions and actions, in P Cohen, J Morgan, and ME Pollack (eds), Intentions in Communication. Cambridge: Bradford Books.

_____ (1990b). Consciousness, explanatory inversion, and cognitive science. Behavioral and Brain Sciences, 13: 585-596.

_____ (1992). The Rediscovery of the Mind. Cambridge: MIT press.

_____ (1995). The Construction of Social Reality. New York: Free Press.

Seeley, TD (1983). Division of labor between scouts and recruits in honeybee foraging, Behavioral Ecology and Sociobiology, 12: 253-259.

_____ (1986). Social foraging by honeybees: how colonies allocate foragers among patches of flowers, Behavioral Ecology and Sociobiology, 19: 343-354.

_____ (1992). The tremble dance of the honey bee: message and meanings, Behavioral Ecology and Sociobiology, 31: 375-383.

_____ (1995). The Wisdom of the Hive. Cambridge: Harvard University Press.

_____ (1997). Honey bee colonies are group-level adaptive units, The American Naturalist, 150 (supp): 22-41.

_____ (2003). Consensus building during nest-site selection in honey bee swarms: the expiration of dissent, Behavioral Ecology and Sociobiology. 53: 417-424.

Seeley, TD and SC Buhrman (2001). Nest-site selection in honey bees: how well do swarms implement the "best-of-N" decision rule? Behavioral Ecology and Sociobiology, 49: 416-427.

Seeley, TD, S Camazine, and J Sneyd (1991). Collective decision-making in honey bees: how colonies choose among nectar sources, Behavioral Ecology and Sociobiology, 28: 277-290.

Seeley, TD and WF Towne (1992). Tactics of dance choice in honey bees: do foragers compare dances? Behavioral Ecology and Sociobiology, 30: 59-69.

Seeley, TD and PK Visscher (2003). Choosing a home: how the scouts in a honey bee swarm perceive the completion of their group decision making. Behavioral Ecology and Sociobiology, 54: 511-520.

Selden, G (1912/1995). Psychology of the Stock Market. Wells, VT: Fraser publishing company.

Sellars, W (1963/1991). Science, Perception and Reality. Atascadero, Ca: Ridgeview Publishing Co.

Smith, A (1776/1976). An Inquiry into the Nature and Causes of the Wealth of Nations. Oxford: Oxford University Press.

Sober, E and DS Wilson (1998). Unto Others: The evolution and psychology of unselfish behavior. Cambridge: Harvard University Press.

Stich, S (1979). Do animals have beliefs? Australasian Journal of Philosophy, 57: 15-28

Sugden, R (1993). Thinking as a team: toward an explanation of nonselfish behavior, Social Philosophy and Policy, 10: 69-89.

Sumpter, D (2006). The principles of collective animal behaviour, Philosophical Transactions of the Royal Society of London: Series B, 361: 5-22.

Sutton, J (2006). Distributed cogntion: Domains and dimensions, Pragmatics and Cognition, 14 (2): 235-47.

Thom, C, TD Seeley, and J Tautz (2000). Dynamics of labor devoted to nectar foraging in a honey bee colony: number of foragers versus individual foraging activity, Apidologie, 31: 737-738.

Titchener, E (1901-1905). Experimental Psychology: A manual of laboratory practice. New York: McMillan.

_____ (1912). The schema of introspection. American Journal of Psychology, 23: 485-508.

Tuomela, R (1992). Group beliefs, Synthese 91: 285-318.

Turing, A (1950). Computing machinery and intelligence. Mind, 59: 433-60.

Tye, M (1997). Ten Problems of Consciousness. Cambridge: MIT Press.

Ullman, S and E Sali (2000). Object classification using a fragment-based representation, Biologically Motivated Computer Vision: 73-87

Velleman, D (1997). How to share an intention, Philosophy and Phenomenological Research 57: 29-50.

Watkins, JWN (1952). The principle of methodological individualism, The British Journal for the Philosophy of Science, 3: 186-189.

Weber, M (1914/1968). Economy and Society (3 vol), G Roth and C Wittich (trans). Somerville, NJ: Bedminster Press.

Wilson, DS (1975). A theory of group selection, Proceedings of the National Academy of Science, 72: 143-146.

_____ (1997a). Altruism and organism: Disentangling the themes of multilevel selection theory, American Naturalist, 150 (supp.): S122-S134.

_____ (1997b). Incorporating group selection into the adaptationist program: A case study involving human decision making, in J Simpson and D Kendrick (eds), Evolutionary social psychology (345-86). Hillsdale, NJ: Erlbaum.

_____ (2002). Darwin's Cathedral: Evolution, religion and the nature of society. Chicago: University of Chicago Press.

Wilson, DS and LA Dugatkin (1997). Group selection and assortative interactions, American Naturalist 149: 336-351

Wilson, DS and E Sober (1989). Reviving the superorganism, Journal of Theoretical Biology, 136: 337-356.

Wilson, DS, et al (2000). Gossip and other aspects of language as group-level adaptations, Cognition and Evolution (347-365). Cambridge: MIT Press.

Wilson, DS, et al (2003). Cognitive cooperation: When the going gets tough, think as a group, Human Nature, 15: 225-250.

Wilson, R (1995). Cartesian Psychology and Physical Minds: Individualism and the sciences of the mind. Cambridge: Cambridge University Press.

_____ (2001) Group-level cognition, Philosophy of Science, 68 (supp): S262-S273.

_____ (2004). Boundaries of the Mind: The individual in the fragile sciences: Cognition. Cambridge: Cambridge University Press.

_____ (2005). Collective memory, group minds, and the extended mind thesis, Cognitive Processing, 6: 227-236.

Wilson, R and A Clark (forthcoming). How to situation cognition: letting nature take its course, in M Aydede and P Robbins, (eds), The Cambridge Handbook of Situated Cognition.

Wittgenstein, L (1953/2001). Philosophical Investigations. Oxford: Blackwell Publishing.

Wegner, D (1986). Transactive memory: A contemporary analysis of the group mind, in B. Mullen and G Goethals (eds), Theories of Group Behavior (pp. 185-208). New York: Springer-Verlag.

_____ (1995). A computer network model of human transactive memory, Social Cognition, 13: 1-21.

Wegner, D, T Giuliano, and P Hertel (1985). Cognitive interdependence in close relationships, in W Ickes (Ed), Compatible and Incompatible relationships (pp. 253-276). New York: Springer-Verlag.

Wegner, T and Wegner, D (1995). Transactive memory, in A Manstead and M Hewstone (eds), The Blackwell Encyclopedia of Social Psychology (pp. 654-656). Oxford: Blackwell.

Wundt, W (1911/1912). An Introduction to Psychology. R Pintner (trans). London: George Allen & Unwin.

Young, A, et al (1993). Face processing impairments and the Capgras delusion, British Journal of Psychiatry, 162, 695-698.