Do You See What We See?
An Investigation of an Argument Against Collective Representation

Bryce Huebner
Department of Philosophy
University of North Carolina Chapel Hill

We often speak as though collectivities can be in mental states. Israel *fears* further attacks from Hezbollah; Microsoft *intends* to unveil a new product in the spring; and, Starbucks *believes* that it needs to present a friendlier and less overtly monopolistic image. However, two major difficulties stand in the way of understanding such ascriptions of mentality literally. First, there is a prevalent commonsense intuition that such attributions of mentality should be understood as figurative or metaphorical. Second, although the social psychology of the "group mind" was a prominent research project in the social psychology of the late 19th century (cf., Le Bon 1895/2002, Freud 1914/1975, and Durkheim 1895/1982 and 1895/1997), there is a broad consensus among philosophers and cognitive scientists that these accounts of collective mentality failed. The positivism and behaviorism of the early 20th century trimmed away the ontological excess of group minds, and the cognitive revolution suggests little reason to extend the mind beyond the individual. Yet, despite a prevalent hostility to collective mentality, a growing number of philosophers (e.g., Gilbert 1989 and Petit 2003) and cognitive scientists (e.g., Wegner 1986, Hutchins 1995, and Knorr Cetina 1999) have began to develop arguments for the existence of collective mentality. Viable theories of collective mentality are still in their infancy, and as such they continue to face serious objections. In this paper I address the most troubling of these objections: the argument that any appeal to collective representation is explanatorily superfluous.[1]

I argue that although many appeals to collective representation are explanatorily superfluous, deploying the claim of explanatory superfluity to readily risks either proving too much, giving us reason to adopt eliminativism about individual mentality and abandon the idea of person-level representations, or risks proving too little, demonstrating precisely the continuity between individual and collective representations that would warrant attributions of collective mentality.

2. Explanatory Superfluity

Contemporary philosophers of mind and cognitive scientists agree on few claims about the nature of mind. However, one point of broad consensus is that a number of mental states and processes (e.g., beliefs, hopes, wishes, memories, and dreams) are best understood as representational.[2] Most cognitive scientists also agree that appealing to a system of internal representations capable of guiding a systems behavior by standing-in for various features of that systems environment is necessary in order to give an adequate explanation of the complexity and flexibility of a cognitive system's behavior. Moreover, a long tradition in the philosophy of mind and psychology has held that the capacity for representation is a necessary condition on mentality. It is commonly assumed that the capacity for mental representation requires, first, the presence of some states or processes that are intentional, and second, that these states or processes provide a system with information about its environment *because* they have the function of conveying information about some feature of the world to that system. Building on this understanding of the representational nature of mental states and processes, there are two ways in which collectivities could fail to posses the capacity for mental representation, and thus two ways in which a plausible argument for the explanatorily superfluity of collective representations could be developed. First, collectivities could fail to possess any representational states or processes whatsoever; second, there could be collective representations, but these could fail to play the relevant functional role within the collectivity to count as genuinely mental representations.

2.1 As-If Collective Representation

The first version of the argument for the explanatory superfluity of collective representation could be developed by demonstrating that although we often speak as-if collective behavior depends on collective representation, the states and processes that actually explain collective behavior are not, in fact, representational states *of the collectivity*. There are many cases where we speak figuratively in attributing mental states. For example, my father once told me that you should freeze seeds the night before you plant them because it will make them *want* to grow—he claimed that they would *think that* winter was ending. In this case, my father was speaking as-if the seeds would act on the basis of its representational states, even though there is no reason to think that this appeal to representational states is warranted. In order to develop this sort of argument, the proponent of the argument from superfluity will attempt to demonstrate that appeals to collective mental states are analogous to such appeals in the case of seeds. If our use of collective mental language rests merely on as-if representational states or processes, then appeal to such mental states, although it may initially seem explanatorily fruitful, will prove superfluous as we develop more empirically adequate explanations of the mechanisms producing collective behavior.

There are numerous collective phenomena that are usefully predicted and explained *as collective* that do not require an appeal to collective representation. The most interesting class of systems such are those collectivities in which the dynamic updating of individual representational states in response to changes in local environment produces self-organization that yields emergent behavior that cannot be exhibited by any of the individuals composing that collectivity. These emergent phenomena suggest that some collectivities have states that are interesting in their own right even in the absence of a central coordinating system; however, they also lend credence to the claim that the positing of collective representations is explanatorily superfluous. Consider the canonical case of such self-organizing behavior—termite nest construction.

Within a termite colony, no individual builds an arch by itself, nor does any individual plan the construction of an arch. Moreover, no termite in the colony coordinates the behavior of the colony in order to construct the intricate system of

chambers, pillars and arches that we find in termite nests. Nonetheless, intricately structured nests emerge. Upon first observing the complex structure of a termite nest, the construction seems as though it must rest on a purposeful and integrated set of intentions; however, there is good reason to think that this collective behavior is better explained by appeal to the simple behaviors and representations of individual termites in aggregation. The explanation runs roughly as follows.[3] Individual termites, specialized in nest construction, make mud balls and drop them randomly. However, when a termite drops its mud ball it also deposits a pheromone trace, thereby modifying the local environment in a way that is conducive to constructing pillars and arches. Termites are inclined to drop their mud balls wherever the pheromone trace is strongest. So, as termites drop their mud balls near those of other termites, the pheromone traces continue to become stronger. This leads termites to drop their mud balls on top of one another, leading to the construction of columns. When two columns are close enough to one another, and when the air currents are right, pheromones drift in such a way that mud balls will be deposited preferentially on the side of a column nearest to an adjacent column. This process continues until the two columns meet—eventually producing an arch. Through this, and other relatively simple aggregative mechanisms, intricate nests are produced by termite colonies.

Someone may initially suppose that the presence of collective behavior in nest construction requires that "swarm intelligence" be produced through the self-organizing behavior of termites. However, although the behavior of termite colonies appears to be unified and purposeful, there is no collective representation motivating nest construction. The colony does not represent the eventual structure of the nest, and neither does any individual within the colony. The emergent phenomena in this case is best explained by the fact that individual termites are capable of producing mud balls and depositing pheromones, representing the location and strength of pheromone traces, and depositing their mud balls near the strongest pheromone trace. Because we can adequately explain the structure of the nest by appeal to the behaviors and intentional states of individual termites, and because nothing at the level of the collectivity needs to be represented in order for the structure of the nest to emerge as it does, appealing to a system of collective representations produced in order to build an intricate network of tunnels and arches is explanatorily superfluous. Nest building behavior in termites emerges because of the way in which the individual termites respond to local changes in their environment and we don't need to appeal to any sort of collective representation to explain this.

Similar considerations are likely to explain the phenomenon of racial segregation in large cities. Using a simple mathematical model, and a handful of dimes and pennies, Thomas Schelling (1971) demonstrates that a relatively small in-group bias can explain the emergence of racial segregation in a community, even where this move towards segregation departs radically from the sort of community that would be desired by each of the individuals that compose a community. Moreover, Schelling demonstrates that once movement towards segregation begins in a particular area, it will be self-sustaining because of the way in which in-group biases lead individuals to resist living in a neighborhood where members of an out-group outnumber them. Although segregation phenomena are interesting and important *as emergent phenomena*, Schelling's model suggests that no explanatory gains will be made in positing collective mental representations at the level of a community. It seems that there is no residual phenomena to explain by appealing to anything beyond the representational states (e.g., the representation of a person as a member of an in-group) of the individuals that compose a community and a demonstration of the ways in which these representational states produce behaviors collective behaviors because of the changes in the individual's local environment.

In both these cases, and in numerous others, there is no reason to the appeal to representational states at the level of the collectivity. Once the socially embedded behaviors of individuals are explained, and once the rules for aggregation of individual states are explained, there are no further phenomena that call for explanation. We are not left with a question about why the termite colony decided to build the nest in one way rather than another, or why it believed that there should be a pillar here and a mound there—the colony makes no such decisions and it has no such beliefs. Similarly, there is no question about why a neighborhood "decided" to be primarily Armenian or Irish beyond the decisions of the individuals. Claiming that a neighborhood intended to be segregated offers no explanatory purchase on institutionalized racism; however, a better understanding of individual representations, coupled with a better understanding of local environments and an account of the way that individual states are updated in response to changes in their local environment *might*.

By looking to the way in which collective behavior can be explained without an appeal to collective representation, we begin to see why someone might think that positing collective representation would be explanatorily superfluous. However, the fact that *these emergent phenomena* do not require an appeal to representational states or processes should not, by itself, suggest that no collectivities possess representational capacities. After all, a number of philosophers have recently developed theories of collective intentionality that purport to demonstrate the existence of collectivities that are structured in such a way that they do produce collective representations. The question, then, is whether the sorts of collectivities discussed within the literature on collective intentionality are operating on the basis of collective mental representations.

2.2. Collective Intentionality and Derived Representations

A second version of the argument for the explanatory superfluity of collective representation can be developed by recognizing that although some collectivities represent some features of their environment, the representational states and processes operative in these collectivities are not really mental representations. Within the philosophy of mind, a distinction is often drawn between those representations that have semantic content because of the causal relations obtaining between the representation and a feature of the environment, and those representations that derive their semantic content from the contentful states of another system.[4] For example, a person can have beliefs about the three 'Xs' and the

'skull-and-crossbones' printed on the label of the moonshine in her cupboard. However, the three 'Xs' and the 'skull-and-crossbones' represent the potency of the moonshine only because of the way in which these symbols are produced and interpreted. Using this distinction, an argument from explanatory superfluity can be developed by demonstrating that collective representations only have derived semantic content.

If all collective representations have derived semantic content, then although appeals to the representational states of a collectivity will have some explanatory value, they will have this explanatory value only when they are recognized as deriving their content from the intentional states of the individuals composing a collectivity and the rules for their expression in a public language. In this case, appealing to collective mental representations will be explanatorily superfluous because a collective behavior will be completely explicable in terms of the structure of the collectivity, the representational states of the individuals that compose the collectivity, and the rules for the public expression of these aggregated individual mental representations. Hence, if collectivities only possess derived representations, then positing collective *mental* representations will be explanatorily superfluous. Consider what it would take to demonstrate the existence of collective mental representations on the basis of these public language structures.

On 6 September 2007, the United States Department of Transportation (USDOT) fulfilled a requirement of the North American Free Trade Agreement and granted Mexican trucking companies the permission to haul cargo anywhere in the United States. The International Brotherhood of Teamsters (IBT), however, *believes* that there are serious problems with allowing these trucking companies into the United States without ensuring that they meet the same safety standards as American trucking companies (e.g., safety inspections on the vehicles and mandatory drug testing on the drivers). Unlike the cases of emergent behavior surveyed in the previous section, in this case we also find genuine representations, public language representations in the form of press releases issued by the IBT following USDOT decision.

Within the literature on collective intentionality, public language structures such as press releases and court decisions are often suggested as plausible candidates for collective representation. Margaret Gilbert has been one of the strongest proponents of the claim that we ought to understand these representations as providing evidence for the existence of collective mentality. Gilbert (2004) claims that whenever the members of a collectivity form a "plural subject" by jointly committing to engage in come action, their collective actions will best be understood as arising from a collective intention. However, the members of a collectivity are to be understood as jointly committed to engaging in an action as a group *just in case* they have all agreed to act as a single body in order to bring about this action in so far as possible. So, according to Gilbert, it is only when each of the individuals adopts an intention to act as part of a collectivity that this sort of collective action can emerge. Moreover, whether a collectivity will be successful in any activity in which they are engaged always depends, according to Gilbert, on the activity of the individual members of the collectivity, each of whom is acting in such a way that is directed towards bringing about the collective goal.

At this point, it seems unclear how such an account of the formation of a collectivity could ever warrant an attribution of a mental state to a collectivity. However, there are a number of cases in which collective intentions are distinct from the intentions of the individuals that compose a collectivity, and it is in these cases that Gilbert finds her strongest cases for collective mentality. Consider the IBT's belief that opening borders is wrong. In this case, we find a number of individuals, each of whom is a member of the union. These individuals, at least insofar as they are members of the IBT, have reached an agreement about the status of the opening of borders. This does not, of course, mean that each of the individuals agrees with the claim that opening the borders is wrong. Rather, it means that each of the individuals is willing to assent to the claim that opening borders is wrong, at least in her capacity as union member. On the basis of this decision, each member of the IBT has committed to act *in the contexts that are relevant to their membership in the IBT* in accordance with this belief. Many members of the IBT may have no beliefs at all about the status of opening borders. However, because they are members of the IBT, they agree to act in conformity with the beliefs espoused by the IBT, *at least in the relevant circumstances*.

A model underlying this sort of reasoning about the possibility of collective mentality has also been developed on the basis of a formal model of collective decision making by Philip Pettit. Pettit (2003, 167) opens a recent paper with the claim that the organization of some collectivities makes them "subjects in their own right, giving them a way of being minded that is starkly discontinuous with the mentality of their members." His argument is constructed on the basis of a phenomenon he refers to as the discursive dilemma. The discursive dilemma takes the following form. Suppose a court, consisting of three judges, is to decide whether a defendant is liable for a breach of contact. Also, suppose that the judges will decide against the defendant if and only if there was a valid contract and the defendant's behavior violated this contract. Now suppose the decisions of the members of the court come out as follows (cf., Pettit 2003, 169):

	Contract?	Breach?	Liable?
Judge 1	Yes	No	No
Judge 2	No	Yes	No
Judge 3	Yes	Yes	Yes

The dilemma arises because *if the decision of the court is a simple majority vote on the question of liability* (the conclusion procedure), the decision about liability will be "No", but *if the decision of the court is based on the decision about each matter* (the premise procedure), the decision about liability will be "Yes".

Pettit contends that this sort of conflict between the premise procedure and the conclusion procedure is likely to emerge in many sorts of collectivities. Petit then argues that many collectivities make a collective decision to adopt the conclusion procedure. This often produces decisions that the majority of members of a group

would not individually accept, but that the group, as a whole, does accept. Because these decisions are not the decisions of the individual members of a collectivity, Pettit claims that we are forced to understand them as the mental states of the collectivity itself.

Unfortunately, a rather serious argument from explanatory superfluity can easily be developed against the theories of collective intentionality advanced by Gilbert and Pettit. Generalizing from the arguments offered by Gilbert and Pettit, Robert Rupert (2005) argues that any defense of collective intentionality that appeals to public language structures such as press releases and court opinions as the representational states of collectivities will face an objection from explanatory superfluity. Here is the argument in brief. For any collectivity that operates on the basis of a decision procedure—which includes all of the cases that have typically been taken to lend credence to the possibility of collective intentions—taking the representations produced by these collectivities to be the *mental* representations of the collectivity is unlikely to provide any explanatory or predictive advantage beyond what could be gained by a more thorough understanding of the psychological states of the individuals that compose these collectivities and a more thorough understanding of the decision procedures that are used by these collectivities in aggregating individual decisions. In fact, Pettit has given us just such an account of how this is possible. Generalizing from these sorts of cases, Rupert (2005: 5) argues that wherever we posit collective mental representations, we will find that:

> 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.

Understood in this way, Rupert's version of the argument from explanatory superfluity has two conditions, either of which is sufficient to rule out collective representation. First, Rupert claims that positing collective mental representations is explanatorily superfluous wherever the production and use of a collective representation proceeds by brute physical causation. Second he claims that attributing mental representations to a collectivity is explanatorily superfluous where the production and use of a collective representation proceeds by causal processes involving the mental states of individuals. These two conditions, however, beg for elaboration.

First, consider the effect of the qualifier "brute" in the first condition. The first major difficulty faced in defending this condition on explanatory superfluity is this. Any viable naturalistic theory of mental representation will require that every process involved in the production of individual mental representation will proceed by some causal or otherwise physical processes. In fact, every naturalistically plausible theory of mental representation holds that the production of *every mental representation* will necessarily proceed by some causal or otherwise physical process. So, unless there are unique difficulties faced by positing collective representations, this condition has the untoward consequence that any denial of collective representation on the basis of explanatory superfluity will double as an argument for denying the possibility of a naturalistically plausible theory of individual representation. But unless the denial of individual representation is on the table as a viable option (which I'll assume for the purposes of this paper it is not), something has gone wrong in offering this condition on explanatory superfluity.

However, there may be reason to suppose that there *are* unique difficulties raised by the realization of collective representations on causal processes involving the mental states of individuals that are importantly distinct from worries about the realization of individual representations. This, then, makes the second condition do all of the work in denying the explanatory value of collective representation. What seems to be important in claiming that collective representations are explanatorily superfluous is the dependence on individual mental states. There are, however, numerous ways in which the production and use of a collective representation can proceed "by causal processes involving the mental states of individuals". Understood weakly enough, this condition would beg the question against the possibility of collective representation—ruling out *a priori* any case in which a representation is produced by a system composed of subsystems that are themselves capable of producing representations. However, Rupert cannot understand this condition in this way. After all, as Rupert is quick to recognize, the possibility of collective representation is an empirical question that cannot be settled merely by an appeal to an intuition that minds are not themselves composed out of minds.

Fortunately, there is a more promising way of understanding this condition. Rupert (2005) claims that the reduction of individual representations to neurological states and collective representations to psychological states of the individuals that compose a collectivity differ greatly 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. Unfortunately, however, this suggests that the current status of our scientific knowledge is all that prevents us from eliminating individual representations.

In analyzing collective representations, we currently know how to look for the individual representations involved in the production and interpretation of collective representations. However, even our best neuroscience isn't nearly well enough developed to identify the physical states on which any individual representation is realized. Yet, it's not clear that this is a difference that makes a difference. As our understanding of neurophysiology and its relation to other sorts of physical explanation becomes more refined, there's reason to suppose that a coherent story about the realization of individual representations on physical states could become apparent, at least on the supposition that some naturalistically plausible theory of mental representation turns out to be true. 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.

It thus seems that there is a relatively straightforward way to dispense with this initial objection to collective representation on the basis of explanatory superfluity. However, there is a deeper worry about collective representation. While individual representations are realized on *physical processes*, collective representations are realized on individual *representational states*. The problem here is that nothing new *in kind* is introduced in moving from individual representations to collective representations—the relevant states and processes 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 are realized do not. The proponent of the argument from explanatory superfluity is, thus, likely to argue that the crossing of explanatory levels is significant in the case of individual mental representations *precisely because* intentional states have semantic content. Individual representations cannot be made superfluous by scientific discovery—we need them to explain the semantically significant states and processes that we find in 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 explanation. And it is this possibility that underwrites the most promising version of the argument for the superfluity of collective mental representation.

Briefly put, a more promising way of understanding this difference in "inter-level relations" is as a claim about the way in which the semantic properties of public language structures, such as court opinions and press releases, diverge from the psychosemantic properties of mental representations. Naturalistically plausible theories of mental representation rely on nomic relation between perceptual (or quasi-perceptual) processes and properties of things in the world that are unlikely to be present in the production of public language structures. These causal relations are supposed to explain how neurological states are capable of indicating or carrying information about properties of things in the world. However, court opinions and press releases, do not seem to indicate or carry information about anything except insofar as they are understood the output of some aggregative algorithm operating over the mental states of the individuals that compose these collectivities. Court opinions and press releases *derive their semantic content from the representational states of the individuals that produce and interpret them*.

Even if it seems reasonable to attribute collective beliefs that diverge radically from the beliefs of the individuals that compose a court or a corporation, these collectivities only exhibit *apparently cognitive activity* and this activity is reducible to "the cognitive states of individuals (including their construction of rules for combining individual activity in a principled way)" (Rupert Personal correspondence). Thus, in the case of court opinions and press releases, positing collective representations is explanatorily superfluous because the content of these public-language structures is reducible to the content of individual representational states (including their construction of rules for combining individual representations in a principled way). At this point, things look bad for the proponent of collective representation.

3. The Constituent Structure of Individual Representations

In the remainder of this paper, I argue that even on this most sophisticated reading, the argument from explanatory superfluity relies on too simplistic a view of the inter-level relations in individual cognition. While it is true that some nomic relations obtain between neurological states and properties of the world (e.g., in edge detection, color detection, phoneme detection, etc), most person-level representations derive their representational content from lower-level states that are themselves *already semantically contentful*. Let's begin by looking at a pair of examples: perceiving one's mother and perceiving motion.

Consider the claim that mental representations are grounded on perceptual or quasi-perceptual states. A perceptual representation of one's mother clearly counts as a case of mental representation, and it can plausibly be seen as a person-level representation built up from simple sensory representations. However, by looking carefully at the process by which a mental representation of one's mother is constructed, a number of interesting things that typically remain unnoticed about this reliance become more obvious.

Suppose that Alyssa sees her mother standing in the doorway of her favorite bar talking to the bouncer, and she turns to the person sitting next to her and says, "That's my mother!" In this case, Alyssa is in a state of perceptually representing the person in the doorway, and so long as the lighting conditions are suitable, we'll have good reason to take Alyssa's 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 Alyssa's cognitive architecture. Suppose we want to know what it is about Alyssa that gives here the capacity to represent someone *as her mother*. When we consider familiar sorts of representational states like representing one's mother, it is hard to these states as more than simple, homogenous lexical items—perhaps lexical items in a language of thought. On the basis of this supposition, there has been a tendency among philosophers to take a representation like Alyssa's and claim that this is just a complex thought built up from a MOTHER concept, some sort of demonstrative concept (for the 'that is' component of her thought), and some sort of possessive concept. 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 looks like.

Consider a person with Capgras delusion. 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, 1998, p. 1856). They seem psychologically intact, aside from the fact that they have an unshakable commitment to the belief that an imposter, a robot, or an evil twin has replaced someone to whom they are close (typically a parent, a spouse, or a pet). Suppose Alyssa has a stroke induced by a drug overdose and awakens in the hospital seeming

perfectly normal—until her mother walks in. At this point, suppose she 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 drug use. Now, it is no longer true of Alyssa that she believes that this is her mother, and the fact that her perceptual representation has changed calls for an explanation of what causes this change in representational capacities. The most widely accepted theory of the mechanism underwriting Capgras syndrome suggests a failure in the binding of visual representations and the affective representations that drive a feeling of familiarity.[5] In the presence of this sort of breakdown, we begin to see that the perceptual representation of one's mother is more complicated that we might have initially thought. In order to have a perceptual representation of someone *as one's mother*, at least two things need to occur. First, the visual system must be functioning properly in order to produce a representation that can be categorized as a representation of one's mother; but second, it also must be true that the affective response is the correct affective response.

Now, consider what would happen if Alyssa awakens with localized damage to her fusiform gyrus. In this case, Alyssa 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—that is, despite the fact that she can no longer see faces. The interesting thing about this case, however, is that the affective response to her mother's face may continue even though she is failing, on any sort of conscious level, to represent a face belonging to her mother at all (cf., Bauer 1984). In this case, the feeling of familiarity may persist at a fairly low level even though the she might no longer have the visual representation of this person as her mother.

The important thing to notice in these breakdowns is that representing a person *as one's mother* relies on a number of component processes. Even at the level of the individual, many of the representational states that are required in order to making our way through the world are composed from the outputs of various subroutines that are already producing representations. To put the point another way, many of the representations that we want to take to be genuine mental representations with primary intentionality in the case of the individual already supervene on component structures that are themselves already intentional. If the explanation of Capgras syndrome and prosopagnosia 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 one's mother since every step in the construction of a representation of one's 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 Jerry Fodor would likely put the point, if a theory can't allow representations of mothers, it's the end of the world!

To further bolster my case, consider a second example. Dale Purves and his colleagues (1996) offer a psychophysical explanation of the wagon wheel illusion (in which a wheel that is obviously moving in one direction suddenly appears to reverse direction) by appeal to the normal functioning of the visual system in producing representations of motion. They argue that rather than monitoring a continuous flow of information, a subcomponent of the visual system processes information about motion by producing a sequential ordering of discrete representations of a single object across temporal episodes. Here's how their account of the illusion works. Consider the sequential ordering of representations of the placement of spokes on a wheel with four equally spaced spokes. If the number of degrees that the wheel turns between samplings of the scene by the visual system generates procession (e.g., 95°), the movement will be viewed as a forward rotation. However, if the number of degrees that the wheel turns between samplings of the scene by the visual system generates precession (e.g., 85°), the movement will be viewed as reverse rotation (cf., Purves et al, 1996, p. 3694).

If this explanation is correct, the visual representation of the wagon wheel as moving in a direction reversed from its actual direction of movement can be fully explained in terms of 1) the properties of discrete and static representations, and 2) rules for the sequential ordering of representations of a single object across temporal episodes. Moreover, if Purves and his colleagues' are correct in their explanation of the visual representation of motion, this is how the visual system processes information about motion *in all cases*. However, the adoption of a superfluity argument would again raise a serious worry. There seems to be no need to posit a person-level representation of motion since every step in the construction of a representation of motion 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 sequential ordering thereof. Here the states are all representational, just as they are in the reduction of collective to individual representations. So, if superfluity arguments succeed in precluding collective representations they will also succeed in precluding individual representations of motion. Again, this seems disastrous for a cognitive science that operates on person-level representations.

4. Collective Representation and Collective Action

At this point, proponents of superfluity arguments against collective representation 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 a visual representation of motion and a visual representation of discrete, 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. 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. The question, however, is: if we take cognitive science to be directed at the explanation of behavior, is there good reason for retaining collective representations as explanatorily valuable structures.

Thus far, I've suggested that a theory of individual representation must allow for causal mediation by lower-level representational states; but I have yet to say much about the relation of this representational mediation to the production of action. Taking representation to be a relation of indication, or information-bearing, 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. As I suggested earlier, 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 and coordinating with one another 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 motor routines. 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 is incapable of representing 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.

Cognitive science 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, at no point during the production of the visual state prior to the final output in conscious monitoring do we have an adequate account of 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 ordinary philosophical accounts of individual representation have gone astray by focusing on high-level linguistic processes, the defense of collective representation has gone astray by taking static representations (e.g., court decisions and press releases) to be the only place worth looking for collective representation. Unfortunately, these states aren't the most promising avenue for developing an account of collective representation 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 representation is to consider 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 Edwin Hutchins case of the collective representation of the "fix cycle" in contemporary pilotage. The "fix cycle" is the cycle of activity in a navigation crew that establishes the location of a ship in relation to various sorts of landmarks, facilitating the computation of the trajectory of a ship (Hutchins, 1995, p. 117). The "fix cycle" is a computation that is implemented through the coordination of a number of computational processes, some of which are internal to individual crewmembers, others of which are external to these individuals. The interesting thing to note, however, is that the representation of a ship's location can only be produced through the interaction and coordination of a number of different lower-level processes—each of which is already in the business of producing representations. Moreover, although the media that are produced by each of the subroutines varies wildly, they are nonetheless capable of being brought into coordination with one another in order to produce a representation that can be used to direct the behavior of the ship.

The navigation system of a ship consists of a number of systems designed to be sensitive to a variety of one-dimensional constraints (Hutchins, 1995, p. 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 a state of coordination—often by way of taking repeated measurements—that a representation of a ship's location is recorded on the chart as a representation of the location of the ship by the navigation system. That is, it is only by way of the coordinated activity of a variety of systems that the a representation of the location of a ship can produced in such a way that it is usable for setting a course for the ship.

While it is true that none of the individuals on the ship actually does represent the location of the ship on her own, it is also true that none of the individuals *can* represent the location of the ship on her own. Because of the way in which training occurs in the U.S. Navy, the representations produced by various different subsystems are typically capable of being understood only by those who are trained to take measurements using a particular device. The crewmembers 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 ship's 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 that none of the individuals in the navigation crew represents the position of the ship. It's only the output of the navigation crew as a whole that represents the location of the ship.

A second case that suggests a promising model for distributing cognition across individuals becomes apparent in considering 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, p. 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 more 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 this data is 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 understanding of things leaves much to be desired. The propagation of information through a CSI team 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, p. 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.

5. A Final Objection

At this point, the view I am advancing faces a serious objection. Although these sorts of collective representations are produced by an integrated network of systems that produce semantically evaluable representations of their own, it is not clear that these collective representations depend on the representational states of individuals *in the same way* that a representation of one's mother or the visual representation of motion depend on lower-level components of the visual system. It is clearly unreasonable to deny that the representational states of the individual members of the crew are *necessary* for producing an authoritative representation of the location of the ship or a narrative for prosecution in the case of CSI. However, it is not obvious that we should treat the crewmembers of a ship and the members of a CSI team as components of a cognitive system *in the same sense* that edge detectors, face detectors, and affective response mechanisms are components of the visual system.

At this point, it will help to consider what it takes for something to count as a component of a mental process. We might begin by distinguishing between background states that are necessary for producing a particular mental state, and the core processes that make it the sort of state that it is (cf., Block, forthcoming). In developing his account of the mind as a cybernetic system, Daniel Bateson (1972) argued that the capacity for representation is best understood as the capacity to distinguish things resulting from the flow of information through a system. Building on this thought, we can begin to make sense of the move to systems-level neuroscience that underwrites my appeal to component processes of a mental state. For example, although many sort of processes are necessary for any mental activity whatsoever (e.g., unless the brain is regulating heart rate and representation a person will be unable to represent anything at all), other processes are necessary only for engaging in a particular cognitive task (e.g., representing a person as one's mother). At any point, numerous processes will be active in a brain; however, only some of these will be relevant to the production of a particular sort of representation. For example, states dedicated to regulating heart rate and respiration, facilitating

digestion, and proprioceptive updating are unlikely to be relevant in considering many of our representational capacities. These states will often be necessary for the continued operation of representational systems, however, does not entail that they are components of the system that produces a representation of motion. In order to count as a component of the representational system that is dedicated to the perception of motion or the perception of one's mother, something more must be the case.

Simple representations will be explicable in terms of structures with less component structure. For example, when examining the system responsible for distinguishing vertical lines from horizontal lines, the best explanation will be in terms of structures in the striate and extrastriate cortex that function to detect orientation. In this case, merely appealing to orientation detectors will be sufficient to explain how a person can distinguish vertical from horizontal lines. However, as the phenomena to be represented become more complicated, so do the mechanisms required in order to make the relevant distinctions. In order for a person to be able to distinguish a photo of her mother from photos of an FBI agent, her mail carrier, and a nonsense image with the same spatial frequencies as a picture of her mother, the visual cortex must produce representations, but so must the fusiform face area and the areas dedicated to the producing an affective responses to an image. Moreover, these component systems must be coordinated in such a way that they produce a complex representation *of her mother's face*. The point, here, is that when we explain what is required for a person to recognize someone as her mother, we must explain this in terms of the systems that are coordinated in such a way that she is capable of distinguishing her mother's face from faces of people who are not her mother and non-faces.

The lesson here is that making sense of a system as fulfilling a particular cognitive function is only possible by making reference to a particular task and a contrast class. Keeping these lessons in mind, we can begin to get a handle on why it is that the crewmembers of a ship and the members of a CSI team ought to be understood as components in the same way.

Suppose we want to explain how the crew of a ship is able to chart a course in a way that facilitates successful docking and prevents the ship from running aground or hitting another ship. Here, we have an intentionally specified task (docking the ship) as well as a contrast class (running the ship aground). In order to explain this phenomena, we must appeal to the crew's capacity to continually represent the location of the ship as it moves through the water as well as it's capacity to avoid shallow water and avoid other ships. There will be numerous things going on in a ship at any given time. For example, there will be people working to ensure the propulsion systems are operational, deckhands will ensure that the ship is presentable when it lands, and there must be people who are monitoring the flow of energy through the ship. Despite the fact that these people are active, and perhaps even must be active if the ship is to be able to land, the activity of these people does not distinguish the capacity of the crew to chart a course from anything else that the ship might do. However, in order to successfully chart a course that will allow a ship to successfully dock, systems must be in place that are dedicated to monitoring the rate at which the ship is moving and turning. Moreover, systems dedicated to monitoring the depth of the water must be operational, as must systems dedicated to tracking the location and course of other ships. Here is the important point. It is only through the coordinated activity of these various people that the intentionally specified action of successfully docking a ship can be explained. If any of the people fail to contribute the relevant representations, no authoritative representation of the location and trajectory of the ship will be produced, and this is why the crewmembers ought to be seen as components of the cognitive system that we might call the "navigation crew" *in the same way* that the systems dedicated to face perception and affective response must be seen as components of a person's representation of someone as her mother.

Analogous considerations hold for CSI. In order for a CSI team to produce a successful narrative for prosecution, as opposed to an unconvincing narrative filled with inadmissible evidence mere speculation, each member of the team must produce the sorts of representations that can be coordinated to produce a viable narrative. Because the task of a CSI team is to produce such a representation, informational structures must be in place in order to successfully engage in such a prosecution. These representational capacities are typically distributed across numerous individuals, and more importantly, the difference between producing an adequate narrative and an inadequate narrative turns on the coordinated activity of a variety of people, none of whom is capable of producing the narrative on her own.

With these considerations in mind, I suggest that collective representation should not be ruled out as a possibility within our best cognitive science. These cases of ship navigation and CSI suggest that there are collective behaviors that are intimately tied to the way in which information is propagated through the collectivity. Moreover, there are sufficient analogies between individual and collective representation in this regard to warrant further investigation into the genuinely cognitive activity of collectivities. 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. We could also abandon the level of explanation at which the location of the ship is determined through the "fix cycle". However, in doing so, we must make a parallel move in the case of the individual. The account of cognitive specialization at play in the collective case parallels the specialization we find in human psychology. If this argument succeeds, retaining individual representations because of their explanatory value warrants retaining collective representations because of their explanatory value as well. Since I find individual mentality to have great explanatory value, I propose that the explanatory value of positing collective representations should not be ruled out by appeal to superfluity arguments.

6. Acknowledgments

I would like to thank the following people for helpful an insightful comments on various drafts of this paper: Jacek Brzozowski, William G. Lycan, Jesse Prinz, Robert Rupert, Susanne Sreedhar, and three anonymous reviewers.

7. References

Baber, C., P. Smith, J. Cross, J. Hunter & R. McMaster (2006). Crime Scene Investigation as Distributed Cognition. *Pragmatics and Cognition*, 1, 357-385.

Bateson, D. (1972). *Steps to an Ecology of Mind*. New York: Random House.

Bauer, R. (1984). Autonomic Recognition of Names and Faces in Prosopagnosia: A Neuropsychological Application of the Guilty Knowledge Test. *Neuropsychologia*, 22, 457-469.

Block, N. (forthcoming). Consciousness, Accessibility, and the Mesh between Psychology and Neuroscience. *Behavioral and Brain Sciences*.

E. Bonabeau, G. Theraulaz, J.L. Deneubourg, N. R. Franks, O. Rafelsberger, J.L. Joly, & S. Blanco (1998). A Model for the Emergence of Pillars, Walls and Royal Chambers in Termite Nests. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 353, 1561-1576.

Dretske, F. (1981). *Knowledge and the Flow of Information*, Cambridge: MIT Press.

Durkheim, E. (1895/1982). *Rules of Sociological Method*. New York: The Free Press.

Durkheim, E. (1897/1997). *Suicide*. New York: The Free Press.

Freud, S. (1914/1975). *Group Psychology and the Analysis of the Ego*. New York: W.W. Norton and Co.

Gilbert, M. (1989). *On Social Facts*. Princeton: Princeton University Press.

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

Heylighen, F., M. Heath, & F. Van Overwalle (2004). The Emergence of Distributed Cognition: A Conceptual Framework. *Proceedings of Collective Intentionality IV*. Siena, Italy.

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

Knorr Cetina, K. (1999). *Epistemic Cultures*: How the Sciences Make Knowledge. Cambridge: Harvard University Press.

Le Bon, G. (1895/2002). *The Crowd: A Study of the Popular Mind*. New York: Dover.

Pettit, P. (2003) "Groups with Minds of Their Own," in F. Schmitt (ed.), *Socializing Metaphysics*, Oxford: Rowman and Littlefield.

Purves, D., J. Paydarfar, & T. Andrews 1996. The Wagon Wheel Illusion in Movies and Reality. *Proceedings of the National Academy of Science*, 93, 3693–3697.

Ramachandran, V.S. (1998). Consciousness and Body Image: Lessons from Phantom Limbs, Capgras Syndrome, and Pain Asymbolia. *Proceedings of the Royal Society of London B,* 353, 1851-1859.

Ramachandran, V.S. (2004). *A Brief Tour of Human Consciousness: From Imposter Poodles to Purple Numbers*. New York: Penguin.

Resnick, M. (1994). *Turtles, Termites and Traffic Jams: Explorations in Massively Parallel Microworlds.* Cambridge: MIT Press.

Rupert, R. (2005). Minding One's Cognitive Systems: When Does a Group of Minds Constitute a Single Cognitive Unit? *Episteme: A Journal of Social Epistemology*, 1, 177-88.

Schelling, T. (1971). Dynamic Models of Segregation, *Journal of Mathematical Sociology*, 1, 143-186.

Wegner, D. M. (1986). Transactive memory: A contemporary analysis of the group mind. In B. Mullen & G. R. Goethals (Eds.), *Theories of group behavior*. New York: Springer-Verlag, 185-208.

Young, A, I. Reid, S. Wright & D. Hellawell (1993). Face Processing Impairments and the Capgras Delusion. *British Journal of Psychiatry*, 162, 695-698.

---

[1] Although numerous objections must be met in defending collective mentality, answering many of these worries is ancillary to responding to the argument from the explanatory superfluity precisely because mentality requires representation. I contend that if the explanation of collective behavior *does not require* an appeal to collective representation, then appeals to collective mentality ought to be ruled out as a genuine possibility. In this case, the presence or absence of other purportedly mental features in collectivities will be irrelevant for the practice of cognitive science.

There are, of course, other ways in the argument that collectivities cannot possess representations could be developed. For instance, someone could argue that collectivities can't represent because they don't possess some feature that is exhibited by paradigmatic cognitive systems. Thus, a biosemantic theory like Ruth Millikan's were the most plausible account of individual mental representation, the objection would be that collectivities don't have the right sort of evolutionary history to count as representational systems. If Jerry Fodor's asymmetric dependence theory were most plausible, the objection would be that no states of collectivities are capable of standing in the right sorts of nomic relations to things in the world. Such arguments are troublesome only if the most promising theory of individual representation requires some feature that cannot be possessed by a collectivity. My task in this paper is to argue that this isn't the case.

Perhaps more troubling is the claim (advanced by philosophers such as John Searle, Collin McGinn and Uriah Kriegel) that the capacity for mental representation is contingent upon the possession of phenomenal or qualitative consciousness. I concede that there is a prima facie worry that collectivities can't be conscious, and if these philosophers are right, the lack of consciousness will be immediately damning to the defense of collective mentality. For a number of reasons, I don't find these arguments compelling. A more robust defense of collective mentality would require me to establish either that individual cognition does not require consciousness—so that the lack of consciousness in a collectivity does not rule out the possibility of collective mentality—or that collectivities can possess phenomenal or qualitative consciousness. I have arguments for both of these claims. However, developing them here would take me far beyond the bounds of this paper.

[2] There are, of course, disagreements about whether all mental states and processes are representational. The class of states and processes over which there is typically dispute are phenomenal states of consciousness and I will not address the possibility of collective consciousness in this paper.

[3] I say "roughly" because the environmental changes that affect the actual structure of a termite nest are incredibly complicated and bound to the particular history of the environmental structure. Things such as convection air streams within the nest, and the rate at which the queen emits pheromone are also important in explaining the eventual structure of a termite mound (cf., Bonabeau et al, 1998). However, focusing on the simplified model I use here does not impact the philosophical point, and since this model is familiar I will not complicate the discussion with a more complete model of nest construction.

[4] It is not clear that this distinction can be made in a way that *always* distinguishes between derived and underived content. However, we do need to distinguish states and processes that have their content in virtue of conventions for interpretation, and states and processes that have their content in virtue of something more intrinsic. The distinction here is clear enough, though the possibility of collective representation introduces a tension that might eventually require a new understanding of this distinction. After all, there might be some states or processes that have intrinsic content relative to one level of explanation, but derived content relative to another. My arguments in the remainder of this paper should make this worry clear.

[5] There is, of course, dispute over the precise nature of the mechanism. 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 the demonstration that this is a localized breakdown in binding affective information to face perception is concerned. 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.