Coordinated Behavior, Emergence, and the Explanatory Salience of Collective Representations

Whit Schonbein

College of Charleston
schonbeinw@cofc.edu

[Presented at the 2009 annual meeting of the Pacific APA, Vancouver, Canada, 10 April 2009. This version has an added footnote, but is otherwise identical to the presented version.]

We sometimes attribute intentional states to groups to explain their behavior. For instance, we say that Ford cut jobs because the company *expects low earnings*. How should we view such attributions? Is there a sense in which Ford *does* represent its future earnings, or is the attribution merely shorthand for an explanation that does not appeal to such 'collective representations' (CRs)?

Rupert (2005) argues that such attributions are likely superfluous, for two reasons: First, explanations appealing to CRs can be reworked to use the more 'conservative' ontology of individual representational states and causal relations between persons and their environment. Second, positing group mental states requires positing group minds, but our best theories of mental content are unlikely to apply to group states, suggesting that groups do not instantiate intentional states. Huebner (2008) addresses the first of these arguments. According to Huebner, Rupert's argument is too strong: If sound, it implies that appeals to individual representations (IRs) are *also* explanatorily superfluous. Furthermore, by drawing on Hutchins' (1995) analysis of ship navigation, Huebner argues that the states of groups engaged in coordinated behavior are no different in kind than IRs. Consequently it is no more superfluous to attribute content-bearing states to groups than it is to attribute them to individuals.

I pursue two theses: First, I argue there is an alternative interpretation of Rupert's argument that avoids Huebner's objection, *and* applies to the ship navigation example. Second, I attempt to defend the navigation argument by drawing on Clark's (1997) discussion of emergence in coordinated group behavior. Unfortunately, this initially promising move fails as well.

1. Explanatory Superfluity

Rupert (2005) holds that attributions of intentional content to groups are 'superfluous' in the sense that they are at best shorthand for explanations that do not posit such CRs. Huebner proposes that we understand this argument in terms of 'derived' versus 'non-derived' content. The content of a representation is derived if it is a principled function of the contents of the representation's constituents (Huebner, 2008, p. 100). According to Huebner, Rupert's claim is thus that attributions of CRs to groups are attributions of

derived contents, and hence can be analyzed into their constituent (non-derived) IRs.[1] Furthermore, this reasoning does not apply to IRs, because, " [whereas] nothing new *in kind* is introduced in moving from individual representations to collective representations … in moving from the physical to the intentional, something new in kind *is* introduced." (Huebner, 2008, p. 100, original emphasis) That is, appeals to IRs resist Rupert's argument because IRs have properties that are 'new in kind' and hence not tracked by other terms in the theory.

Huebner points out in response that this argument depends on IRs not having derived contents. But many IRs do. For example, he argues that cognitive deficits in facial recognition suggest that our MOTHER representation (an IR) has constituents that are themselves representational, including visual and affective representations (Huebner, 2008, p. 102). More generally, he notes, "[e]ven at the level of the individual, many of the representational states that are required in order to make our way through the world are composed from the outputs of various subroutines that are already producing representations." (Huebner, 2008, p. 102). If Huebner is correct, then Rupert's argument implies that many IRs are as superfluous as CRs.

Now, Huebner's point about our MOTHER representation is a (non-Fodorian) version of the view that mental states have a compositional semantics. A system of mental representation is compositional if the meanings of complex representations are a function of the meanings of the constituent representations along with their structure. Huebner's example clearly fits this mold: The content of MOTHER is a function of the contents of the constituent visual and affective representations, along with the structural relations between these constituents (Huebner, 2008, p. 102). So Huebner's challenge can be restated in more familiar terms: Many IRs have a compositionally derived semantics, and Rupert's argument implies that attributions of such contents are superfluous.

However, there is an alternative interpretation that avoids Huebner's objection. Rupert argues by example, discussing one popular putative CR: Court decisions. He points out that "the construction of the court's opinion, its legitimacy, and its effect on society can all be explained without invoking anything beyond the conservative ontology of individuals and their states (together with photons, sheets of paper, ink, etc.)" (Rupert, 2005, p. 6). For example, the procedure for calculating the CR is through a strict majority; cases where the court decision deviates from this mapping are explained by invoking other IRs including beliefs about the role of precedent; and the causal efficacy of the decision is explained by appeal to additional IRs, e.g., beliefs about the court's authority. Rupert concludes, "[p]ositing group cognitive states of the court does no causal-explanatory work, for there is none left to do" (Rupert, 2006, p. 7).

Notice that Rupert is concerned with explaining how the decision is (i) a function of individual mental states, and (ii) causally efficacious. This suggests the following (admittedly rough) principle:

---

[1] For example, Ford's expectation of low earnings can be analyzed into beliefs and desires of the managers and accountants comprising the company. Huebner describes Rupert's argument as follows: "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)." (p. 100-101). Note that the proposal being addressed in this passage is that CRs are linguistic entities such as press releases. However, the same argument could conceivably be deployed in opposition to other candidate CRs.

(R) An appeal to a representation is explanatorily superfluous when (1) there is a explanation of how the content of the representation is a function of its constituent contents, and (2) the causal efficacy of the representation is explained by appeals to additional representations of the same type as those constituents.[2]

When these conditions are satisfied, we understand how the hypothetical CR is related to members' mental states, and we understand why the CR has the causal role it does, all without appealing to anything above and beyond IRs and causal interactions.[3]

Huebner's MOTHER case does not satisfy both clauses of (R). The first is presumably satisfied by virtue of the representation's content being compositionally derived. But the second is not, because there are situations where the causal efficacy of MOTHER cannot be understood except insofar as it is a representation of one's mother. Suppose, for example, Ted is called to the hospital to identify a body. Upon viewing the deceased, Ted breaks down in tears. We don't explain Ted's behavior by saying he had some set of visual representations plus an associated affective representation. Rather, we say he broke down in tears because he recognized the deceased person in front of him *as his mother*. It is the fact that the IR has *that* content that explains Ted's behavior.[4]

At the group level, however, we must descend to the level of IRs to understand why a CR has the causal consequences it does. Suppose at their weekly meeting a club decides to order pizza, and this decision causes the club to call the local pizza parlor. Unlike the MOTHER case, where an appeal to constituent representations renders the explanation unintelligible, in this case such an appeal promotes our understanding: A club member makes the phone call because she wants to participate in the club, believes she should do as the others in the club ask, believes that the club should order pizza, and so forth. In short, the causal efficacy of the putative CR is best explained at the level of individual representations.

If this is correct, Rupert can reply by challenging Huebner's interpretation. Rather than relying on a distinction between derived and non-derived content, his argument rests on the fact that an understanding of the causal roles of both IRs and CRs resides at the level of individual representations.

---

[2] [Added after the conference presentation.] This principle as stated is clearly defective. First, there is a typo in (1), which should read 'an' rather than 'a'. Second, clause (2) is badly worded. The basic idea behind (2) is that, to be explanatorily salient, the causal efficacy of the putative collective representation cannot be explained by reference to individual representations of the same type. In particular, if the collective representation is a propositional attitude, then its causal role cannot be explained by appeal to the causal roles of the propositional attitudes of the members of the group.

[3] I make no attempt to defend principle (R) in this paper. My arguments are thus conditional on some version of the principle being defensible.

[4] Here's another example, borrowed from Dennett (1981): While playing chess, Ted brings his queen into play early in the game. Why did Ted move his queen out at that time? We might try to explain his behavior by appealing to the contents of the constituent representations comprising his QUEEN concept (whatever those are), to the contents of the constituent representations of other relevant concepts, and to his affective states. But such an attempt is doomed to fail: The best explanation is that Ted moved his queen early because he believed that by doing so he could achieve some goal (e.g., taking a rook, etc.). If we try to explain Ted's behavior by appealing to the constituent representations, we end up missing the relevant patterns of behavior, e.g., that Ted is engaged in a strategy the goal of which is to win the game.

2. Emergence

Rupert and Huebner agree that the issue is an empirical one, so there may nonetheless be cases where appeals to CRs are not eliminable. Huebner considers several candidates, the most well-developed being ship navigation (Hutchins, 1995). Huebner's strategy is to argue that states of the navigation crew are of the same type as IRs. Since it is not explanatorily superfluous to appeal to IRs, and CRs are in the same category of theoretical entity, it follows that it is not explanatorily superfluous to appeal to CRs.

Consider the MOTHER representation. As mentioned above, Huebner argues that MOTHER depends on the coordinated activity of task-specific representation-producing subsystems. Furthermore, if any of these subsystems malfunctions, the representation will no longer correctly guide behavior – its causal role is also a function of the component representations. The same is true of a ship's navigation crew, argues Huebner. Acquiring a representation of a large ship's location requires many specially trained crewmembers, each dedicated to gathering and transforming distinct types of information, e.g., the speed of the ship and the depth of the water. Huebner writes:

> 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. (Huebner, 2008, p. 109, original emphasis)

For Huebner, since both putative CRs and IRs depend on the coordinated activity of task-specific representation-producing subsystems, they are of the same general type, and hence appeals to either are equally respectable.

One might reply by challenging the relevance of this comparison between IRs and CRs. In particular, when we explain the behavior of the group, can the explanation be couched solely in terms of IRs and causal relations? The *prima facie* answer is *yes*: The process of generating the representation is highly structured and constitutes a procedure for calculating the CR from the intentional states of the crew, e.g., their beliefs about the readings they are getting from their instruments, how to use those instruments, and who needs to receive the information they acquire. Breakdowns in this process are explained by appeal to malfunctioning instruments, accidents, or to an individual's mistaken perceptions or beliefs. Similarly, the causal consequences of the CR are also understood as operating at the individual level: For example, the representation causes changes in the ship's direction, and this outcome is best explained by reference to, among other things, the captain's beliefs about the accuracy of information presented to her by the navigation crew, about her duties, and about the destination of the ship. In short, nothing in the passage quoted above suggests that an appeal to CRs in the case of ship navigation avoids superfluity as defined by principle (R).

Perhaps we can avoid this criticism by focusing on the notion of *coordinated behavior*.[5] Cases such as navigation are often given as justification for a claim about the realization of cognitive processes: A groups of individuals, each of whom *lacks* certain cognitive capacities, can coordinate their actions to manifest collective behavior inviting description in terms of these missing cognitive capacities.[6] Consider nest building by termites, effectively described by Clark (1997):

> All the termites make mud balls, which at first they deposit at random. But each ball carries a chemical trace added by the termite. Termites prefer to drop their mudballs where the chemical trace is strongest. It thus becomes likely that new mudballs will be deposited on top of old ones, which then generate an even stronger attractive force. … Columns thus form. When two columns are fairly proximal, the drift of chemical attractants from the neighboring column influences the dropping behavior by inclining the insects to preferentially add to the side of each column that faces the other. This process continues until the tops of the columns incline together and an arch is formed. (Clark, 1997, pp. 74-5)

Note that we can easily imagine invoking a CR to explain the behavior of the termite colony: Why are the termites building arches? Because *the colony wants to build a nest*.

Clark emphasizes that "[a]t no point in this extended process is a plan of the nest represented or followed. No termite acts as a construction leader. No termite "knows" anything beyond how to respond when confronted with a specific patterning of its local environment." (Clark, 1997, p. 75). In other words, individual termites do not token IRs representing the goal of their activity. This effectively *blocks* appeals to IRs with that content when formulating explanations of the colony's behavior – we should not endeavor to explain the nest building by attributing representations of some architectural plan to the individual termites for the simple reason that they don't *have* any such representations.

This also holds for the navigation example. Clark writes of ship navigation, "The overall (ship-level) behavior is not controlled by a detailed plan in the head of the captain. The captain may set the goals, but the sequence of information gatherings and information transformations which implement the goals need not be explicitly represented anywhere." (Clark, 1997, p. 77). This is echoed by Huebner, who writes, "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" (Huebner, 2008, p. 106). If no individual represents the location of the ship, and the only way to understand the behavior of the navigation crew is to posit such a representation, then we must make use of a non-superfluous CR.[7]

---

[5] At this point I am departing from Huebner's presentation in an effort to pursue a promising lead that seems implicit in his appeal to ship navigation.

[6] These cases are also sometimes offered in support of a related thesis, namely that an individual's cognitive processes are best understood as *extended* in an environment. The idea here is that, for example, the instruments used by crewmembers in taking measurements are not mere tools for assisting in thought, but rather are *constitutive* of those very cognitive processes. Since this is a thesis about *individual* representations or processes, I won't discuss it here.

[7] There's a slight difference here that bears mentioning: Clark's point is that no member of the crew represents the procedures involved in gathering and processing the information used to construct end

The argument can be further clarified. On the basis of these and other examples, Clark distinguishes between different types of emergent phenomena. The most promising one for the CR proponent is based on the notion of an *uncontrolled variable*. Uncontrolled variables "track behaviors or properties that arise from the interaction of multiple parameters and hence tend to resist direct and simple manipulation" (Clark, 1997, p. 110). An example concerns a mainframe computer that 'stumbles' when 35 or more users are connected. There is no explicitly represented user limit coded into the operating system – so the 'stumbling' threshold can't be directly manipulated. Instead, this variable arises from the interactions of the computer's components. Clark thus offers the following account of emergence: "Emergent phenomena [are] any phenomena whose roots involve uncontrolled variables … and are thus the products of collective activity rather than of single components or dedicated control systems." (Clark, 1997, p. 110).[8]

Deploying this notion of emergence, our revised reply is that a CR of the ship's location is an *emergent* phenomenon, and as such has no explicit existence at the level of any subsystem of the navigation crew. Therefore it cannot be explained away by appeal to IRs, and appeals to CRs are not superfluous.

Unfortunately, this form of emergence does not give us non-superfluous CRs. In fact, it suggests how to *avoid* positing CRs. Consider two interpretations of the termites: (1) While no individual termites represent a plan for the nest, the colony does; and (2) The colony constructs the nest without relying on any representation of that goal. The former interpretation posits a CR while the latter does not. Unfortunately, Clark's definition of emergence clearly favors (2): In the same way that no component of the mainframe explicitly represents the maximum number of users, no component of the termite group explicitly represents the nest. The building of the nest is not under the direction of a distinct 'variable' (an explicitly represented plan at the individual level). In fact, it's not under the control of *any* representation of the nest, even at the group level. The fact that the phenomenon is emergent suggests that to understand the behavior of the colony we should eschew the use of global representations and appeal *only* to IRs (e.g., individual's percepts of mudballs) and causal interactions (e.g., the influence of the chemical markers on the termites' sensory systems).

The same holds for the navigation example. The two interpretations of the navigation case are: (i) The navigation crew realizes a CR of the ship's location; and (ii) The navigation crew does its job without relying on representations of the set of processes by which this achieved. Now, it is true that the navigation crew produces, as an output, a representation of the location of the ship. But this representation is analogous to the termites' nest – it is not a candidate CR.[9] As in the case of the termite

---

product (the representation of the ship's location). For Clark, some crewmembers probably *do* represent the location of the ship – the captain, who sets the navigational goals, is a good candidate. Huebner, in contrast, claims that no crewmember participating in the information gathering and processing represents the location of the ship. I assume that Clark and Huebner are in agreement on all three of these points: (i) some crewmembers *do* represent the location of the ship, (ii) no member of the navigation crew represents the location of the ship, and (iii) no crewmember represents the entire set of procedures for gathering and processing information for the generation the representation of the ship's location.

[8] Clark goes on to slightly modify this definition to handle problem cases where it is possible to manipulate the presumed emergent phenomenon by adjusting a single parameter (such as convection rolls in a liquid being controlled by temperature). However, the revision does not affect the present discussion.

[9] As Huebner points out, "the defense of collective representation has gone astray by taking static representations (e.g., court decision and press releases) to be the only place worth looking for collective

colony, the crew's behavior is not under the control of any set of procedures represented explicitly at the individual or group levels. Instead, the behavior can be explained as the result of various IRs (which will differ depending on the subtask performed) and causal interactions (such as between instruments and individuals). It therefore seems that elaborating on the coordinated behavior aspect of Huebner's appeal to navigation fails to secure a reply to our initial concern: It is still the case that explanations appealing to CRs can be replaced by explanations that make use of a more conservative ontology of IRs and causal relations.

3. Conclusion

If my arguments are sound, then Rupert's initial skepticism is justified. But this does not mean that emergent behavior cannot be explanatorily salient. For example, we might explain the collapse of a house by appealing to the nest building of a colony of termites. Similarly, it may be that emergent phenomena do serve as content-bearing vehicles. But whether an emergent state is contentful is likely to be an independent issue. In this respect, Rupert's second argument – that no prominent theories of mental content apply to groups of humans – is germane: There must be an additional ingredient beyond coordinated behavior for intentional content to enter the picture. Thus the question is not whether groups exhibit emergent behavior, but rather whether they possess this additional ingredient.

References

Clark, A. (1997). *Being There: Putting Brain, Body and World Together Again*. Cambridge, MA, MIT Press / A Bradford Book

Dennett, D. C. (1981). 'True believers: the intentional strategy and why it works', in A.F. Heath (ed.), *Scientific Explanation: Papers based on Herbert Spencer Lectures given in the University of Oxford*, Oxford, Oxford University Press.

Huebner, B. (2008). 'Do you see what we see? An investigation of an argument against collective representation', *Philosophical Psychology* **21**(1): 113-127.

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

---

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" (p. 105). A candidate CR would thus be an intention to produce a representation of the ship's location, not the representation itself.

Rupert, R. (2004). 'Minding one's cognitive systems: when does a group of minds constitute a single cognitive unit?' *Episteme* **1**(3): 177-188. Page numbers refer to the version available on Rupert's website: http://spot.colorado.edu/~rupertr/Papers.htm