

Hybrid collective intentionality (extended abstract)

Thomas Brouwer¹, Roberta Ferrario² and Daniele Porello³

¹School of Philosophy, Religion and History of Science, University of Leeds, Leeds, UK

²Laboratory for Applied Ontology, ISTC-CNR, Trento, Italy

³University of Genova, Italy

Abstract

This is an extended abstract of Brouwer, T., Ferrario, R. & Porello, D. “Hybrid collective intentionality”. *Synthese* 199, 3367–3403 (2021). <https://doi.org/10.1007/s11229-020-02938-z>

Though there are different kinds of collectives, in this paper we focus on those that are not mere collections of individual agents, but display a behaviour that is unified enough to be aptly described as the behaviour of a single agent. In the literature it is fairly widely recognised that to this kind of agent one can ascribe mental attitudes. Such attitudes have been called by List and Pettit “corporate attitudes”.

In the philosophical debate, it is a fairly agreed upon fact that collective intentionality is not a metaphysically basic phenomenon, so the first research question concerns the basic facts that *ground* collective intentionality or, in other words, what are the facts which collective intentionality supervene on? Or, again, what are the facts in virtue of which collective intentionality is what it is?

Many approaches hold that what grounds collective intentionality are the individual intentionalities of the components of the collective. In the paper we argue that these are not enough, there are other facts, that may be heterogeneous in type, at the basis of collective intentionality. For instance, often the decision of the collective are taken thanks to some elements that is external to the collective. As an example, one can think about the coach of a football team choosing who will be the captain in a specific match, or a company that retains a legal firm to react on its behalf to any legal troubles it may encounter, not having the competences to take such decisions alone. So, the first research question is: what grounds collective intentionality?

The second research question is: can we build an approach on collective intentionality that can account for hybrid collectives, where for “hybrid” we intend collectives whose components are both human and artificial?

Our proposal for the synchronic account is to extend the metasemantic approach of radical interpretationism initiated by David Lewis and recently developed further by Robbie Williams.

In radical interpretation, we imagine an idealised interpreter who has access to all the basic facts about the interpretee and their situation—all the facts that don’t presuppose the target facts. They know the physical facts about them, they know about their behaviour (including the sounds and signs they make) and they know about the environment they’re in. But when it comes to the contents of their attitudes, the interpreter starts from zero (hence the term ‘radical’). The interpreter is idealised in certain respects: they are omniscient in the abovementioned spheres, and their ability to process information is unlimited. Their task is to ascertain the contents of the interpretee’s attitudes. In other terms, they stand in for an abstract interpretation function.

In the individual version of radical interpretationism, given certain preferences on the interpretee’s part, and certain beliefs about the world, the decision theory allows to predict that the interpretee, if rational, will take certain actions when receiving certain evidence. Knowing their actual evidence and actions, and on the assumption of rationality, one can reverse-engineer their beliefs and their desires. This, then, allows one to determine the interpretation that makes the interpretee most rational, which is, *ceteris paribus*, the best interpretation.

For the synchronic account of collective intentionality, we have adapted radical interpretationism, but many questions arise concerning what counts as the collective’s actions, evidence and intentions. With respect to actions, these should include:

- i*) A collective agent might perform an action by means of some individual agential component (i.e. a human agent, a non-human agent, or a partial agent) of the collective performing an action on behalf of the collective;
- ii*) A collective agent might perform an action by means of some subset of the collective’s components, which itself constitutes a smaller collective agent within the larger collective agent, performing an action on the larger collective’s behalf;
- iii*) A collective agent might perform an action by means of some subset of the collective’s agential components acting jointly on the collective’s behalf, but without constituting a full-fledged collective agent within the larger collective;
- iv*) A collective might perform an action by means of all of the components (agential or non-agential) acting jointly on behalf of the collective.

Among the states which realise collective intentions, we may find the following:

- i*) Intentions formed by either individual agents within a collective or by collective agents that are part of the collective, when those agents are acting in an official role associated with that collective.
- ii*) Joint intentions (as described by theories of joint action such as e.g. Bratman’s) formed by all or some of the collective’s members.
- iii*) Decisions reached by authoritative members of a collective (individually or jointly) the execution of which is delegated to other members of the collective.

Evidence and intention states work a certain way for a given collective because that collective is organised so that information flows into it and through it in certain systematic ways, and affects its behaviour in certain systematic ways. To put it another way, a collective agent has a certain cognitive structure, and insofar as this structure is stable, what makes for evidence and intention for that agent is also stable. The requirement that the ideal interpreter have access to (the correct characterisation of) the collective agent's evidence and intentions could also be expressed as the requirement that they know the agent's structure.

To fulfil their task of rendering the collective agent rational, the interpreter should, all other things being equal, assign that structure to the agent which allows them to maximise the rationality of the agent's behaviour. For the diachronic account, instead, we take Kit Fine's notion of qua-individual and extend it to the collective case.

We start by introducing Fine's notion of a qua-object. According to Fine, qua-objects describe the relationship between a material thing and the matter by which it is constituted. Interestingly, he includes social roles and their players among qua-objects, even though his theory isn't concerned with them especially. In a nutshell, Fine's qua-object is defined in this way: "Given any object x and description (property) ϕ possessed by x , we shall suppose there is a new object x qua ϕ , or x under the description ϕ . [...] Given such an object as x qua ϕ , we shall call x the *basis* and ϕ the *gloss*."

If we apply Fine's terminology to the problem of appropriately ascribing evidence and actions to collectives, we might say that the contentful states that matter in this regard are attitudes and actions of qua-individuals (or, more generally, the features of qua-objects).

The notion of a qua-object can also be applied at the collective level, by constructing a qua-entity out of a collection of entities, i.e. a qua-entity the basis of which is given by a collection of entities. In this case, the gloss is a complex description that states the relations holding among the entities that compose the collective. We have the abstract beginnings of an ontology of collectives: they are embodied collectives, i.e. complex qua-objects that are the result of imposing a certain gloss (a compound of relational properties) on a basis which is a plurality of appropriate individuals. Note that nothing in the picture so far forces us to think of the individuals as human agents; they could be any object that has some properties to contribute to a collective.

In its theory of variable embodiment, Fine also considers material things that can lose or acquire parts throughout their existence, like a car that requires to substitute parts periodically to maintain it in order. These are cases in which it is necessary to talk about a relation of temporary parthood, for which Fine introduces the notion of variable embodiment, to represent objects whose constitution can vary in time.

Our proposal is to apply the theory to describe temporal states of embodied collectives, i.e. the fact that at a certain moment (or period) t , some embodied collective is formed by members a, b, c, \dots (the basis) arranged in functional structure R , their relational principle, in Fine's terms (the gloss).

When considering the interpreter's endeavour diachronically, the task of the interpreter becomes more complex. In this case, we have to postulate the collective as a variable embodiment, manifesting possibly radical changes in its relational structure and its members throughout its history. Therefore, to single out the correct evidence and action at each time, the interpreter has to rely on the information encoded in the principle of variable embodiment, which provides, at each time, the correct structure R . That is, the task of the radical interpreter is now to select, among the candidate functional principles of variable embodiment, those that provide a chronicle of evidence and actions that makes the collective, through time, maximally rational.