The Intentional Synchronization of Individuals

Or: Do Groups Really Exist?

Irina Spiegel*

16.06.2026

1. Introduction

In this paper[1] I want to contrast an evolutionary-anthropological conception of the intentional synchronization of individuals with the philosophical conception of so-called *group agency*. The former is based on the research of the American anthropologist and developmental psychologist Michael Tomasello—specifically on his book *A Natural History of Human Morality* (2016)—while the latter rests on *Group Agency* (2011), the book by the philosophers Christian List and Philip Pettit. The juxtaposition I present involves a trivial but essential asymmetry. Whereas Tomasello and his colleagues use conceptually profound experiments to describe the evolutionary social mechanisms by which moral, distinctively human group behavior arises, List and Pettit attempt to render the epistemic reality of collective entities[2] plausible by means of conceptual-logical arguments and analogies. Tomasello's aim, then, is to integrate experimental findings on human collective behavior into a consistent hypothesis together with a coherent interpretation of it. List and Pettit's aim, by contrast, is to shore up—by the resources of linguistic logic—the conceptions and intuitions about group agency (i.e., collective agency; the terms *agent* and *agency* are defined in Part Two) that are anchored in everyday language. They posit collective entities (e.g., "the UN," "Russia," "Facebook") that possess person-like properties such as responsibility. The two philosophers thus proceed not from a merely metaphorical and pragmatic reading of sentences such as "Russia is the aggressor and

*irina.spiegel@lrz.uni-muenchen.de

[1] The article will be published in German in the anthology edited by Jörg Noller: Kollektive Personen. Mentis, 2026.

[2] The quotation marks are used here on this single occasion to indicate that, at this point, it cannot yet be known what is and will be meant by this concept, and that one should not succumb to a preconceived definition which—as the argument's thesis holds—will prove untenable. Where terms are taken to be used with a meaning that is clearly not endorsed here, and where the point is precisely a revision—one that is, or ought to be, the upshot (e.g., "intentional states" or "the mental")—these too are placed in quotation marks on their first use. Where, however, a term carries a thesis that is explicitly at issue and that is not endorsed, the term is not placed in quotation marks (e.g., group agency or collective entity). The same holds for terms that are not explicitly at issue but whose use is contested (e.g., responsibility). The aim is to signal that using these terms does not automatically secure assent to them. In addition, quotation marks are used for sentences or parts of sentences that are specifically referred to in the text (e.g., quotations or titles).

must be sanctioned as a collective agent," but from a literal one. On Tomasello's reading, what happens here is merely that individuals synchronize with one another in waging a brutal war of aggression in violation of international law. From this perspective, the synchronized individuals are to be distinguished from the posited group entity "Russia," with its collective personhood and responsibility. It will emerge that the logical attempt to render the posited collective agents plausible as entities—or as collective persons—is ultimately unconvincing. And it is, moreover, scarcely compatible with the natural and social sciences.

2. An Evolutionary, Social-Psychological Account of the Intentional Synchronization of Individuals (*Tomasello*)

What justifies introducing a scientific account into a philosophical argument in the first place? What legitimates appealing to current approaches from the scientific study of collective human behavior within a non-empirical thesis—namely, the thesis that the existence of collective persons or entities cannot be established by purely conceptual-logical means? This scientific orientation rests on so-called scientific realism, which the modern sciences have presupposed (usually implicitly) since Galileo Galilei at the latest. The basic core of this implicit (though not necessarily naive) realism is that whatever our best empirical—and, in application, also mathematical—theories describe as consistently as possible is assumed, or is to be assumed, to exist. The best theories are those that rest on falsifiable hypotheses and, through measurement and replication, elicit knowledge from reality, so as to fit it consistently into a larger or broader scientific context.[3] One of the best theories of the emergence of life, and of social life as well, remains the theory of evolution. One of the best evolutionary hypotheses of collective life, in the sense just given, is in my view the one advanced by Michael Tomasello and colleagues. It is called the interdependence hypothesis. In his book *A Natural History of Human Morality*,[4] Tomasello assembles the currently relevant findings from primate research and from research on human infants in order to trace, as coherently as possible, a scientific theory of the evolution of morality. He understands morality as a special cooperative—and hence collective—form of behavior of *Homo sapiens sapiens*.[5] At its core, the interdependence hypothesis holds that human beings, as ultra-social creatures, were able to adapt to their environment and survive because they synchronized in a special, that is, cooperative way. *Homo sapiens sapiens* is distinguished by a specific—"moral"—form of cooperative group behavior, one that stands in tension with the competition-based or aggressive forms of

[3] Further reflections on the question "What is a good theory?" seem no more necessary here than usual, since that question is not specifically at issue.

[4] Tomasello 2016.

[5] "[. . .] [H]uman morality is a form of cooperation [. . .]." Tomasello 2016, 2.

human group behavior.[6]

In the interdependence hypothesis, the concepts of shared (or joint) intentionality and of collective intentionality play a key role. In using them, Tomasello does not employ *intentionality* in a mentalistic sense. The expression broadly designates nothing more than a general feature of goal-directed cognition—of the cognitive mechanisms of goal pursuit—in humans and other animals. Because philosophers all too often scent reductionism, or even scientism, in the scientific use of such mentalistic terms, Tomasello has met with much criticism from philosophy—criticism that, however, often turns out to be unwarranted. For Tomasello does not, at the theoretical level, touch the problem of consciousness or of the mental. There is a large debate in philosophy about what "intentional states" really are and how they relate to physically irreducible phenomena of the mental. But it is not Tomasello's business to pursue the (often overly scholastic-seeming) disputes in the philosophy of mind. Nor do they play any role in my paper. Tomasello is not concerned with a theory of mental or intentional states, but with an evolutionary theory of the social—and especially the moral—behavior of human beings as an ultra-social species. He thus uses mentalistic expressions such as *intentionality*, but also collective expressions such as *plural-agent "we"*,[7] not literally and hence only pragmatically—that is, in a simplifying and abbreviating manner—without marking them with quotation marks throughout. This is often unnecessary, since on first use he introduces these expressions precisely, by means of behavioral-biological, social-psychological, or cognitive-scientific concepts. So when Tomasello uses such terms as cannot be cashed out empirically—for example, *intentionality*, but also *plural agent*—he is not thereby explicitly positing their existence as entities. It will become apparent that the philosophers List and Pettit, by contrast, want to derive from the use of these terms the existence of the phenomena so designated, and then to buttress that existence further with the formal language of logic.

Back to the interdependence hypothesis: for the anthropologist and developmental psychologist Michael Tomasello, shared (or joint) intentionality and collective intentionality, as social-psychological forms of behavior, constitute two essential stages in the evolution of humans and of human culture. Only this two-step development of social cognition—of social behavior—gave rise to moral, or more generally normative, group behavior. Tomasello formulates this as follows:

> The interdependence hypothesis [...] is that this took place in two key steps, both of which involved new ecological circumstances that forced early humans into new modes of social interaction and organization: first collaboration [based on joint intentionality, I.S.] and then culture [based on collective intentionality, I.S.]. The individuals who did best in these new social circumstances were those who recognized their interdependencies with others and

[6] Ibid.

[7] Quotation marks Tomasello's; Tomasello 2016, 3.

acted accordingly [...].[8]

2.1 Shared and Joint Intentionality

The first step came with the emergence of the genus *Homo* about two million years ago, when a change in ecology confronted the early hominids with the alternative of hunting together or starving. In order to coordinate their joint activities, distant ancestors of present-day humans developed the complex capacity for shared intentionality, which enabled them to synchronize with partners toward a common goal (joint intentionality) and then to pursue it together (joint agency). The groups of these "early humans"[9] are still relatively small. Their collaborative behavior is dyadically structured: everyone coordinates with everyone else, everyone knows everyone and their individual characteristics, and everyone recognizes everyone else as of equal standing (equivalence between "me" and "you"). They know that everyone knows that this is their common aim or intention. They synchronize toward a common goal—for example, "We are now hunting this antelope." In doing so, each one, with his own individual characteristics, takes on a role in this joint enterprise: "Each individual is both the 'we' that is pursuing with her partner a joint goal (in joint attention) and at the same time an individual that has her own role and perspective."[10] Participation in the collaboration bound the partners to one another in a unique way: they now formed something that Tomasello refers to metaphorically as a "we" or "group agent." This special group behavior does not arise where there is no coordinated, common intention. When people run to the same shelter in the face of a storm, for instance, the search for shelter is not their common intention but the individual intention of each. Systematic observation of chimpanzees, the closest relatives of present-day humans, suggests that chimpanzees synchronize without joint intentionality (that is, without a common goal), much like human behavior during a storm. It has been shown that when chimpanzees in the wild hunt capuchin monkeys, they pursue no common goal and form no common intention; rather, each has the egoistic goal of catching the prey for himself, while a spatial coordination occurs ("individualistic coordination"[11] or "group behavior in I-mode"[12]). This interpretation is plausible because it is shown with significant frequency that, after seizing the prey, chimpanzees each try to grab it entirely for themselves, with the most dominant and strongest animal in the group, as a rule, getting the most. According to Tomasello, this is not cooperation defined by a common goal.[13] Collaborative

[8] Tomasello 2016, 3.

[9] The quotation marks used above ("early humans") indicate that a term or group of terms is not to be understood literally, but is self-explanatory in context and used in a merely abbreviating, if intuitive, way. Tomasello sometimes uses them similarly, to indicate that a term is meant to denote a new kind of cognition or behavior of individuals that arose (over the course of time, i.e., evolution).

[10] Quotation marks Tomasello's; Tomasello 2016, 50.

[11] Tomasello 2016, 26.

[12] Tomasello 2016, 27.

[13] Ibid.

experiments with chimpanzees, compared with the same experiments with human infants, suggest that an essential distinguishing feature of humans is that they cooperate—that is, that they can synchronize dyadically toward a goal with a common aim. According to Tomasello, this set off a dynamic toward culture in human evolution. When, in dyadically structured groups, the cooperating partners have accomplished something together (on a hunt, say), this experience becomes common ground—that is, they now mutually know what each is to do at the next antelope hunt. Each knows his own role *and* the role of the other:

> These common-ground role ideals may be thought of as the original socially shared normative standards. These ideal standards were impartial in the sense that they specified what either partner, whichever of us that might be, must do in the role. Recognizing the impartiality of role standards meant recognizing that self and other were of equivalent status and importance in the collaborative enterprise.[14]

2.2 Collective Intentionality

The second evolutionary step, according to the interdependence hypothesis, consists in the further development of joint intentionality into collective intentionality—a new form of group behavior, or of the socio-cognitive capacities of individuals.[15] It occurred about 150,000 years ago, as groups grew ever larger and the number of human populations increased rapidly. Groups split into smaller ones that remained united at the tribal level, or they formed new tribes that then competed with other tribes for resources. Individuals now acted within their own cultural group without having to coordinate with one another personally any longer. The dyadic common ground was no longer necessary, nor was it any longer sufficient for social life in an unsurveyably large cultural group. In order to coordinate group activities cognitively and to secure, motivationally, a certain measure of social control, the modern ancestor of present-day humans developed new cognitive capacities:

> Cognitively, what modern humans did to adapt their new social reality was to transform a joint intentionality geared for dyadic collaboration into a collective intentionality geared for cultural collaboration.[16]

The dyadic, personal I–you equivalence (egalitarian partnership) in the joint intentionality of the small groups from the first developmental step evolved into an anonymous equivalence or similarity among all members of a much larger cultural group. The

[14] Tomasello 2016, 4.
[15] Tomasello 2016, 92 ff.
[16] Tomasello 2016, 92.

result was that the conventional practices of a culture contained normative standards (*Normierungen*) that everyone knew (or had to know), and everyone knew that everyone knew that they held for everyone as a cultural common ground:

> Collective or cultural common ground is the basis of—it is almost the definition of—culture.[17]

The result is collective intentionality: the socio-cognitive capacity to deal with the implicit normative standards of a larger (cultural) group (cultural common ground)—a behavior that did not have to be coordinated dyadically in advance between particular individuals, because the implicit knowledge of "how *we* are to behave in this or that situation" is already cognitively encoded and, as such, is trained into the young early on. This cultural common ground, consisting of the conventions and normative standards (norms, laws, and unspoken rules) of a culture, is thus, as a rule, anchored implicitly in people's behavior and cognition. According to Tomasello, individuals perceive such normative standards as "objective."[18] That is, they actually perceive them consciously only very rarely—up to the point at which mature group members either more or less consciously affirm them or explicitly (more or less radically) contradict them. These two stages of human evolution thus encompass highly complex socio-cognitive and moral-psychological processes (obligations, social controls, sanctions, etc.) that are today being described in ever greater detail and in interdisciplinary-anthropological terms. Despite this complexity and interdisciplinarity, there is a common denominator here—a general pattern that recurs throughout nature: the aggregative pattern of the synchronization of the many. Thus evolutionary anthropology, too, works with aggregative—that is, mathematical—theories and models, for example in evolutionary game theory and in theoretical or experimental approaches within collective choice theory. Tomasello himself is not a mathematical modeler,[19] but the group phenomena he describes can be translated into aggregative patterns, so as to describe more accurately and precisely the basic structure of the moral, or normative, group behavior of individuals.

2.3 Aggregative Modeling (*Judgment Aggregation Theory*)

If one abstracts from this (behaviorally well-supported) description of evolutionarily developing collective intentionality as a special group behavior of individuals, its fundamental moral—or, more generally, normative—structure[20] comes into view: the aggregation of individual intentions into a collective intention. It thus turns out that this

[17] Tomasello 2016, 93.

[18] Tomasello 2016, 85 ff.

[19] Tomasello does, however, speak of a "mathematical point"—a special point at which the balance between egoistic and cooperative rationality tips over into the special cooperation of morality, with egoistic interests being sublated (*aufgehoben*) in cooperation. Cf. Tomasello 2016, 17.

[20] In his book, Tomasello is explicitly concerned with the natural history—that is, the evolution—of human morality.

normative behavior can be defined and modeled aggregatively. This is, in fact, nothing new. Aggregative-dynamic modeling has long been commonplace, for example, in evolutionary game theory.[21] What Tomasello (alongside many proponents of evolutionary ethics) attempts to support experimentally, and what becomes relevant for what follows, is that moral behavior is only an aggregative (group) behavior of individuals—albeit a highly complex one, finely tuned by evolution. As such, it can be captured precisely and modeled mathematically. This approach dispenses with any philosophical postulation of further entities, including collective ones. The underlying ontology is anchored solely in individuals and is therefore not inflated unnecessarily. The idea to be pursued further in the future, in order better to understand or model normative phenomena, is to use the most general aggregation theory—judgment aggregation theory (JAT)[22]—to explore the basic possibility of grasping norms, or morality, aggregatively, as a special, intentionally synchronized group behavior. JAT need not—and cannot—be introduced here.[23] The following basic schema of aggregation is sufficient for what follows:

Input = individual intentions → Output = synchronized intention = norm

With respect to this schema, joint and collective intentionality—as two specific kinds of human group behavior—differ in their input (*agenda*), in the *constraints* that restrict the agenda, and in their respective aggregation rules. The input may consist of various intentions of the individuals (e.g., some want to hunt hares, some antelopes). The *constraints* are logical or factual conditions on a consistent combination of different intentions. Thus it may, for example, be really impossible for a given group, under given environmental conditions, to hunt hares and antelopes at the same time, because the particular group or environmental properties are simply as they are. The group would have to decide; otherwise, hunting both animals at once would be instrumentally irrational. A rational decision or aggregation rule might then be one that, in the (egalitarianly organized[24]) groups, mostly takes one of the many forms of the majority rule. In the light of this basic aggregative schema, the synchronized or normative behavior of individuals—with respect both to joint (dyadic, personal) and to collective (anonymous, impersonal) intentionality—can now be understood as explicit (conscious, articulated) or implicit (unconscious, tacit) aggregation. Individual group behavior, as the behavior of the individuals of a group, is the result of an explicit or implicit aggregation of the intentions of the interdependent human individuals, who, explicitly or implicitly and unanimously, synchronized at least on this: the wish to be a *member* of a (this) group (intention). The thesis, minimalistically stated, is therefore that, for joint or collective intentionality, there must be at least

[21] For example: Smith 1982; Sigmund 1993.

[22] The approach is more general than social choice. Cf. Dietrich 2007.

[23] A compact introduction to this theory can be found in Grossi/Pigozzi 2014.

[24] Cf. Boehm 1999. He seeks to show that early human tribal societies were more egalitarian in structure than the troops of chimpanzees (our closest relatives), among whom the aggregation rule "dominance of the alpha animal," or "dictatorship," remains decisive to this day.

the one unanimity—that of being a group—since otherwise group behavior, or intentional synchronization, would not come about in the first place. Every further group behavior, or every further group property, then follows aggregatively, by means of an aggregation rule generally more complex than the unanimity rule. This intentional aggregative behavior is thus the source of all moral or normative phenomena. Over the course of natural (cultural) evolution, they form the norms, including the moral norms (*output*).

List and Pettit, too, develop their arguments against the background of judgment aggregation theory. They interpret the *output* of the aggregation quite differently, however—as a *group agency* detached from individual intentions and activities, to which they also ascribe collective responsibility, entirely in the sense of *collective personhood*.[25] For them, *group agency* and *collective personhood*, as *real* collective entities, supervene on the individual intentions, just as the mind supervenes on the physical substrate (neurons or elementary particles)—that is, a new, second (group) level *becomes* real and is not identical with the individual level. The following, second part now examines whether the central arguments that List and Pettit develop for this interpretation prove sound. Is the view that group properties are more than merely the properties of the individuals of that group really convincing?

3. The Philosophical-Logical Justification of the Existence of Collective Agents (*group agency*)

In the preface to their book, List and Pettit pose the following key question:

> Can groups be unified rational agents over and above their individual members?[26]

To pursue this, they first characterize *agency* quite generally as follows. *Agency* is a *system* that exhibits three features:

- It can represent its environment (that is, it has *beliefs* about how things in the world *are*).
- It has motivations or intentions (that is, *desires* about how things in the world *ought* to be).
- It has the capacity to process its *beliefs* and *desires* in such a way that it is able to *intervene* in its environment when the latter does not accord with its *desires* (instrumental rationality).[27]

[25] List/Pettit 2011, 170.
[26] List/Pettit 2011, vii.
[27] Ibid., 20.

They now transfer these three features to the group:

> It is a group that exhibits the three features of agency [...]. However this is achieved, the group has representational states, motivational states, and a capacity to process them and to act on that basis in the manner of an agent. Thus the group is organized so as to seek the realization of certain motivations in the world and to do so on the basis of certain representations about what that world is like.[28]

Building on this, List and Pettit attempt, in several steps, to establish that this transfer is not merely metaphorical—even though everyday speech often makes metaphorical use of "collectives" (that is, collective designations), as in "Germany has decided to supply Ukraine with tanks." For both philosophers, a non-literal usage is not yet a reason to deny group agents reality in general, as the proponents of eliminativism (or non-realism)[29] and of singularism[30] do.[31] List and Pettit also reject so-called emergentism with respect to *group agency*: in the end, it presupposes a mysterious force such as Rousseau's *volonté générale*.[32] List and Pettit take themselves to escape both extreme positions—the Scylla of emergentism and the Charybdis of eliminativism—by means of supervenience:

> If emergentism makes group agents into non-supervenient realities, eliminativism denies the reality of group agents altogether and so denies that the issue of supervenience even arises in any interesting way. Where emergentism makes group agents into hyper-realities, eliminativism makes them into non-realities. On this view, group agents are not entities of any kind and thus not even candidates for an emergent or non-supervenient status.[33]

The conception of *group agency* that List and Pettit advocate is inspired by rational and social choice theory, and these theories proceed from so-called "methodological individualism," which Karl Popper already espoused.[34] The core of this methodological position is to grasp and describe collective phenomena solely in terms of individuals—that is, without any reference to "collectivist" forces or entities. List and Pettit now want to show that this individualism is compatible with their form of non-eliminativism regarding the existence of *group agency*. To this end, the two philosophers develop three main lines of argument, by which the *existence* of group agents is to be rendered plausible and logically derived without any new mysterious force—that is, in keeping with methodological

[28] Ibid., 32.
[29] On group reality, cf., e.g., Quinton 1975.
[30] On the sole reality of individuals, cf., e.g., Gilbert 1989.
[31] Ibid., 3 ff.
[32] Ibid., 73.
[33] Ibid., 75.
[34] Popper 1945.

individualism.[35] They call the first strand the "positive,"[36] the second the "negative,"[37] and the third the "normative claim."[38] In what follows, these three lines of argument are sketched *in nuce* and immediately called into question.

3.1 The Positive Argument (*parsimony*)

In this line of argument, List and Pettit claim the following:

> From a positive perspective, one can often achieve greater descriptive and explanatory parsimony by viewing a collective as a single agent, acting in pursuit of a single set of desires, in accordance with a single set of beliefs.[39]

The two philosophers thus assume that the behavior of a group can be described, or even explained, better and more parsimoniously if one disregards the attitudes and intentions of the individuals of that group. Unfortunately, they give no example here that might show how a group agent can be described better and more parsimoniously entirely without reference to the individuals involved. They offer only the following analogy, varied several times: just as one could not describe an individual's mental and intentional states on the basis of the brain's neuronal activity, so too one could not describe group activity on the basis of the activity of individuals.

Now, this "analogy" is, in general, the least suitable one imaginable—something taken up in more detail in Section 3.2—and to choose it specifically here is quite unnecessary, since this argument is, after all, only about parsimony. If, however, something similar is indeed seen here, then it is clear why the path appears so positive. Drawing on the brain–mind problem as a prime example is, in this light, indeed very parsimonious.

While we have, to date, no (natural-)scientific theory of the mental, there are social-scientific and anthropological theories and hypotheses that can plausibly derive collective behavior from the description of individual behavior. Tomasello's interdependence hypothesis is one such hypothesis—even if, pragmatically, one occasionally abstracts from individuals (for instance, in describing the interaction between different groups) in order to speak more compactly. In principle, however, interactions within a group as well as between two groups can be traced back to the behavior of individuals. It may, under certain circumstances, merely become complicated to put into words, but not impossible. In short: this positive, or epistemic, argument is of a purely pragmatic nature and does not necessarily suggest the existence of "literal" group agents or collective persons. The epistemic (and precisely purely pragmatic) *parsimony* concerning collective agents that

[35] That the two philosophers are concerned with a derivation of the existence of collective agents is betrayed by the following sentence: "Hence group agents exist." List/Pettit 2011, 75.
[36] Ibid., 5.
[37] Ibid., 6.
[38] Ibid., 11.
[39] Ibid., vii.

List and Pettit propagate stands opposed to the ontological *parsimony* of the scientific realism of synchronously acting individuals.

It is obvious that List and Pettit's approach is not a fundamental theory of human group behavior or of *group agency*—if only because it is not a theory attuned to empirical research. Perhaps the following analogy can illustrate the point of criticism: the laws of thermodynamics are more parsimonious with respect to certain intended applications. But they can be reduced entirely to the far more complex laws of statistical mechanics. It is far more complex to describe the behavior of each particle statistically than merely to describe the aggregate of particles by means of thermodynamics with its ideal gas laws. It does not seem to be ruled out in principle that, with more far-reaching, AI-assisted anthropological research, the collective behavior of human individuals will become ever more describable in terms of their individual agency—even if the "protocol" will become ever more complex—with the consequence that the individuals of the groups can be better understood. And that is the only thing available for interaction or intervention.[40] In this respect, the pragmatic orientation of List and Pettit's positive argument is, after all, to be welcomed here:

> [The] positive claim [...] is that this explanatory ascription of agency to groups opens up a distinctive way of relating to and interacting with such groups, as is implicit in the role that *talk of group* agents plays in ordinary as well as social-scientific *discourse*.[41] (emphasis I.S.)

And further:

> Although their agency depends on the organization and behavior of individual members, as individualism requires, they display patterns of collective behavior that will be lost on us if we keep our gaze fixed at the individual level. And to lose sight of those patterns is to lose an important *source of guidance as participants* in the social world.[42] (emphasis I.S.)

Their concern is not to let the commonalities among individuals of different groups, and among individuals of one group, go unrecognized, and to discern patterns for the orientation of individuals in the social world ("source of guidance as participants in the social world"). There is nothing wrong with that, except that opening up levels of lifeworld description does not lead to levels of existence. This can lead to serious misunderstandings, as becomes evident in the "analogy" of the so-called body–mind problem. But this is immaterial here, since "talk of group agents" is, after all, only about parsimony and pragmatism, which obviously cannot be an argument for existence.

[40] See: "Agency is a system," (b).
[41] List/Pettit 2011, 5.
[42] Ibid., 6.

3.2 The Negative Argument (*supervenience*)

To defend the view that group agents really exist, List and Pettit must show how a unification of individual intentions can take place at the collective level, and this despite the fact that the group members have their own (often divergent) intentions. To explore this logical possibility, they draw on judgment aggregation theory (JAT), within which the majority rule, as an aggregation rule, has been studied most intensively.

The already far-advanced (theorem-based) exploration of the logical space of majoritarian aggregation via JAT[43] shows that there is a surprising, recurrent logical impossibility of deriving a majority-based group intention from individual intentions. This impossibility is termed the *discursive dilemma*: namely, the finding that even when all individuals bring consistent intentions to the vote, it is impossible, by means of the majority rule, to generate a consistent and unambiguous collective intention:

> The 'discursive dilemma' explored [...] shows that even perfectly consistent individuals can deliver inconsistent majority votes on connected issues. Thus a collection of people committed to act on majority opinion across a variety of issues may find itself committed to conflicting courses of action. Such a collection, as we show, could not generally satisfy the conditions for agency.[44]

In their book *Group Agency*, List and Pettit suggest that this result, proven by several theorems, is only the tip of the iceberg. They take this result as a fixed logical fact and use it to ground the emergence of the real, existing, autonomous group agent:

> [...] majority rule is not the only form of organization that fails to offer an independent, implementable voice individuals might authorize. It turns out that no plausible, equally simple process can provide such guidance, where the collective attitude on each issue is determined as a straightforward function of the individual attitudes on it. What we describe as the 'discursive dilemma' is just an illustration of a more general difficulty. And that difficulty arises with any plausible, issue-by-issue method of aggregation, majoritarian or otherwise, that is designed to deliver a suitable collective output on each issue from individual inputs. [...] If a group agent is to display the rationality that agency requires, its attitudes cannot be a majoritarian or other equally simple function of the attitudes of its members. The group agent has to establish and evolve a mind that is not just a majoritarian or similar reflection of its members' minds; in effect, it has to develop a mind of its own. This gives rise to the kind of autonomy that we ascribe to group agents.[45]

[43] In the literature, it is also referred to simply as "democratic aggregation." Cf. List 2011.
[44] List/Pettit 2011, 8.
[45] Ibid.

Thus, by way of the impossibility theorems of aggregation, a "group mind" is introduced ad hoc as a necessity, so that a group agent can come about—rather than allowing that the group members, whether aware or unaware of the impossibility of being a group in the strict sense, nonetheless explicitly or implicitly assent to the (an) output, as evidently happens in reality.

They nonetheless develop further the thesis that the group agent must now possess a mind ("The group agent [...] has to develop a mind of its own."), by now transferring considerations of the supervenience relation from the philosophy of mind into their argument—that is, into the negative results from JAT concerning the discursive dilemma—construing the aggregation rule between input and output as *supervenience*: the properties of a collective agent supervene on the properties of the individual agents (their respective individual intentions) just in case a change at the collective output level is always accompanied by changes at the individual input level, but not necessarily vice versa.[46] In short, List and Pettit attempt to reformulate the special (purely logical) dependence between input and output by means of supervenience, so that—in the sense of methodological individualism—individual attitudes or intentions do play a determining role for the group mind, while the latter also detaches itself from individual determinations and thereby becomes *autonomous*. In their eyes, the becoming-independent of the intentional group agent or group mind occurs entirely without any mysterious or metaphysical emergentism. But what, in fact, is supervenience? It seems to be merely a logical (not causally explanatory) special material implication, or else it additionally expresses a special modality, or modal force, though it is not clear what is actually being described or explained thereby. Or it simply designates an emergence which, given an ontological turn, would once again suggest a mysterious force, if one were not content with the ignorance here—as, indeed, List and Pettit are. If, however, we were to accept the logical reading for the moment, where would it really lead with respect to the reality of *group agency*?

To develop this argument, the two philosophers formulate the following central thesis:

> The supervenience thesis. The attitudes and actions of a group agent supervene on the contributions of its members.[47]

They apply this thesis to a few important impossibility findings from JAT, run it—with respect to supervenience—through all the relevant theorems, and arrive at the following (cryptic) result:

> Under the premise-based procedure, the supervenience relation between individual and group attitudes has not only a holistic as opposed to proposition-

[46] The supervenience relation in the simple, modal definition: "A set of properties A supervenes upon another set B just in case no two things can differ with respect to A-properties without also differing with respect to their B-properties. In slogan form, 'there cannot be an A-difference without a B-difference."' McLaughlin/Bennett 2018, 1. In a non-modal definition, "can" would be dispensed with.

[47] Ibid., 66.

> wise character but also an additional property. The individual attitudes on the premises alone are sufficient to determine the group attitudes on all the propositions. Thus the group attitudes on the conclusion are 'autonomous' in relation to the members' attitudes in two senses. The individual attitudes on the conclusion are both insufficient for determining the group attitudes on it and unnecessary.[48]

On the supervenience interpretation, it is again merely postulated that there is something supervenient, whose existence could then be confirmed by means of a supervenience relation, in that a non-breaking-off of the actually impossible aggregation leads to the autonomy of a group agent. This is simply a self-fulfilling demand: there *is* a group agent, which can then distinguish itself by a kind of autonomy as a new property—rather than simply declaring the group formation to have *logically failed*.

The supervenience reinterpretation adds nothing new to JAT's notorious difficulty of aggregating an unambiguous collective output from the individuals' input. It is only now interpreted somewhat differently (given a positive turn): the result allegedly supports the widespread *intuition* that a group intention that has become independent—or an autonomous group agent—supervenes over and above the individuals. A relation between individual agents and the group agent would have been interesting only if a nomological—that is, "causal"—necessity of the latter were to result.

Moreover: is the interpretation of the impossibilities in JAT by means of supervenience even the only possible one? No: in a mathematical model such as judgment aggregation here, the transition from input to output is always a *mapping*. If it is carried out correctly and yields no unambiguous (consistent) result, then one may consider weakening or fine-tuning the *constraints* or *conditions* in such a way that a normative aggregation could take place which would also find majority assent. This "fine-tuning" or "relaxing" of the desiderata is already taking place in current research, though not yet to the necessary and desirable formal and experimental extent, nor with the corresponding applications. It thus turns out that supervenience is not the only interpretation of the notorious impossibility results in JAT.

The supervenience interpretation therefore does not support the existence of *group agency* or *group mind*. But List and Pettit seem to know this themselves: the negative arguments from JAT, and their reinterpretation in the light of supervenience, provide no ground for the existence of *group agency*. The negative argument has (as with the first *positive claim*) a "merely" epistemological orientation:

> The arguments for our claim about the autonomy of group agency are *epistemological* [emphasis I.S.], bearing on the difficulty of deriving a group agent's

[48] Ibid., 71.

> attitudes from the attitudes of its members. But this should not be surprising, since it is similarly epistemological arguments that support the stance of many psychologists, as well as the commonsense view, that individual agents cannot be understood merely as complexes of neuronal organization. Under standard approaches to the individual subject, mind and agency supervene on what happens at the neuronal level. But the difficulty of predicting from a neuronal base what an agent does provides a justification for making sense of the agent in terms that abstract from the way its neuronal parts operate. Analogously, the difficulty of predicting from an individualistic base what a group agent does provides a justification for making sense of the group agent in terms that abstract from the way its members perform. It is true that in the case of a group agent we make sense of parts and whole in the same intentional vocabulary, whereas we do not use that vocabulary with reference to neurons. But this disanalogy does not undermine the fundamental similarity between the two cases.[49]

Indeed, it is presumably precisely this assumption that will be the reason for the far-reaching and also dangerous misunderstanding (see Section 3.3, The Normative Argument): "But this disanalogy does not undermine the fundamental similarity between the two cases." It is firmly assumed that the analogy holds: if the mind ("members' mind") exists, then the group mind ("group mind of its own") exists as well.[50] Apart from the fact that no proof from analogy was given—that is, the analogy was not used in the supervenience account and constitutes merely a claim of plausibility without any explanation—there are, in the case of the so-called body–mind problem (without opening that debate here), indications of the additional existence of the mental, with the problems bound up with it. This does indeed make an account of the dependence relations necessary. If such an analogy is not assumed—if, that is, it is not assumed that the supervenient and the supervenience base exist—then the hitherto unsolved problem does not arise: there is no *group agency*, that is, no *group beliefs*, *group desires*, and *group actions*.[51]

A third, normative argument (*normative claim*), finally, concerns the normative necessities of assuming the existence of *group agency*: lifeworld practice allegedly shows that, as soon as people assume group agents, they begin to interact with this entity detached from individuals (the group agent)—for instance, by criticizing it or placing certain demands on it—in a way that would not be possible if the group, as a group, had no real, independent *agency*. One refers to a group agent even without attending to what happens at the level of the individuals of that group. When one criticizes a group agent, for

[49] Ibid., 78.
[50] Ibid., 8.
[51] Ibid., 20.

example, one does so *separately*, even entirely without criticizing the individuals involved in that group agent. Such, allegedly, is the established normative *practice*, or lifeworld.[52]

3.3 The Normative Argument (*collective responsibility*)

In their third attempt, then, List and Pettit try to show, by means of *collective responsibility*, that realism about group agents is also *normatively* necessary. The existence of group agents must be presupposed, they argue, because otherwise the ascription of responsibility or guilt would not be possible in certain cases. This is often the case in situations where there is an offense but the individuals responsible cannot be identified, because the group, or the deed, is too large.[53] An example: for violations of privacy, above all in connection with the Cambridge Analytica scandal, the US consumer-protection authority imposed a penalty of around five billion dollars on Facebook.[54] The ruling does not proceed against individual persons, but against Facebook as the group agent responsible here. The point List and Pettit make is that it is necessary to assume the existence of this group agent here, because without it Facebook could not be found responsible and condemned.[55] It is thus condemned as a "bad social arrangement"[56] and, by means of a penalty, compelled to change its arrangement in accordance with data-protection laws.

This argument, too, seems to be of a purely pragmatic or instrumental nature: while it reflects our legal and normative practice, it does not constitute a sufficient reason to assume Facebook as an existing entity. It was all the employees who had synchronized in this way at Facebook—or some knew nothing of the machinations of particular individuals in this corporation, which was presumably the case. Owing to the high complexity arising from the large number of individuals involved at Facebook, it may, under the circumstances, be practically impossible to identify the individuals responsible. One has therefore agreed to sanction all individuals, condemning them as a *bad social arrangement*. Whom the sanctions actually reach—for whom, specifically, they have negative consequences (and that is the only thing that matters)—is not specified and hence not controllable either. It may strike the unknowing and uninvolved employee who has to be laid off because of a subsequent drop in profits, or the unknowing and uninvolved stakeholder who loses his investment.

So if it is practically impossible (because of the complexity) to identify the individual

[52] Ibid., 6.

[53] After the Second World War, the so-called "collective guilt" of the Germans for the Holocaust and other crimes against humanity during the Nazi regime was discussed controversially worldwide. One of the most convincing critiques of "collective guilt" was developed above all by Hannah Arendt (Arendt 2014).

[54] Cf. Welt, 12 July 2019. Cambridge Analytica collected sensitive data from around 87 million Facebook users—without their consent, with the aim, among others, of manipulating their votes in the 2016 US elections in the interest of the "clients."

[55] Ibid., Part III, 153 ff.

[56] Ibid., 106.

wrongdoers in the group, and we are under *normative* compulsion to ascribe responsibility and carry out the punishment, this does not yet mean that the particular individuals were reached who did or failed to do this or that unlawfully within the group Facebook *and* who are now (supposed) to feel the negative consequences. List and Pettit do not dispute this either; they say only that *not only* the individuals involved—in our example, at Facebook—are to be held responsible (even if they could be identified), but also that the badly designed system should be criticized, condemned, and sanctioned. If one did not condemn the *bad social arrangement* Facebook as a group agent, then the individuals involved could not be condemned either:

> To refuse to ascribe responsibility to the group as a whole, on the grounds that the evil done was done entirely by the spokesbody, would be to miss the opportunity to put in place an incentive for members of the group to challenge what the spokesbody does, transforming the organizational structure under which they operate: making it into a structure under which similar misdeeds are less likely.[57]

But if both particular individuals of the group can be held responsible for their deed *and*, in addition, the group itself, does this not thereby give rise to a kind of *over-responsibility*? The two philosophers deny this. They distinguish two different levels of responsibility. On the one level, particular individuals are responsible for the deed of the group "Facebook"—for instance, when particular persons in key positions decide to sell the non-anonymized user data to *Cambridge Analytica*. The other level is when all the other members (the employed members not involved in the decision) become accomplices or fellow travelers of this *bad social arrangement*. Here, too, the symbolic practice of disciplining takes hold—one that is, however, ultimately directed at the *individuals* after all:

> By finding the group responsible, we make clear to members that unless they develop routines for keeping their government or episcopacy in check, they will share in member responsibility for what is done by the group and may also have a negative form of enactor responsibility for allowing it to be done. We may also make clear to the members of other similar groupings that they too are liable to be found guilty in parallel cases, should their collective body bring about one or another ill.[58]

To stay with our example: the individuals employed at Facebook are made co-responsible, or punished—for example, with wage cuts or dismissals (so as to offset the multimillion-dollar penalty, say)—*merely* because they are, or were, employees of Facebook. There is

[57] Ibid., 169.
[58] Ibid., 169.

no instrument by which the members of a group could each be held responsible according to the degree of their involvement. Do such group condemnations really work well, then, and are they just? Do they not prevent an individual examination of those presumably responsible (e.g., Mark Zuckerberg)? Does one not, rather, punish the worst-off in the group? What would happen if one punished not Facebook, but those responsible in this monopoly? Would not the group behavior of the individuals at Facebook then improve—and a *bad social arrangement* thereby become a better one—if those actually responsible were condemned and sanctioned? Is it not simply a scientific—and not merely a legal—task to investigate and describe these phenomena (e.g., the corruptibility and greed of the powerful in a powerful group) empirically, by closely examining the individual behavior of these powerful persons and their fellow travelers, or even anthropological regularities concerning people in positions of power or of fellow-travelership?

Empirically speaking, in the end it is always individuals who are punished—whether purely arbitrarily or in a targeted way, whether as a mere member of the group or, additionally, as a concrete (decision-making) perpetrator. Would it not be better to prescribe, in a targeted way, transparent organizational structures, so as to install a so-called *take-charge responsibility* which, in the case of non-fulfillment, would be sanctionable in a person-specific manner? Instead, the established practice—when a pro-forma group agent is assumed—is the collective and not particularly effective blanket punishment of *all* individuals of that group, where, empirically, it tends to be those who are worst-off there who must pay, and who, as a rule, bear a "co-responsibility" only by virtue of their membership or employment.

It can thus be argued that this practice is unfair and possibly also counterproductive. But with that, the alleged *normative necessity*—which is surely an empirical question—of the group agent presupposed for it, as a real collective person together with its collective responsibility, would fall away. If the normative necessity cannot be established, then, on List and Pettit's own terms, the third argument for the existence of group agents—with their group beliefs, group desires, group intentions, and, finally, group actions—would become void. In any case, however, the normative argument, too, is unsuited to making the existence of *group agency* even so much as plausible. *De facto*, a group is never sanctioned; rather, it is always the individuals of a group—just as a group never interacts or intervenes, but always the individuals of a group themselves, then possibly with the assent of all—or they leave the group.

4. Conclusion

In the first part, it was shown that Tomasello's interdependence hypothesis—that is, an evolutionary description of the intentionally synchronized behavior of humans as ultra-social beings—dispenses with any postulation of the collective entities group agent and

collective person. Does this mean that collective and normative phenomena (group properties, or group behavior) can be reduced, or traced back, to individual behavior, or to the properties of the individuals themselves? Tomasello shows: it is possible in principle. In his work, collective concepts are introduced empirically via individual properties and unfolded both phylogenetically and ontogenetically. Even if he may not always comply with the compulsion to set quotation marks, it always becomes clear that he uses the collective expressions metaphorically, or pragmatically. They could, in principle, be replaced by precise (if possibly very complex) aggregative definitions, and the *real* aggregations could be investigated empirically and concretely accounted for—even if what is then at issue is not a pure, logical form in the sense of, say, judgment aggregation theory. What is decisive is the fact that an explicit or implicit agreement of the individuals came about—that they declared themselves to be in accord with the output as to which property each individual of the group is to have.

In the second part, the three main arguments of List and Pettit proved too weak to establish the existence of intentional group agents or collective persons. Descriptive and explanatory parsimony, the supervenience of existences already postulated by way of the analogy of the body–mind problem, and, finally, normative habit do not suffice to render plausible the assumption of the existence of group agents.

In order to be able to describe and sanction the behavior of synchronized individuals, a postulation of the existence of *group agency* is not necessary—and possibly not even desirable. Further research is needed in order to be able to set up a more fundamental theory here, one that describes the dynamics of the synchronization of individuals. Human individuals synchronize toward a certain new property located in the individuals themselves—that of being a member of this or that group. In doing so, at least one property or intention of the individuals involved must be shared unanimously by all the individuals involved (implicitly or explicitly), so that they can synchronize further and thereby bring forth different cultures or forms of life, together with their respective cultural and moral norms.

Finally, examples can be adduced of an unfair and unjust practice of sanctioning, one grounded in the widespread intuition that groups exist and that group responsibility and collective guilt are therefore possible. This practice prevents, among other things, a precisely targeted intervention or prosecution. On closer consideration and investigation, however, it might turn out—in a normatively attuned (that is, aggregative) way—that in the future we will not consider this unjust practice desirable or worth preserving.

References

Arendt, Hannah 2014, *Eichmann in Jerusalem. Ein Bericht von der Banalität des Bösen*, München.

Axelrod, Robert 1985, *Die Evolution der Kooperation*, Berlin.

Boehm, Christopher 1999, *Hierarchy in the Forest: The Evolution of Egalitarian Behavior*, Cambridge.

De Waal, Frans 1996, *Good Natured: The Origin of Right and Wrong in Humans and Other Animals*, Cambridge.

Dietrich, Franz 2007, "A Generalized Model of Judgment Aggregation," in: *Social Choice and Welfare* 28, 529–565.

Grossi, Davide; Pigozzi, Gabriella 2014, "Judgment Aggregation: A Primer," in: *Synthesis Lectures on Artificial Intelligence and Machine Learning* 8(2), 1–151.

List, Christian 2011, "The Logical Space of Democracy," in: *Philosophy & Public Affairs* 39(3), 262–297.

List, Christian 2012, "The Theory of Judgment Aggregation: An Introductory Review," in: *Synthese* 187, 179–207.

List, Christian; Pettit, Philip 2011, *Group Agency*, Oxford.

McLaughlin, Brian; Bennett, Karen 2018, "Supervenience," in: *Stanford Encyclopedia of Philosophy*, Stanford. `https://plato.stanford.edu/entries/supervenience/`.

Popper, Karl R. 1945, *The Open Society and Its Enemies*, London.

Sigmund, Karl 1993, *Games of Life*, Oxford.

Smith, John Maynard 1982, *Evolution and the Theory of Games*, Cambridge.

Tomasello, Michael 2016, *A Natural History of Human Morality*, Cambridge.

Tomasello, Michael 2018, "Précis of A Natural History of Human Morality," in: *Philosophical Psychology* 31(5), 661–668.

Tomasello, Michael; Vaish, Amrisha 2013, "Origins of Human Cooperation and Morality," in: *Annual Review of Psychology* 64, 231–255.