

We-mode as Layered Agency

Lukas Schwengerer

Penultimate Draft, please cite the published version:

<https://onlinelibrary.wiley.com/doi/10.1111/josp.12600>

Abstract

In this paper, I explore a new approach to we-mode agency drawing on the concept of layered agency. I argue that agents can shut out their personal attitudes in favour of a perspective jointly established with other people. I can act as a member of the philosophy department aiming for what the department agreed on, even if that might conflict with my personal beliefs. I can shut out these personal beliefs for a moment and reason from the group's standpoint. While this observation itself is not novel, I provide an original account for this phenomenon building on C. Thi Nguyen's proposal of layered agency. I then use this new account of the we-mode as a basic building block for conceptualising social groups and group attitudes. I end by showing how the proposal deals with common cases from the literature.

1. The Big Picture

We humans are social beings.¹ As such, our agency can take on goals and means that are not only determined by an agent individually but by cooperating and coordinating with other people. Agents that are part of groups can act with the group in mind. They can act as group members taking on the group's motivational structure. I can act as a member of the philosophy department aiming for what the department agreed on, even if that might conflict with my personal beliefs. I can shut out these personal beliefs for a moment and reason from the group's standpoint. And I can also switch from my reasoning as a group member back to my individual point of view when the situation requires it. I suggest that this ability to switch between agencies is best described as a form of layered agency – a *we-mode layer of agency*.

The idea of layered agency has been proposed by C. Thi Nguyen (2019; 2020) in his seminal

¹ An earlier version of this paper was presented in Raphael van Riel's research seminar at the University of Duisburg-Essen and at the 'Group Minds and Collective Agency' workshop at the University of Leeds. Thank you to all attendees for their feedback. Further thanks to the reviewers of this journal.

work on striving play – play we partake in with the aim of experiencing a struggle. I argue that the idea can also be explanatory powerful in the social realm more generally. How we reason in many cooperative activities is best described as taking on a new layer of agency that is jointly established with other people. The ability of agents to submerge themselves in this jointly established layer of agency is the key to understanding actions in a group mindset, the nature of social groups, and the mental states of group agents.

Much of our social life is shaped by this ability to take on a we-mode as a temporary layer of agency. I can smoothly transition from practical reasoning based on my own individual goals to reasoning based on goals established jointly with other people, such as the goals of the philosophy department. I can also transition between different we-modes as different temporary layers of agency. At one point I reason as part of the philosophy department, at another point, I reason as part of a football team. The goals of these groups are different, and some accepted cornerstones that guide my reasoning will also be different. These cornerstones are goals and attitudes that provide fixed points for practical reasoning as a department member or a member of the football team. As a philosophy department, we might have jointly accepted that Katherine is a good job candidate and, in my reasoning as part of the department, this acceptance is held fixed to guide practical reasoning. But as part of the football team, nothing requires me to hold that acceptance fixed to guide my reasoning. The football team is not interested in good job candidates for a philosophy department. There has not been any agreement on job applicants at all for the football team. My practical reasoning has different cornerstones depending on what layer of agency I take on board: the philosophy department layer or the football team layer. I can switch between these different kinds of agencies with their individual cornerstones depending on what the situation calls for.

I suggest that the we-mode layer of agency is the basic building block of social groups and collective mental states. The we-mode layer itself is only dependent on joint commitments and my abilities of submersion and layering of agency. It does not depend on the prior existence of groups or collective mental states. Moreover, the we-mode layer does not entail any form of plural subject as is accepted in other accounts using the notion of joint commitment, such as Gilbert's (1989; 1990; 2009) proposal. A social group is then narrowly defined in terms of agents sharing a we-mode agency with the same cornerstones by joint commitment. A collective mental state is defined in terms of a sufficient ratio of the same we-mode attitudes in relevant group members (where the relevancy is determined by the cornerstones of the shared we-mode agency).

I end these introductory remarks with a note of caution: not every detail nor potential problem for the proposed account will be discussed in detail. My aim is to sketch a big picture view of how to think about the we-mode in a new way. It is an exploration of using the idea of layered agency in the social realm and more work will be needed to fully flesh out the proposal.

2. We-mode as a Layer of Agency

Agents can adopt a we-mode as individuals. Unlike other proposals of a we-mode (e.g. Tuomela (2005; 2013), Schmitz (2017; 2018; 2023), Schmid (2017)) I understand the we-mode as a temporary layer of agency that allows agents to take on disposable goals and means. These goals and means are based on agential cornerstones jointly committed to by multiple agents under conditions of common knowledge. This is the picture I want to spell out in more detail.

Let me start with the idea of layered agency put forward by Nguyen (2019; 2020) to analyse the agency one takes on when playing a game.² As a player of a game, I take on the goals and means given to me by the rules of the game (and indirectly the game's designers).

When I play a game of Super Mario Kart, I adopt a temporary agency that wishes to win the race. I make that temporary agency dominant. I immerse myself within it, but within certain limitations. [...] But so long as the background monitoring processes of our full agency haven't broken through with such a cancellation, we let the temporary agency regulate our decisions and dominate our consciousness. (Nguyen, 2019, p. 443)

For striving play I look for a particular experience of struggle. It is not the winning itself that is my ultimate goal in playing *Super Mario Kart*, but the struggle experienced while trying to win. This is the core of striving play in contrast to achievement play, which would be playing for the sake of winning. What makes striving play so interesting is that I cannot aim for the struggle throughout my practical reasoning while I am playing the game. If struggling was my guiding aim while playing, I would intentionally play worse. Nguyen illustrates this with a story: a ten-year-old enjoyed his experience of beating his father in *Monopoly* so much that the boy ended up offering free cash to his father to prolong the game. The kid missed something essential about playing a game (Nguyen, 2019, p. 437). A player with the kid's reasoning can never wholeheartedly play the game, trying to win. In a sense, it becomes dubious whether a mature adult using this kind of reasoning would be playing at all. The constitutive goal of the game would not be adhered to. To wholeheartedly take part in striving play the goal set by the game has to be taken on in a way that allows it to temporarily dominate the player's agency. The

² Nguyen (2019; 2020) himself also talks about layers of agency as 'modes' already.

background goal of experiencing struggle has to be shut out from practical reasoning for a while. I have to fully submerge myself into the goals and means provided by the game.

This is a remarkable ability. Not only can I take on temporary goals and means, but I can do so in a way that changes my practical reasoning and my actions fundamentally. In doing so I create a new layer of agency.³ This is a motivational two-step process. I have the overall goal of experiencing striving play and in order to do that I need to take on the temporary goals and means in a way that keeps the overall goal itself temporarily out of my practical reasoning. This ability does show itself in other areas as well. Nguyen (2020, p. 55) gives the example of relaxing and clearing one's head from work stress. I cannot clear my head and relax by willing myself to do that. Instead, I need to submerge myself in some other task that gets my mind away from work. Perhaps going for a walk, cooking or reading. And I need to do this in a way that temporarily shuts out the overall aim from my current reasoning and action. I have to wholeheartedly focus on the activity. Just like in the case of games, I manage to take on a temporary layer of agency with its goals and means and I let this agency become dominant.

I suggest that adopting a we-mode is another instance of this kind of layered agency. What makes the we-mode a particularly interesting instance of layered agency is that the goals and acceptances that constitute the cornerstones for my reasoning in the newly adopted layer of agency are not determined by myself alone (as in the relaxation case), nor are they determined solely from the outside (as typically in the games case). The cornerstones are determined jointly with other people. They are established via joint commitments under common knowledge. I temporarily seal myself off from my ordinary interests and ends in order

³ Camp (2023) argues that accepting such an ability to create layers of agency is not going far enough and that Nguyen's proposal leaves room for opponents to argue that these agential layers are mere pretense. Camp herself provides arguments against goals in the temporary agency being merely pretense. I will bracket this discussion.

to make the new joint commitments dominant in both my reasoning and motivational structure. I do that in order to achieve some overall, individual aim which I can best (and sometimes only) achieve if I take these new, jointly established goals and means on board. I need to fully submerge myself in that new layer of agency with its unique structure of practical reasoning in order to follow my individual goal successfully. Just like in the case of playing a game, it is a motivational two-step process. I start with a usual, individual I-mode goal that motivates the creation of a new layer of agency. In the we-mode case it is an I-mode goal that I cannot (or not easily) fulfil by myself. This I-mode goal leads me then to joint commitments together with other people that establish a particular we-mode. I take on this we-mode as a new layer of agency and make it dominant. I reason and act in the new layer of agency based primarily on the goals and means jointly committed to. My common I-mode goals and reasons still exist in the background, but they are temporarily shut out and only show up again when the newly adopted layer of agency becomes questionable – either in its functionality or with some unpalatable consequences, or perhaps simply because something more important needs to be taken care of. When I take on the we-mode layer of agency my action is temporarily guided only by the new goals and means jointly committed to, plus my individual representations of the world (as long as they are not incompatible with the joint commitments).

Take Gilbert's (1990) example of two people going for a walk together. I cannot achieve the goal of going for a walk together with someone else just by myself. Moreover, I need the other person to cooperate in a way that is more than just both of us wanting to go for a walk at the same time at the same place. It requires particular commitments to each other that set up tacit rules and expectations. For instance, going for a walk together establishes normative demands that no person just randomly runs off without notice or explanation. In order to go

for a walk together we jointly commit to do so. Going beyond Gilbert, I take this to be well described by each of us adopting a temporary layer of agency of us walking together as a unit. And in contrast to Gilbert's proposal, my model does not entail any sort of plural subject that is formed in virtue of those joint commitments. By jointly committing to go for a walk together we each seal off our individual agencies and set some cornerstones for each of our new we-mode agencies, but these new we-mode agencies are still located within each of us individually. There are individual subjects who act and reason in a we-mode agency based on the set cornerstones, but there is no plural subject in anything like Gilbert's sense. The normative demands of the social practice of going for a walk together are such cornerstones, which each of us individually now takes on board as guides for practical reasoning. For instance, each of us accepts not to leave without voicing the plan to leave and the reasons for doing so. And each of us accepts not to move too fast or too slow compared to the other person. These are examples of cornerstones that now shape the individual we-mode agency for each of us. My actions are determined by my practical reasoning based on these cornerstones and my beliefs about the world (that are compatible with the cornerstones jointly committed to).

This idea of one's reasoning and one's actions being partially determined by cornerstones and partially by personal beliefs also fits well with an observation made by Carol Rovane (2002): Joint endeavours do not require that all involved parties achieve *overall* rational unity, but rather rational unity to a sufficient degree. With 'rational unity' Rovane refers to a rational equilibrium of all-things-considered judgments by an agent. An agent has overall rational unity when their all-things-considered judgments have the least contradictions and conflicts possible. It is a maximally rational, maximally coherent set of such judgments. This includes judgments about the world and evaluations of actions available to the agent. Part of rational

unity is to choose those actions that fit best with the agent's psychological attitudes and to judge those actions as appropriate for the agent. Rovane takes overall rational unity to be a normative demand on individual agents (Rovane, 2002, p. 215). Agents ought to show overall rational unity – they ought to resolve contradictions and conflicts within their judgments.

For group agency and the we-mode, rational unity comes into play as a notion that also applies to people acting and reasoning together. In order to reason and act together the involved parties need to show some rational unity. Group members cannot hold wildly contradicting beliefs and desires if they want to act and reason together. However, Rovane observes that rational unity to a degree can be sufficient for joint deliberation and action – group members do not need to be perfectly aligned in all judgments. Let me show this with her example of two undergraduate students working together to answer a philosophical question (Rovane, 2002, pp. 218-219). In order to successfully do that, they have to build on what each one of them says, work out consequences, and so on. This can only be done if they behave as if they have only one rational point of view with respect to their task. They need to work with the same set of base assumptions and the same goal guiding their joint deliberation. However, this does not require that both students need to share *all* goals, desires and beliefs, nor that the students cannot have any contradicting beliefs or desires. They only need to share things that pertain to their joint endeavour. For instance, both need the same goal of answering the philosophical question and they need to agree on a methodology (among other things). But it does not matter if one wants to do so because they are genuinely curious and the other one has a completely different motivation (Rovane, 2002, p. 219).

Similarly, two people walking together might want to do so for different reasons, but they need to align to a sufficient degree with respect to the desire to walk together and in (at least some of) their beliefs what walking together entails. Only then they can form a rational point

of view together. This is exactly what happens when they commit to particular cornerstones for the we-mode. Those cornerstones capture the jointly accepted point of view and establish a sufficient degree of rational unity between the involved parties. There are many beliefs and desires that each individual party has besides those cornerstones, but for a we-mode a degree of rational unity is necessary and that rational unity is established insofar as all parties take on the cornerstones as fixed points for reasoning and acting. Those fixed points then determine which other beliefs held by individuals can also be used to aid judgments in the we-mode.

Different joint endeavours require different degrees of unity between the involved parties and hence more or less cornerstones that shape the we-mode. Hence, one might say – as Rovane (2002, p. 221) does – that the joint endeavour imposes normative requirements on the people involved. Some complex tasks require all parties to take on many cornerstones, others only require a few cornerstones to be done successfully. Some tasks and cornerstones are rather general, others specific. We-mode agency is shaped by the task at hand that prompts people to join forces.

The proposal of a we-mode as a form of layered agency with cornerstones can shed light on how easily such we-modes can be established, how they relate to I-mode agency and how they fit into the broader picture of our ability to shift between agencies more generally. Layered agency shares some characteristics with Tuomela's (1992; 2013) proposal of a we-mode with positional beliefs (or positional views) – the beliefs (views) that a position-holder has qua a position-holder. My layer of agency proposal and Tuomela's idea of a we-mode approach share that an individual can reason and act from different points of view: their own personal point of view, and a point of view jointly established with other people. In contrast to Tuomela, I am reluctant to put this jointly established point of view in terms of group membership. I want to use the jointly established layer of agency to explain social groups,

rather than explaining a we-mode with reference to groups. Tuomela understands the we-mode as including the agents intentionally functioning together as a group for the same authoritative group reason, satisfying the criterion of collective commitment and a collectivity condition (Tuomela, 2013, pp. 23-24). This collectivity condition requires that the group members stand or fall together in we-mode activities and attitudes. For Tuomela, a group member cannot satisfy a we-mode goal or attitude if that goal or attitude is not satisfied for all other group members as well. He takes that to be a constitutive principle of the we-mode (Tuomela, 2013, p. 29).

My layer of agency model for we-modes does not accept a general collectivity condition. An agent can have we-mode goals and attitudes that can be satisfied only for the agent, as long as they are not cornerstones for the layer of agency. The primary reason for this is that individual agents operating in a we-mode layer of agency can form new goals and attitudes based on the jointly established cornerstones. These newly formed goals and attitudes are still in the we-mode, as they operate fully within the temporary layer of agency. But they need not be shared between everyone who jointly committed to the cornerstones and took on a we-mode layer of agency. In such a case one agent A might have a we-mode goal that is not present in another agent B, and therefore A can satisfy that goal without any such goal satisfaction taking place for B. I might jointly commit with my football teammates on overall aims and strategies for the football game. When on the pitch, I act and reason within the we-mode layer of agency established by such joint commitments. Nevertheless, I can form intentions and goals individually in that we-mode while playing football. I can form a we-mode goal to pass the ball and I can satisfy that goal by acting accordingly. But there might not be a corresponding we-mode goal present in my teammates. They do not have the goal of me passing the ball – neither as an I-mode goal, nor as a we-mode goal. My teammates' we-mode

layer of agency has the same cornerstones as mine, but lacks the relevant non-cornerstone goal. There is nothing to be satisfied for them. However, it is still a goal of mine in the we-mode, because it was formed in the we-mode layer of agency. That is a difference between Tuomela's understanding of a we-mode and my layer of agency proposal. Tuomela takes we-mode goals to only be those that are established jointly, either directly by joint commitment or indirectly as a result of joint commitments on positions within a group. In contrast, I take we-mode goals to be *any* goals present in a layer of agency with cornerstones that were established in a joint commitment.⁴

Even though I reject the strong collectivity condition, I accept a weaker version. A weakened version holds that a we-mode goal can only be satisfied if the goal is also satisfied for all other group members *who have the same we-mode goal*. The weaker version respects that we-mode goals can differ between individual members and come into conflict. Forming goals and intentions from the group's point of view is no guarantee that no such conflicts show up. They regularly do. Many intradepartmental conflicts stem from parties forming incompatible goals while nevertheless having the group's best interest in mind. These conflicts constitute practical challenges that prompt some form of conflict management, bargaining and joint deliberation. A strong collectivity condition as a necessary condition for we-mode attitudes cannot capture these non-ideal intragroup conflicts. A weaker version recognises these conflicts, but nevertheless captures that we-mode goals that have exactly the same conditions of satisfaction will all be fulfilled at the same time. Two we-mode goals with the same

⁴ Some we-mode goals in my account would be qualified as we-mode proposals rather than we-mode goals by Tuomela. We-mode proposals are formed by individual group members from the group's point of view, but are not – or not yet – jointly committed to by the group. Hence, they are not group we-mode goals or attitudes but can become those. Cf. Tuomela (2013, p. 105).

conditions of satisfaction cannot differ in their satisfaction. If two agents have the same we-mode goal, it is impossible for only one of them to achieve their we-mode goal.

The model of layered agency has clear advantages over Tuomela's model of a we-mode with positional beliefs. The layered agency model is independently applicable to behaviour that has nothing to do with groups or social positions. Remember Nguyen's (2020) example of clearing one's head by taking a walk and trying to relax, and the central role of layered agency for the ability for striving play. It is a distinct advantage of the layered agency proposal that the same capacity can explain Nguyen's examples and also the we-mode. Positional beliefs (or views) on the other hand cannot do that. Even with the most charitable interpretation, they do not fit the example of clearing one's head. Moreover, layered agency provides a good explanatory story for how, when and why one switches between different agencies. I can submerge myself in a temporary layer of agency while in the background my I-mode agency is still present – sensitive to situational factors that demand me to abandon the temporary layer of agency if I am confronted with something of higher priority. The temporary layer of agency is taken on in order to fulfil a goal that could not (or not easily) be achieved otherwise. The whole temporary layer of agency is justified in the broader agency of the agent. When that is no longer the case, or when the costs of staying within the temporary agency become too high, I abandon it. Moreover, I can regularly check whether that is the case. There will be instances at which I ask myself whether I still want to act with the jointly committed goals and means in mind. This is also something that has parallels outside of the we-mode or any form of cooperation. When I play a game, a situation might come up in which I step back and ask myself whether I want to keep playing – whether it is still fun. Or when I am trying to clear my head by taking a walk I might stop to ask whether the walk is actually working and helps me clear my head. In all these situations I step back from the layer of agency I took on temporarily,

ask whether reasoning and acting within that layer is still promising for my overarching goals, and then decide whether to abandon that layer of agency or submerge myself back into it.⁵ I monitor – either consciously or unconsciously – whether I should stay in a particular layer of agency. This is exactly what happens when I act under joint commitments with other people as well. There will always be a normative demand to keep acting according to the joint commitment, but that demand can be overridden in favor of leaving the we-mode layer of agency. Again, layered agency captures more than positional beliefs (or views) can. The we-mode is just one particular form of layered agency – one based on joint commitments.

3. Why Joint Commitments?

I suggest that the we-mode is a form of layered agency with cornerstones determined by joint commitments. However, one may wonder whether it is necessary to rely on the joint commitment idea to get my proposal off the ground. One might think that something less demanding, perhaps akin to Bratman's (1993) proposal for shared intentions can be sufficient. His account is stated as follows:

We intend to J if and only if

1. (a) I intend that we J and (b) you intend that we J
2. I intend that we J in accordance with and because of 1a, 1b, and meshing subplans of 1a and 1b; you intend that we J in accordance with and because of 1a, 1b, and meshing subplans 1a and 1b.

⁵ In a similar way one might get into situations in which one is drawn to both layers and keeps jumping between them. For instance, I might have jointly committed to ϕ , but I have moral concerns related to actually ϕ -ing when it comes to it. I might feel drawn to act according to the commitment, but also drawn to act against it for moral reasons. And it can be difficult to decide on how to act in such a situation.

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

Shared intentions are here understood as interlocking intentions, meshing subplans and common knowledge. In this framework shared intentions are explained on an individualistic basis. There is nothing fundamentally collective in the same way that joint commitments are. Can I use a similar, purely individualist basis for my proposal? Perhaps interlocking mental states akin to Bratman's account determine the cornerstones for the we-mode layer of agency. If so, then the proposal functions just the same, except it loses the strong collective nature established with the reference to joint commitments. One might argue that such an account becomes more parsimonious and ought to be preferred. It might even be a better fit with the primarily individualist psychological explanation used to describe an agent's action in the we-mode.

Unfortunately, such an account will not do the trick. Joint commitments come with personal cornerstones in the we-mode layer of agency for all parties. But at the same time, joint commitments are more than that. They perform double duty as personal cornerstones in we-reasoning and as a normative constraint on the parties that jointly commit to ϕ . By jointly committing to ϕ the parties take on a we-mode layer of agency and are committed to reason and act within that layer of agency. If someone does not reason or act in a way that is within that layer and fits with its cornerstones, then that person fails to fulfil the joint commitment and is therefore blameworthy. For instance, when we jointly commit to going for a walk together but I end up running away for personal reasons incompatible with the cornerstones, then I fail to fulfil my commitment because I act from I-mode reasoning rather than we-mode reasoning.⁶ This explains the normativity that comes from joint commitments observed by

⁶ Of course, someone can be excused from these normative constraints (e.g. if my life is threatened while on a walk it is perfectly reasonable to just run away), but ordinarily, these constraints on the parties involved hold.

Gilbert (1989; 1990; 2017). A similar account that establishes the we-mode layer in virtue of a Bratman-like individualistic ground has difficulties capturing this normative component. This is at the core of the disagreement between Gilbert and Bratman. I find Gilbert's (2008; 2009) arguments convincing, showing that personal intention as proposed by Bratman will not be enough to generate the right sort of normativity that is involved in joint action. Merely pointing to a demand for stability of intentions, as done in Bratman (1993), is not enough. The distinct normativity present in groups and joint action is best explained by a joint commitment. However, it is worth emphasising here again that endorsing joint commitments and giving them an important role in the model does not entail any sort of plural subject as is usually associated with Gilbert's concept of joint commitments (1989; 1990; 2017).

4. Cornerstones and Affordances

So far, I suggested that the cornerstones jointly committed to include goals and acceptances. However, their transformation of one's agency can go beyond those. Some goals and acceptances might change one's *affordances*. Affordances here refers to action possibilities that are offered to the agent by the environment.⁷ This possibility for layers of agency coming with new affordances is known from the case of games. When I start to play the board game *Clank!* a field on the board that meant nothing before now suddenly becomes an action possibility within the game. It becomes a spot to move to and take a game action (e.g. get a coin). This was not possible without taking on the new agency provided by the rules of the game. Accepting the rules of the game changes one's abilities such that the affordances change. Something similar can occur when I adopt a we-mode. For instance, there could be a joint commitment between Kathleen, Graham, and me that assigns me a role which allows me

⁷ The term was coined by Gibson. See for instance Gibson (1966).

to form joint commitments for the three of us by myself. Perhaps I pick the restaurant for the group's dinner by myself. This was not possible until all parties jointly committed to giving me the power to pick the restaurant. My action possibilities changed because of the joint commitments and the new layer of agency. There is likely more to say here, but for now, I merely highlight this possibility of making further joint commitments by myself for other people as an interesting new action possibility that can arise through joint commitments. Gilbert (2017) talks about these cases as involving 'non-basic' joint commitments. Every ability to create non-basic joint commitment is based on a basic joint commitment in some form. I understand how basic joint commitments come about in line with Gilbert (1989; 2017) as follows:

[E]ach of two or more people openly expresses his readiness jointly to commit them all in a certain way, and their having made these expressions is common knowledge between the parties. By this I mean, roughly, that the expressions are entirely out in the open between them, and each knows this. (Gilbert, 2017, p. 131)

There are, however, important differences between my proposal and Gilbert's. For Gilbert, a joint commitment to ϕ establishes that all parties are committed 'as one' to ϕ . They form a plural subject. This is not the case in my proposal. I suggest that a layer of agency is established by joint commitments and that layer of agency explains what happens in individuals after joint commitments take place.

Agents take on a we-mode layer of agency individually based on joint commitments. They do not form a single body in the strong sense suggested by Gilbert. However, they are acting 'as one' insofar as they all now have the same cornerstones for practical reasoning in the we-mode. They all act roughly with the same layer of agency. They are only roughly the same layer

of agency because the cornerstones, that have been jointly committed to, leave gaps. Every individual can still have their own representations of the world (and their own subgoals) as long as these representations (and subgoals) do not conflict with any cornerstone jointly committed to. Two people might jointly commit to going for a walk together, but they can still differ in their knowledge of the outside temperature and their beliefs on whether to wear a jacket. The joint commitment is quiet on these issues and the parties can make individual judgments in the we-mode that differ with regard to what is best to wear for both of them walking together. On the other hand, for issues that are covered or practically entailed by the joint commitments such a difference is not possible. If they jointly committed to walking on Princes Street none of the parties can practically reason with the belief (or acceptance) that it would be best to walk on Rose Street instead.

5. An Account of the We-Mode as Layered Agency

I am now in a position to explain how attitudes in the we-mode should be understood in terms of layers of agency, joint commitments and cornerstones for agency. A we-mode attitude is an attitude A possessed by subject S in a layer of agency L for which the following conditions are true:

- 1) The layer of agency L differs from the ordinary individual agency in goals, acceptances and affordances⁸.
- 2) A subject who has taken on the layer of agency L uses the layer's cornerstones (goals, acceptances and affordances) determined by joint commitments (under common

⁸ I list affordances separately for clarity, even though new affordances within the new layer of agency are strictly speaking a result of the temporary goals and acceptances the agent takes on.

knowledge) with other people as part of the basis for attitude formation within the subject's agency.

- 3) The layer of agency L is responsive to future relevant joint commitments (under common knowledge) with the other agents that have jointly established the layer of agency L. Such a future relevant joint commitment may change cornerstones.
- 4) The attitude A has been either formed as a cornerstone for the layer of agency L, or has been formed within L, or is a personal representation of the world that is taken to be required to achieve the we-mode goals (and is compatible with the cornerstones).
- 5) The attitude A is responsive to a relevant change of the cornerstones of the layer of agency L.
- 6) The attitude A is responsive to changes in other attitudes (including relevant changes in the representation of the world) in subject S that are within the layer of agency L and consistent with the cornerstones, but are not themselves cornerstones.

The paradigmatic story for we-mode attitudes goes like this: several agents have a primary goal that each of them cannot reach by themselves. They decide to work together and jointly commit to that primary goal. In order to reach the primary goal, they further jointly commit to a set of acceptances and secondary goals that they agree to as cornerstones for practical reasoning. Each involved party thereby takes on a we-mode layer of agency with these cornerstones as fixed points in their individual reasoning.⁹ The agents can jointly commit to changes to the cornerstones if needed and their individual we-mode reasoning and we-mode attitudes have to adapt based on changes to cornerstones. They might start with a joint commitment to accept p, but find out later that accepting p as a cornerstone is problematic

⁹ As mentioned earlier, such reasoning is also partially based on individual representations of the world, as long as those representations are consistent with the cornerstones.

and hinders them from aiming for the goal they jointly committed to. In this case, they can jointly commit to amend the prior commitment and now accept not-p. All the previous attitudes that were based on the cornerstone acceptance that p now have to change relative to the adjustment in the cornerstone. Moreover, the we-mode reasoning and we-mode attitudes also have to adapt based on changes to the individual's representation of the world. Suppose I am playing football with my team. We have as a football team a bunch of joint commitments on playing football in general and on the strategy in a particular game. My current we-mode cornerstones about our strategy plus my belief that there is no opposing player in front of me might lead my practical reasoning to judge that I should move forward. But once I see an opposing player come up my cornerstones plus my new perceptual belief might lead me to change my intention and pass the ball to a teammate. The we-mode cornerstones all stayed the same, only my individual representation of the world changed, and my attitude had to adapt accordingly. Such changes to attitudes can be done in individuals in the we-mode, as long as the cornerstones themselves stay fixed. Cornerstones can only be changed by joint commitment (basic or non-basic).

We-mode attitudes in this picture are of the same ontological nature as I-mode attitudes. Their only difference is the layer of agency they operate in. Hence, even though I tend to write of we-attitudes (e.g. we-intention), I thereby merely refer to an attitude in a we-mode layer of agency, not an ontological difference of the attitude itself. Because the difference between we-mode attitudes and I-mode attitudes is only the layer of agency the attitudes are in, it seems at first sight that any attitude that can be present in the I-mode layer can also be present in the we-mode layer. This also entails that the we-mode layer of agency may include beliefs. I am unsure whether to take this on board. Whether beliefs can be established by joint commitments is currently under debate. On one side, Gilbert (1989; 1990) proposes by herself

and in collaboration with Pilchman (Gilbert & Pilchman, 2014) that group beliefs can be established by joint commitments. On the other side, Wray (2001) and Hakli (2007) prominently argue that joint commitments cannot generate beliefs, but merely acceptances. Beliefs cannot be generated at will and generating group beliefs by jointly committing to p seems too close for comfort to believing at will. My proposal itself is neutral on this question. The proposed account works both in a version that allows for all attitudes to be present in a layer of agency established by joint commitments and in a version that restricts the attitudes in such a layer to exclude beliefs. In the latter case, acceptances play a role similar to beliefs to form intentions and lead to action in the we-mode layer of agency.

Cornerstones of we-mode layers – goals, acceptances and affordances – can be *positionalised*. A positionalised cornerstone is one that is relative to a social position (or role) and assigns people to positions. Positionalised cornerstones are important for coordinating more complex goals. For instance, the layer of agency might include cornerstone affordances that are only available to a finance officer and a cornerstone acceptance that Bob is the finance officer. In such a case Bob might be able to make financial decisions and form financial acceptances for a group by himself, because the joint commitments include a cornerstone acceptance that gives Bob that power.¹⁰ If the positionalised cornerstone was jointly committed to, then it is a cornerstone for everyone who was part of the joint commitment, even for those that do not have the position. A work colleague of Bob still has acceptances about the positional cornerstone affordances related to finance officers and the cornerstone acceptance that Bob is a finance officer. This ensures that the colleague's we-mode reasoning is unlikely to conflict

¹⁰ Usually, it will also include an accepted responsibility for that anyone in the position of the finance officer to inform the rest of the group of such financial decisions or beliefs

with Bob's, because the colleague has to take the positional cornerstones about Bob into consideration. It is part of the layer of agency that Bob has these powers and responsibilities.¹¹

I have now provided the core of my account of the we-mode as a layer of agency. (1) to (6) describe conditions for proper we-mode attitudes. However, there can be attitudes that are very similar to we-mode attitudes but fail to satisfy one of the conditions. To keep these defective forms distinct from proper we-mode attitudes I call these *we-layer attitudes*. I will mention two interesting cases of we-layer attitudes:

Unresponsive We-Layer Attitudes

I call unresponsive we-layer attitudes any attitudes that are generated and maintained in an agent such that (1) to (4) are satisfied but (5) and/or (6) are not satisfied.

Unresponsive we-layer attitudes come about in the same way as we-mode attitudes. They can be either themselves cornerstones, or non-cornerstone attitudes that are at least partially based on the cornerstones. But unlike proper we-mode attitudes mere we-layer attitudes are unresponsive to relevant changes in other attitudes in the same layer of agency.

A second important form of defect are mistaken we-layer attitudes:

Mistaken We-Layer Attitudes

I call mistaken we-layer attitudes any attitudes that an agent holds in a layer of agency that the agent takes to be established by joint commitments, when in fact no such joint commitment has taken place.

¹¹ There is an open question here: in how much detail do Bob's responsibilities have to be represented in the cornerstone acceptances? I do not have a fully developed answer at this point, although I take it to be dependent on the concrete situation with the relevant joint commitments.

When an agent mistakenly thinks that a joint commitment has been established they take on a temporary layer of agency with cornerstones and act accordingly. However, if there has not been an actual joint commitment, this new temporary agency does not constitute a proper we-mode layer of agency. In any such case, the normative force of acting in the we-mode will also be missing. I might mistakenly take us to have jointly committed to ϕ . When you then do not provide your part in ϕ -ing, I will attempt to rebuke your actions and demand that you do your part. But you will not feel any normative force in that demand. Instead, you would simply respond by pointing out that we did not agree to ϕ – we did not form a joint commitment. In my individual psychology mistaken we-layer attitudes can play the same role as proper we-mode attitudes. However, they lack the features that synchronize we-mode attitudes between different people and the normative force to act in the we-mode established between the parties of a joint commitment.

We-layer attitudes come close to proper we-mode attitudes, but barely miss the mark in some regard. As such, they can never be the basis for group attitudes either. However, even proper we-mode attitudes do not always amount to group attitudes. We-mode attitudes are located within individuals. What makes we-mode attitudes special is that they are established in a way that synchronises the we-mode layers of agency of multiple agents by setting the cornerstones for we-mode reasoning. This allows we-modes to be the basic building block for social groups and group attitudes. With the we-mode as a layer of agency in place I can analyse the notion of a social group:

Social Group

A social group is a set of subjects with the same cornerstone goals, acceptances and affordances constituting synchronised we-mode layers of agencies with the following conditions being met:

- (1) these cornerstones are determined by joint commitment between the subjects under the condition of common knowledge; and
- (2) these cornerstones serve as (part of) the basis for the subjects' individual attitudes in the we-mode layer of agency.

Groups are based on agents adopting layers of agency that are synced by virtue of having the same cornerstones because of joint commitments. The joint commitments ground the newly adopted agency for all group members. My account, therefore, shares a lot of DNA with Gilbert's (1989; 1990; 2017) proposal. Social groups start with individuals jointly committing to something. However, my account differs from Gilbert's in terms of what these joint commitments do: they establish a temporary layer of agency with the same cornerstones in each of the involved parties and they give rise to normative demands to reason within that temporary layer of agency. How the group members act as group members is explained by that new agency, rather than by features of the joint commitments. Group attitudes are then understood as products of members acting and reasoning within their new layer of agency. Hence, group attitudes are supervening on the we-mode attitudes of group members. Here many options, of how exactly these supervenience relation looks like, are available. A simple attempt identifies group attitudes with a sufficient sum of individual we-attitudes of the group members.

Group Attitude

A group has the attitude A if and only if

- 1) a sufficient number of group members relative to the group size with relevant positional affordances have the attitude A in the relevant we-mode; or
- 2) some group members with relevant positional affordances have the attitude A in the relevant we-mode and some group members with relevant positional affordances would have the attitude A in the relevant we-mode, if they were to consider whether to have the attitude A in the relevant we-mode (after full disclosure of evidence within the group), such that the sum of members with relevant positional affordances that have A or would have A is sufficiently large relative to the group size.¹²

Any we-mode attitude that is a cornerstone for the temporary layer of agency is automatically also a group attitude, because it is shared by all relevant group members. However, attitudes that are not cornerstones can still amount to group attitudes. That happens when they are formed in enough relevant individuals reasoning in the we-mode layer of agency. Sometimes individuals legitimately reason and act for the group without the group members deliberating together on that reasoning or action. Many of these cases can be captured with non-basic joint commitments, e.g. when the group has jointly committed to a finance officer with decision-making power who then makes a financial decision by themselves that still counts as establishing a (non-basic) joint commitment. This is part of my proposal of looking at people with ‘relevant positional affordances,’ which includes the ability to form non-basic joint commitments. But not all cases can be captured that way. Sometimes a group’s actions are

¹² This might look very close to summativist views of group attitudes, who hold that a group has an attitude iff all (or most) members have an attitude individually. And in a sense, it is summativist. But the important difference is that the attitudes that are aggregated are we-mode attitudes, instead of I-mode attitudes. The account can therefore capture the usual divergence arguments against summativism. It is possible for a group to believe that p, even though no single member I-believes that p. However, sufficiently many members have to we-believe that p.

best explained by considering that the group has a particular attitude even if no joint commitment of any sort took place. This happens when enough relevant group members have we-mode attitudes such that the group's behaviour in a situation is determined by those attitudes, and the other group members would not renounce those we-mode attitudes and the group behaviour if they had to decide with disclosed information. For instance, take a philosophy reading group that regularly meets to discuss a recently published book. One of the group members, Molly, happens to meet the author and we-mode reasons that it would be a good idea to invite them to one of the reading group sessions. Molly proceeds to invite the author. There was no joint commitment in place that allows Molly to create non-basic joint commitments about inviting people to a group session. However, based on the discussions in the reading group, Molly knows that everyone would love to have the author visit the group for a session. The discussions in the reading group established we-mode cornerstones that are the basis for the we-mode reasoning to invite the author. It seems best to see Molly as a spokesperson for the group, even if she was not officially given such a position. She still speaks for the group, because she reasons in the we-mode layer of agency and the rest of the group would also we-reason just like her, if they were in her position. The group's behaviour is best explained as the group accepting that they should invite the author, even though the we-mode acceptance (and corresponding intention) is only present in Molly. Things would be different if the we-mode reasoning of other group members would have led to different we-mode attitudes. If other members would have we-reasoned that the author should not be invited, were they in Molly's place, then Molly should not be taken as speaking for the group. And this can be observed by looking at the reactions of the other group members when they get the news of the author's visit. If their we-mode reasoning is in line with Molly's, then they will be happy and not have an issue with the invite. If their we-mode

reasoning conflicts with Molly's, they will be unhappy, renounce the invite and deliberate on what to do with Molly's actions. This is captured in (2) by looking at what people with relevant positional affordances would do, if they were to consider whether to form the we-attitude A in a situation.

Some tricky questions are looming. For instance, how many members would need to form the same attitude as Molly for her to count as speaking for the group? Do counterfactual possibilities involving the individual group members all have the same weight? Can some of them be more important than others? These questions are important, but I doubt they can be answered sufficiently on a general basis. Most of the details depend on the structure of the group in question. A group with strict voting processes and hierarchies as part of the cornerstones of the members' we-mode layer of agency looks different to an informal group of friends. However, my aim is much more modest. I merely aim to provide a first outlook of how group attitudes can supervene on we-mode layers of agency in individuals. And my proposal can do the job well enough.

6. Examples

I will end by going through three examples to further illustrate the proposed we-mode as a layer of agency and the proposals for social groups and group attitudes. I want to emphasise that for each example my proposed account can capture everything we want to capture. It provides a plausible story for why and how the group members act with the group's beliefs and goals in mind; why and how the members reason based on the group's perspective with the established cornerstones as central points for reasoning; and it explains the normative demands on the group members to act in a particular way. My proposal does all this while being as ontologically parsimonious as possible. It only relies on humans having the ability for

layered agency – an ability that already seems plausible for reasons unrelated to my account – and requires no plural subject at all.

Example 1 – Poetry Group

A paradigmatic example of Gilbert’s account is the poetry group (Gilbert, 1989, p. 288). Some people meet regularly to discuss poems. Individuals in the group put forward interpretations, others can deny those and offer alternatives. A discussion emerges until the group agrees on the poem or section in question. For instance, the group talks about a poem by Philip Larkin until they all agree that the ending of the poem is especially moving. In future discussions, the group members can refer back to that agreement as something the group takes to be true. The individuals jointly committed to the ending of the poem being moving.

Gilbert’s description of the case also works in my proposed model. How joint commitments come about is the same, and the normative demands that go with that commitment are similar. The difference is in what the joint commitment does and does not. In contrast to Gilbert, in my account the joint commitment does not establish a plural subject holding a group belief. Instead, the joint commitment establishes a we-mode for all the involved parties individually and a commitment to reason within that we-mode in discussions within the poetry group. That we-mode has as a cornerstone acceptance that the end of the Larkin poem is especially moving. Hence, every individual who reasons within the we-mode has to reason such that there is no conflict with that cornerstone acceptance. And because every party involved committed to reason within the we-mode, every party can refer to that cornerstone acceptance in the discussion as a fixed point. The cornerstone acceptance that the end of the Larkin poem is especially moving is also a group attitude. It is a we-mode attitude present in

every party involved and thereby satisfies the conditions for group attitudes. It can be used to explain and predict the group's actions without postulating a plural subject in Gilbert's sense.

Example 2 – Painting a wall together

Bratman (1993) introduced another paradigmatic example into the debate: two people painting a wall. In my proposal, this is captured by both parties forming a joint commitment to paint the wall together. This includes a joint commitment to coordinate in doing so. These joint commitments lead to cornerstones for a we-mode layer of agency. The intention to paint the wall together, the intention to coordinate and the acceptance of how to do so serve as the basis for further we-mode reasoning within that new layer of agency for both agents. Because both have the same cornerstones for their temporary agency and those cornerstones were established by joint commitment, they now constitute a group. The cornerstones also establish some intentions as group attitudes, because they are found in the we-mode agency of all parties involved. Again, just as in the poetry group example, this way of establishing a group attitude via the we-mode with specific cornerstones in the individual group members does not entail any sort of plural subject as it would in Gilbert's (1989) account.

Based on the joint commitment and the synchronised we-mode the agents can coordinate and bargain how exactly they want to go about painting the wall. Such a bargaining process leads to more joint commitments about subgoals and subplans, some of which are positionalised in each we-mode agency in a way that allows better coordination between both parties. Each agent has positionalised cornerstone acceptances and goals that assign roles and tasks to the individuals. For instance, they each have the cornerstone that agent A is in charge of buying the paint, and person B is in charge of removing the old paint from the wall. This way they coordinate such that they are not acting in ways that conflict with each other, e.g.

they are not using different colours and they are not both buying supplies. At the same time, each party can reason independently and form we-mode intentions and attitudes by themselves, as long as they are not in conflict with any of the cornerstones jointly committed to. For instance, the agent responsible for removing the old paint can form subplans on how to remove the paint in the we-mode layer of agency without the need to deliberate together with the other party.

In terms of bargaining and coordination my account captures the same features that Bratman (1993) aims to capture with the example. However, as discussed in part 3, my proposal differs significantly from Bratman. In my proposal there is more to shared intentions than interlocking personal intentions. Instead, I use the notion of joint commitments to establish a we-mode in each involved individual agent. This joint commitment comes with strong normative demands to then act and reason within that we-mode. Because of these demands my account can better explain why those who paint the wall together have particular normative relations to one another and to their parts in painting that wall. For instance, they have a right to rebuke the other person if they do not contribute as they ought to, given the established cornerstones. This is something that Bratman's model struggles with, but my account can capture by drawing on some of the resources from Gilbert (1989; 1990; 2009).

Example 3 – Philosophy Department

A philosophy department is an example of a highly structured group that assigns different members different roles in the department. These are captured by positionalised cornerstones that are established by basic and – predominantly – non-basic joint commitments. As in the other cases, this works without creating any form of plural subject that is the philosophy department. The philosophy department exists merely in virtue of its

members having synchronized we-mode agencies with the same (often positionalised) cornerstones, which are established via the joint commitments.¹³ Employees are taken on board in a process that involves a non-basic joint commitment that allows a person to hire a new employee. Someone joining the department adopts the we-mode layer of agency of the department. Elaborate onboarding processes are required in order to teach new employees about the relevant cornerstones they need to commit to in order to join the group. These cornerstones capture, among other things, the structure of the department: who is responsible for what, how departmental decisions are made, and so on. Departmental decision-making can allow for complex ways of forming new joint commitments that serve as cornerstones for the departmental we-mode layer of agency. For instance, they might include subgroups with distinct roles (e.g. a committee that deliberates on who to admit for the department's graduate programme), or complex, multi-level approval processes (e.g. when the committee puts forward names that have to be approved by the department head and university administration). In some cases, structural features can directly influence joint commitments. For instance, the committee deliberating on graduate admissions might choose a particular candidate because they take that candidate to be reasonably qualified and the most likely to be approved by the administration. They might do so even though no individual committee member believes that this is the best candidate overall.¹⁴ Part of the we-mode reasoning that individual committee members perform is that the approval process has to be taken into consideration. The committee alone does not have the power to form the non-basic joint commitment of admitting a particular candidate. Only in tandem with the agents

¹³ Of course, there are some other features in the larger context of a society that govern philosophy departments. For example, in modern societies some governmental accreditation for higher education institutions are required in the background for a group to exist as a proper philosophy department. I here am mainly interested in emphasizing the difference to Gilbert's plural subject view and hence bracket these other requirements.

¹⁴ Such a case has been put forward by Lackey (2020, p. 188). The example is a good illustration of a paradigmatic divergence case. Similar cases have been proposed by Gilbert (1989), Schmitt (1994) and Tollefsen (2007).

in the approval position they can establish new, non-basic joint commitments admitting a candidate to the graduate programme.

These examples are not exhaustive. Joint commitments, we-mode layers of agency, and group attitudes come in many shapes and sizes. However, the examples provide a good illustration of how my proposal fits with concrete cases discussed in the literature. For all these cases my proposal can deliver a convincing explanatory story. My account provides tools that help us analyze these different kinds of groups and their behavior.

7. Conclusion

I have argued for a new way to conceptualize a we-mode as a form of layered agency. Drawing on Nguyen's (2019; 2020) work on striving play I suggested that acting and reasoning from a group's point of view should also be understood as forming a new layer of agency with particular cornerstones set up jointly with other people. This we-mode layer of agency only presupposes the ability to fully submerge oneself into that layer and joint commitments. It does not presuppose groups as a foundational part. Instead, the we-mode layer of agency can be used as a conceptual building block to define social groups and their intentional states. I have provided a first attempt at doing that and illustrated it with examples. I am confident that the big picture I presented can provide valuable explanatory work in understanding social groups. I am less confident about the details of putting that big picture to concrete definitional work. Likely some adjustments will be required in the future. However, I believe this is a promising start for the concept of layered agency in social philosophy.

References

Bratman, M. E. (1993). Shared Intention. *Ethics*, 104(1), 97-113.

- Camp, E. (2023). Agency, Stability, and Permeability in "Games". *Journal of Ethics and Social Philosophy*, 23(3), pp. 448-462.
- Gibson, J. J. (1966). *The Senses Considered as Perceptual Systems*. London: Allen and Unwin.
- Gilbert, M. (1989). *On Social Facts*. London: Routledge.
- Gilbert, M. (1990). Walking Together: A Paradigmatic Social Phenomenon. *Midwest Studies in Philosophy*, 15(1), 1-14.
- Gilbert, M. (2008). Two Approaches to Shared Intention: An Essay in the Philosophy of Social Phenomena. *Analyse & Kritik*(30), pp. 83–514.
- Gilbert, M. (2009). Shared intention and personal intentions. *Philosophical Studies*(144), pp. 167–187.
- Gilbert, M. (2017). Joint Commitment. In M. Jankovic, & K. Ludwig (Eds.), *The Routledge Handbook of Collective Intentionality* (pp. 130-139). London: Routledge.
- Gilbert, M., & Pilchman, D. (2014). Belief, Acceptance, and What Happens in Groups: Some Methodological Considerations. In J. Lackey (Ed.), *Essays in Collective Epistemology* (pp. 189–212). Oxford: Oxford University Press.
- Hakli, R. (2007). On the Possibility of Group Knowledge Without Belief. *Social Epistemology*, 21(3), 249-266.
- Lackey, J. (2020). Group Belief: Lessons from Lies and Bullshit. *Aristotelian Society Supplementary Volume*, 94(1), pp. 185-208.
- Nguyen, T. C. (2019). Games and the Art of Agency. *Philosophical Review*, 128(4), pp. 423-462.
- Nguyen, T. C. (2020). *Games: Agency as Art*. New York: Oxford University Press.

- Rovane, C. (1997). *The Bounds of Agency*. Princeton: Princeton University Press.
- Rovane, C. (2002). From a Rational Point of View. *Philosophical Topics*, 30(1), pp. 209-235.
- Schmid, H. B. (2017). What Kind of Mode Is the We-Mode? In G. Preyer, & G. Peter (Eds.), *Social Ontology and Collective Intentionality: Critical Essays on the Philosophy of Raimo Tuomela with His Responses* (pp. 79-94). Cham: Springer International Publishing.
- Schmitt, F. F. (1994). The Justification of Group Beliefs. In F. F. Schmitt (Ed.), *Socializing Epistemology: The Social Dimensions of Knowledge* (pp. 257-287). Lanham, MD: Rowman and Littlefield.
- Schmitz, M. (2017). What is a mode account of collective intentionality? In G. Preyer, & G. Peters (Eds.), *Social Ontology and Collective Intentionality: Critical Essays on the Philosophy of Raimo Tuomela with his Responses* (pp. 37-70). Cham: Springer.
- Schmitz, M. (2018). Co-Subjective Consciousness Constitutes Collectives. *Journal of Social Philosophy*, 49(1), pp. 137-160.
- Schmitz, M. (2023). From We-Mode to Role-Mode. In M. Garcia-Godinez, & R. Mellin (Eds.), *Tuomela on Sociality* (pp. 177-200). Cham: Palgrave Macmillan.
- Tollefsen, D. (2007). Group Testimony. *Social Epistemology*, 21(3), 299-311.
- Tuomela, R. (1992). Group Belief. *Synthese*, 91(3), pp. 285–318.
- Tuomela, R. (2005). We-intentions Revisited. *Philosophical Studies*, 125(3), 327–369.
- Tuomela, R. (2013). *Social Ontology: Collective Intentionality and Group Agents*. Oxford: Oxford University Press.
- Wray, B. K. (2001). Collective Belief and Acceptance. *Synthese*, 129(3), 319-333.