Journal for the History of Analytical Philosophy
Volume 15, Number 6 (2026)

A Moderate Collectivist Reading of Sellarsian We-Intentions

Jeremy Randel Koons
koonsj@georgetown.edu
Georgetown University in Qatar

ABSTRACT

A lively debate has arisen as to whether we should interpret Sellars as an individualist or a collectivist about we-intentions. Individualism holds that an agent's we-intending does not depend on the we-intending of the larger group. The collectivist holds, by contrast, that an individual's we-attitude presupposes the corresponding group attitude. Sellars's view of we-intentions, and his view on reasoning among intentions, underwent significant shifts during his career. Most recent interpretation of Sellars reads him as an individualist. However, I will argue for a moderate collectivism, one that offers a reconciliation between stronger versions of collectivism and individualism while at the same time insisting on the priority of collectivism. On this account, the paradigm or "reference" case—the case most revealing of the nature of we-intentionality—requires actually shared intentions. Individual unshared intentions are possible, but necessarily exceptional, and parasitic on the paradigm case. Thus, the shared intentions of a group have explanatory priority over an individual's we-intention, and the latter is dependent on the former. I will conclude that moderate collectivism is not only textually supported by Sellars's late articles, but that it is the most philosophically sound position as well.

1. Introduction

Wilfrid Sellars offered the first philosophical account of we-intentions. As his account perhaps lacks the sophistication of later accounts, it is often misleading to apply the categories from the contemporary collective intentionality literature to Sellars's own theory.[1] However,

[1] This does not, however, prevent us from attempting what Loeffler (2020) calls a "reconstruction" of Sellars's theory, which involves updating it using contemporary theoretical tools while attempting to maintain its key features and central commitments.

some categories have clear relevance; and a lively debate has arisen as to whether we should interpret Sellars as an individualist or a collectivist about we-intentions.[2] Individualism holds that an agent's we-intending does not depend on the we-intending of the larger group. Defenders of this reading often appeal to passages where Sellars writes, for example, "Let me emphasize that from the fact that Smith values something, X, as one of us, it doesn't follow that 'We value X'" (OIM, 180/§127).[3] The collectivist holds, by contrast, that "an individual's we-attitude presupposes the corresponding group attitude…[This we-attitude] presuppose[s] a group attitude, which itself presupposes either a set of mutual beliefs or mutual expectations" (Koons 2019, 91).

Sellars's view of we-intentions, and his view on reasoning among intentions, underwent significant shifts during his career.[4] Most recent interpretations of Sellars read him as an individualist.[5] However, I will argue for a *moderate collectivism*, one that offers a reconciliation between stronger versions of collectivism and individualism while at the same time insisting on the priority of collectivism. On this account, the paradigm or "reference" case—the case most revealing of the nature of we-intentionality—requires *actually shared* intentions. Individual unshared intentions are possible, but necessarily exceptional, and parasitic on the paradigm case. Thus, the shared intentions of a group have explanatory priority over an individual's we-intention, and the latter is dependent on the former. In the course of my argument, I will address recent individualist readings of Sellars (particularly by Stefanie Dach and Kyle Ferguson). I will conclude that moderate collectivism is not only *textually* supported by Sellars's late articles, but that it is the most *philosophically* sound position as well.

[2] See, for example, Koons (2019), Dach (2023), Ferguson (2023), and Loeffler (2023).

[3] Other examples include: "'We shall do A' need not, as some critics have supposed, be 'chorused.' It is a form of practical thought which can go on *in foro interno*. I can think in terms of *we*…Obviously people need not agree in their logically intersubjective intentions. The point is that they *can* literally agree" (CSDS, 734/§10); and "an individual can have an intention of intersubjective form even if no one else in point of fact shares it" (OIM, 184/§143).

[4] See Dach (2023).

[5] See Dach (2023) and Ferguson (2023).

2. Individualist Readings of Sellars

Individualist readings of Sellarsian we-intentions have dominated the recent literature.[6] One of the most convincing arguments is offered by Kyle Ferguson (2023). As a way of settling on a plausible account of shall-expressions, Ferguson introduces the method of *reportorial ascent*. The method is (in principle) simple: There is an intimate connection between expressions and ascriptions of states such as beliefs and intentions. An expression is sincere and appropriate when the corresponding ascription is true. Thus, we can shed light on various intention-expressions by examining when their corresponding self-ascriptions are true. Ferguson offers the following analogy with belief: "The performance conditions for an assertoric thought-expressing sentence are satisfied when the truth conditions for a corresponding thought-reporting sentence are satisfied. In other words, '*p*' is sincere when 'I think that *p*' is true" (Ferguson 2023, 48). Similarly, Ferguson writes, "'I shall do A' is sincere when 'I intend to do A' is true" (Ferguson 2023, 48).

Hopefully, then, we can get clear on what exactly is expressed by "We shall any of us do A in C" by figuring out what true intention-*ascription* corresponds to sincere and appropriate instances of the former intention-*expression*. After examining (and rejecting) three options, Ferguson provisionally settles on a fourth: "I as one of us intend to do A in C" (Ferguson 2023, 53). Ferguson argues for the adequacy of this analysis as follows:

> Because 'I' is the subject, the reported intentions need not be chorused or shared. That is, these reporting sentences' truth conditions can be satisfied depending on the speaker's mental state. But because of the "as one of us," the reported intentions are shareable. For the we-intending—the state, not the content—is a mental state one has as one of us. Others may join in so intending. (Ferguson 2023, 52)

(Ferguson actually makes one further amendment to this analysis, arguing that to reflect the intersubjectivity of both form and content, the

[6] See, for example, Dach (2023), Ferguson (2023), and Loeffler (2023). Olen and Turner (2015, 2016) arguably offer a collectivist reading of Sellars, but I will not engage heavily with them as (a) they chiefly focus on Sellars's early to middle writings, whereas I want to focus on his later writings; and (b) they are more interested in careful analysis and explication of Sellars's texts and the relation of his thought to that of his intellectual predecessors; they have little interest in applying the concepts of the contemporary debate to Sellars's theory.

relevant intention ascription should read, "I as one of us intend any of us to do A in C" (Ferguson 2023, 53). However, Michael Hicks points out that this amended version is problematic: "If I intend any of us to do A, presumably I intend Smith to do A. But this seems to imply that in taking A to be obligatory, I am committed to enforcing that obligation – doing what it takes to ensure that Smith do A. This is controversial at least" (Hicks 2023, 5). But in any case, my specific worry concerns both of Ferguson's formulations.)

Unfortunately, I think both formulations fail to answer the central question: *What is it to intend as one of us?* On Ferguson's analysis, this is left unexplained. To be fair, Ferguson recognizes this problem, writing later in the chapter, "it remains unclear what it is to intend as one of us. Isn't a much more substantial elaboration on the nature of intending as one of us required for a fuller understanding the state of we-intending?" (Ferguson 2023, 55) He concedes that this is a pressing issue, writing that "A more fully developed individualist account of we-intending must do much more than I have done in this chapter. . .Until we have a more developed account of 'intending as one of us,' we have only a sketch of the reports' truth conditions and thus the expressions' performance conditions" (Ferguson 2023, 55). However, until and unless such an analysis is offered, we are left without an answer to the question, "What is it to intend as one of us? And how does this account for the intersubjectivity of we-intending?" For example, it may be that to intend "as one of us"—as in "I as one of us intend to do A in C"—it is necessary that "and others of us intend to do A in C" enters my intention-content *via* So-be-it[7], or that the parent intention ("We shall any of us do A in C") be (or be derived from) an *actually* shared intention. These would both introduce collectivist elements into the target intention-expression, and Ferguson hasn't ruled them out. (Indeed, I think that the latter analysis approaches the truth of the matter. But more on this anon.)

A reader has commented that to establish his individualist thesis, Ferguson does not have to answer this question ("What is it to intend as one of us?"). It is sufficient that he demonstrate—*via* reportorial ascent—that individualism about we-intentions is true, not *how* it can

[7]For a discussion of So-be-it, see ORAV, 481ff/§§66ff. Presumably, anything that enters the content of an intention-expression *via* So-be-it would also be in the content of the corresponding intention-ascription, since these contents are the same—something Ferguson seems committed to by his *identity thesis* (Ferguson 2023, 41).

be true. I reply, however, that he hasn't established that individualism *is* true if his individualistic analyses fail to account for central features of we-intending, viz., their intersubjectivity. That is the Achilles' heel of such accounts, and reportorial ascent cannot by itself establish the truth of individualism if no individualistic intention-ascriptions can be offered that fully account for all features of we-intention ascriptions. Ferguson's concession—mentioned in the previous paragraph—that we need "a more developed account of 'intending as one of us'" offers us no guarantee that such an account can be offered consistent with individualistic premises.

Let us turn to Dach's analysis. Dach (2023) argues that Sellars's view of what makes an intention intersubjective undergoes a significant evolution during his career, culminating in his final publication, "On Reasoning About Values" (ORAV). I wish to extract one important lesson from her analysis—namely, her conclusion regarding what (in the mature Sellars) distinguishes we-intentions from I-intentions. Dach argues that on the face of it, the intention

> Shall be [each of us scratch my back]

could *either* be a personal intention, *or* it could be something I intend *qua* member of a community (say, the Society of Friendly Back Scratchers). What distinguishes whether this is a personal intention or a we-intention[8] is the *route of derivation*. If I derived the intention from

> Shall [I, the intender, lead a satisfying life]

then it is a personal intention. If, instead, I derived it from

> Shall [each of us members of the Society of Friendly Back Scratchers foster our common welfare]

then it is a we-intention. Thus, Sellars writes,

> If a person *has* the intention
>
> Shall [I do *A*]
>
> and has it *because* it is implied by the intentions

[8]Dach does not use the language of we-intentions, as she argues that "the terms 'we-intention' and 'I-intention' may not be fine-grained enough for a clear reconstruction of what Sellars was up to" (Dach 2023, 18). However, for my simpler purposes, the language will suffice.

> Shall [any of us do *A*]
>
> which he also *has*, we can say that Jones intends to do A *sub specie* "one of us," and flag our representation of his intention with a subscript "we," thus,
>
> Jones intends "Shall_{we} [I do *A*]"
>
> (ORAV, 505)

Like Ferguson, Dach endorses an individualist (rather than a collectivist) view of Sellarsian we-intentions (Dach 2023, 32–33, n 30). However, I will argue that Sellars's account in ORAV encourages the opposite.

3. A Moderate Collectivist Reading of Sellars

Dach identifies a key feature of Sellars's mature view: We-intentions are distinguished by their route of derivation. Sellars articulated a version of this view in a 1976 letter to David Solomon, shortly before the publication of ORAV:

> The moral point of view involves the form
>
> '*We* shall any of *us* do A, if C,'
>
> which entails
>
> 'I shall do A, if C.'
>
> To flag its *origin* (or *ground*) in a logically intersubjective intention, I represent the latter as
>
> 'Shall_{we} [I do A, if C].'
>
> (ORAV, 505/§180)

How does this bear on the question of individualism vs. collectivism in our interpretation of Sellarsian we-intentions? First, let us focus on the intrinsically reasonable intention constituting the moral point of view: "Shall [each of *us* do that which, in the circumstances, promotes the happiness of each and every one of *us*, all relevant things considered]" (ORAV, 505/§194). Crucially, Sellars holds that this intention must be *actually shared* by members of the moral community. As Dach notes, "[Sellars] argues that there is an intention which is not merely shareable but must be shared by all those forming a certain group (in the limiting case, all rational beings)...According to Sellars, a number of people

sharing such an intention is what constitutes groups in the first place (*SM*, chap. VII, §132). Without such a shared intention, there is no group or community" (Dach 2023, 25).

Sellars argues for this latter conclusion stepwise. In ORAV, 506–07/§§184–190, Sellars analyzes what it is to intend, for example, *qua* member of the Whooping Crane Society (WCS). The crucial passages are these:

> ...[T]he purpose or intention expressed by
>
> > Shall [each of us WCS members promote the survival of whooping cranes]
>
> is reasonable *for members of the WCS*, because it is by virtue of that shareable intention that there is such a thing as the WCS.
> . . .This intention defines what might be called "the WCS member's point of view." It is the prime mover intention of this point of view, and generates, given matters of factual implications and premises, subordinate intentions, the relative reasonableness of which is grounded in its intrinsic and non-relative reasonableness as constituting the WCS point of view.
> (ORAV, 506–07/§§186–87)

There are two important things to note about this passage. First, the WCS only exists in virtue of this "founding intention" (as I shall call the shared intention that creates a specific community) spelled out in §186. This intention is seen by members of the WCS as intrinsically reasonable, but—crucially—it must be an *actually existing* and *actually shared* intention for the WCS to exist. There is only a WCS because people jointly intend for there to be one. Second, when I correctly intend *qua* member of the WCS, it is because my intention is derived from this founding intention. Thus, my reasoning can be represented as follows:

> Shall [each of us WCS members promote the survival of whooping cranes]
> [each of us WCS members promote the survival of whooping cranes] implies [each of us does A in C]
> $\text{Shall}_{\text{WCS}}$ [I do A in C]

As per Sellars's letter to Solomon, the subscript "WCS" flags this intention's origin in the founding intention of the WCS.

The next move Sellars makes is to ask whether there is an analogous intention constituting the moral community: "Is there, then, an intention

which defines the moral point of view? Is there an intrinsically reasonable we-referential action intention which stands to a moral community as the intention that 'we' promote the survival of whooping cranes stands to members of the Whooping Crane Society?" (ORAV, 507/§189). Of course, Sellars's answer is "yes"; but the crucial point of the analogy is that not only must this intention be intrinsically reasonable for members of the moral community, but it must *in fact be shared* by these members. Further, to intend *qua* member of the moral community is to entertain an intention which is derived from this shared, intrinsically reasonable foundational intention. Thus, we have the intrinsically reasonable and *in fact shared* intention constituting the moral community:

> Shall [each of *us* do that which, in the circumstances, promotes the happiness of each and every one of *us*, all relevant things considered]. (ORAV, 508/§194)

We-intentions (of the form "Shall_{we} [I do A_i in C_i]") are such in virtue of their derivation from this shared intersubjective founding intention. (See ORAV, 505/§§179–80.)

This element of Sellars's moral philosophy was not a late innovation of his; he held this view at least for the second half of his career. In Chapter 7 of *Science and Metaphysics* (*SM*), he also expresses the view that members of the moral community must *in fact* share the foundational intention constituting this community: "It is a conceptual fact that people constitute a community, a *we*, by virtue of thinking of each other as *one of us*, and by willing the common good *not* under the species of benevolence—but by willing it as one of us, or from a moral point of view" (OIM, 181/§132). Further, moral imperatives are intentions that are derived *via* various nomologicals from this founding intention. (See OIM, 178/§§121–22.)[9]

How does this bear on the individualist/collectivist reading of Sellarsian we-intentions? An individually-held we-intention *presupposes* a collectively held intention: You cannot intend something *qua* member

[9]See also OIM, 184/§142, where Sellars writes, "Can we say that rational beings generally constitute a community? They would do so if they shared the intersubjective intention

> It shall_{we} be the case that each of us rational beings so acts as to promote our welfare."

of the WCS if there is no WCS. Similarly, intending from the moral point of view can only happen if there is an existing moral community (a claim that communitarians and other assorted Hegelians have been defending for decades). The individually-held intention is a we-intention in virtue of its derivation from this collectively-held intention. Thus, Sellars's theory of we-intentions has collectivism built into its very structure. Further, consider the derivation of individual we-intentions. As Sellars writes in Chapter 5 of *SM*, "ethical truths are the projections of 'matter-of-factual' truths. . .into the framework of intentions and purposes" (*SM*, 105/§3). Thus, as indicated above, ethical reasoning proceeds as follows:

> 1. . . Shall [each of us do that which in the circumstances promotes the well-being of each and every one of us, all relevant beings considered]
> 2. . . [each of us promotes the well being of each and every one of us] entails [each of us does A_i in C_i]
> C. Therefore, Shall_{we} [I do A_i in C_i]

Premise 1 is the founding, *shared* intention of the moral point of view (in the version stated in ORAV, 509/§206). Premise 2 is a nomological—a causal law which is true independent of any belief or communal agreement. Thus, the conclusion represents an intention that should also be chorused. On Sellars's view, then, disagreement can only arise out of an error or mistake—one has misapplied the supreme principle, or used a false nomological, or inferred incorrectly. While this represents a simplified view of moral reasoning, one can view it as a model that clarifies why disagreement creates normative pressure to resolve this clash of moral intentions (just as incompatible beliefs creates a rational pressure to resolve the contradiction).

But the essential thing to note is that individual moral intentions *presuppose and are parasitic upon* collective, shared intentions. This is characteristic of the *moderate collectivism* I am attributing to Sellars. Unshared we-intentions are possible, but they (a) presuppose a background of actually shared intentions and (b) only count as we-intentions *because* they are derived from some subset of these background shared intentions.

One might object that I am ignoring some of the nuance of Sellars's discussion in OIM and ORAV. For while Sellars does identify "the moral community with that group with which it is reasonable for us, all relevant things considered, to identify"—viz., "rational beings generally" (ORAV, 509/§206)—he qualifies this by adding, "I am *not* saying that

everybody shares this shareable intention. I am simply saying that it defines the moral point of view, as contrasted with, say the WASP point of view" (ORAV, 509/§207). And in OIM §145 he expresses doubts about his ability to establish the intrinsic reasonableness of the founding intention of the moral community.[10]

We need to distinguish between three separate theses here:

> 1) The founding intention of the moral community is intrinsically reasonable for all rational beings.
> 2) The founding intention of the moral community is in fact shared by all rational beings.
> 3) A founding intention must in fact be shared for a community to exist, and one's we-intending *qua* member of the community is parasitic upon this background shared intention. (This is the thesis of moderate collectivism.)

As a rule, Sellars is fairly clear in distinguishing among these theses—and he is clear that when it comes to whether communities actually exist, what is at issue is (3): "A community exists if the relevant individuals think of themselves as its members" (OIM, 185/§144). If Sellars expresses doubts about his ability to establish the intrinsic reasonableness of the founding intention of the moral community, or about whether all rational beings in fact share this founding intention and thereby constitute a moral community, this doesn't demonstrate that he does not in fact have the collectivist commitment that a community can only exist—and so one can *a fortiori* only intend qua member of that community—if the members of that community in fact share the founding intention constituting that community. The "skeptical" passages don't threaten this thesis.

Consider, for example, OIM, where he is trying to establish (1), the claim about reasonableness—namely, that being a rational agent implies (as a conceptual matter) being a member of the moral community of all rational beings, normatively bound by the founding intention of this moral community. In OIM §144 he states the two central premises of such an argument and writes, "These premises would entail that the concept of oneself as a rational being implies the concept of oneself as a member of an ethical community consisting of all rational being" (OIM, 185/§144). When he admits in the next paragraph that "the

[10] An anonymous referee for *JHAP* raised this objection.

argument for the reality of an ethical community consisting of all rational beings. . .remains incomplete," he is merely admitting that he cannot conclusively demonstrate that this founding intention is intrinsically reasonable for all rational beings. But crucially, he concludes §144 by writing, "Yet if the above premises were true, all rational beings would 'implicitly' think of themselves as members of an ethical community consisting of all rational beings. *But since a community exists if the relevant individuals think of themselves as its members*, the ethical community of rational being would have an 'implicit' existence" (OIM, 185/§144, emphasis added). All rational agents would be rationally committed to seeing themselves as members of the moral community and bound by its founding intention, whether or not they in fact do. But in any case, the community doesn't really exist (as opposed to having "implicit" existence) without the *de facto* commitment of its members—difficulty establishing (1) doesn't entail the falsity of (3).

Consider Sellars's skepticism regarding thesis (2)—as to whether the moral community of all rational beings actually exists, as expressed in the above-quoted passage (ORAV, 509/§207). Again, though, this doesn't in any way show that the moral community can exist if its founding intention is not shared. It is exegetically odd to think (a) that Sellars can express some skepticism over whether the founding intention of the moral community can be shown to be intrinsically reasonable, but also think (b) that he *cannot* express any skepticism over whether the moral community in fact exists (or whether it in fact encompasses all rational beings).

And indeed the most plausible reading of Sellars is that the construction of the moral community is an ongoing project, and that the moral community of all rational beings is (like Sellars's "Peirceish conceptual structure") an ideal whose achievement is likely ever out of reach. For example, in OIM §124 he laments "the tribocentricity of moral judgments in the not too remote past"; it is likely he would concede that this has yet to be fully overcome (and that its overcoming is, as Rorty (1989) would argue, a task that will require our constant striving). I will return to this topic—the scope of the moral community—in Section 4.1.

I conclude that it is plausible to read Sellars as having a significant collectivist commitment. Nevertheless, as Sellars presents his view, I think it understates the collectivist element. Part of this is due to the

simplified model of moral reasoning he uses in his writings on moral theory. Sellars's holistic view of justification is, I suppose, well-known; and by elaborating on this picture as it relates to morality, we will be in a better position to make the (moderate) collectivist case.

4. Sellars as a Proto-Contextualist

Sellars's presentation of his moral theory has the potential to mislead due to its particular structure. The reasonableness of intermediate moral principles—Sellars's "categorical imperatives," which have the form "Shall$_{we}$ (I do A_i in C_i)"—is derivative (*via* nomologicals) from the reasonableness of the supreme principle of morality (Shall [each of us do that which in the circumstances promotes the well-being of each and every one of us, all relevant beings considered]). Sellars is clear elsewhere that moral justification has a "top down" structure. For example, in SE he writes, "Particular moral judgments are justified by subsuming them under intermediate principles or maxims, and. . .intermediate principles are justified by subsuming them under more general principles and, ultimately, first principles" (SE, 100/§16). However, this cannot be the whole story. As I argue at length in my (2019), Sellars's model cannot be correct.[11] This *formalism* about moral reasoning contradicts his deeper commitment to anti-formalism about reasoning, and in particular the emphasis he plays on the role of material inferential principles. Thus, it behooves us to try to understand how moral reasoning would actually work for someone with Sellars's philosophical commitments, and what consequences this has for moderate collectivism.

Sellars is a logical expressivist: Statements such as "If this is copper, then it is metal" make explicit inferential proprieties that are implicit in practice. Analogously, Sellars *should* hold that the goodness of "Shall$_{we}$

[11] For starters, practical premises and conclusions are not connected by nomologicals, for the simple reason that—given my preferences, my current circumstances, and so on—there are in general multiple equally good ways of satisfying a given end: If I am hungry and want to eat dinner, there are multiple equally satisfactory options available to me; there is no lawlike connection between "Shall (I eat dinner)" and any particular means of fulfilling this intention. One consequence of this is that the simple model of moral reasoning Sellars uses cannot be accurate, since individual we-intentions cannot simply be derived from the supreme principle of morality. See (Koons 2019, chap. 6).

(I do A_i in C_i)" is parasitic on the goodness of material inferential proprieties of the form:

> I am in C_i
> Therefore, $Shall_{we}$ (I do A_i)

It is not good because it is derived from supreme principle of morality. Thus, "$Shall_{we}$ (I do A_i in C_i)" is a *material principle of inference*, one whose goodness expresses an existing practical inferential propriety.

These material principles of inference have a *prima facie* authority, as expressing practical inferential proprieties. Sellars makes this clear in "Some Reflections on Language Games" (SRLG) when he is discussing how one comes to occupy a position in a language game. The two most obvious ways are through language-entry and language-language transitions. But there is a third way. The language game contains certain 'free positions' or 'auxiliary positions,' which are sentences "which can properly be occupied at any time if there is any point to doing so" (SRLG, 330/§25). These auxiliary positions are of two sorts: formal and material principles of inference. The latter are of particular concern to us here. Material principles of inference such as 'Here is smoke, therefore here is fire' and 'All colours are extended' (SRLG, 331/§28) may be taken up whenever appropriate, and Sellars is explicit that "one comes to be at [auxiliary] positions without having moved to them from other positions" (SRLG, 330/§25). That is, they are not and need not be inferred from any other position. Your entitlement does not depend on having done so. Thus, I am entitled to assert, "If I strike this dry, well-made match, then it will light," or "If my friend is in trouble and open to receiving aid, then ceteris paribus I ought to help her," even though I haven't inferred either principle from another proposition (or set of propositions).

The picture of moral justification that emerges is an ethical anti-foundationalism that prefigures the ethical contextualism of later authors such as Mark Timmons (1999). Consistent with ethical contextualism and Sellars's comments from SRLG about "auxiliary positions," we can adopt a "default and challenge" view about moral claims. Expressions of we-intentions (like assertions) come with a default assumption of reasonableness that only needs to be defended in the case of legitimate challenge. In the case of such a challenge (or in the case of conflict among intermediate moral principles), we can appeal to higher-order

principles to resolve the disagreement. However, this is consistent with regarding the supreme principle of morality as a higher-level dialectical posit we use to resolve conflicts among intermediate moral principles, a move that preserves the hierarchical structure of Sellars's moral theory.[12]

Thus, we can paint a more holistic picture of justification on Sellars's moral theory. We can justify intermediate principles by appeal to the supreme principle of morality. In this sense, justification is flowing "top down"—from the highest-order principle to lower order principles. But a key feature of Sellars's anti-foundationalism is (or should be) that justification never flows only in one direction—lower-level principles can be used to justify or revise higher-order principles. Thus, one can move from a particular claim ("You ought to visit Belinda in the hospital") to the level of an intermediate moral principle ("*Ceteris paribus*, you ought to provide comfort to your friends when they are in need") to the level of the supreme principle of morality. Indeed, it is commonly acknowledged that *particular* experiences can lead to the revision of more *general* moral commitments, as when bigots revise their views upon spending time with the objects of their bigotry.[13]

This (patently revisionary) reading of Sellars as a moral contextualist might seem like a digression, but it is important for a number of reasons. First, as it stands, the structure of Sellars's moral theory is simply incompatible with his deeper commitments, particularly concerning the authority of material inferential principles. Thus, to bring his theory of moral justification in line with his other epistemological commitments, it is important to revise the "upside down foundationalism" of Sellars's theory—where all justification ultimately traces back to a single principle—in favor of a more holistic picture.

Second—by Sellars's admission—in moral reasoning people generally do not have a particular first principle of morality up their sleeve[14]; and if they do, they do not explicitly reason from it. Perhaps this is why

[12] I am grateful to Stefanie Dach for suggesting how to present several of the ideas in this paragraph.

[13] See, for example, Nguyen's (2018) fascinating re-telling of the story of Derek Black, who was raised as a neo-Nazi. While Black was in college, a fellow undergraduate (Matthew Stevenson) began inviting Black to his family Shabbat dinners. Black eventually left the neo-Nazi movement, even becoming an anti-racist activist.

[14] "Though one who makes a singular ought-statement is committed to support it with a universal ought-statement, it is not necessary that he have the latter 'up his sleeve'" (IIOR, 370/§8).

Sellars often writes of "derivable" intentions rather than of "derived" or "deduced" intentions. This doesn't rule out that the first principle is the ultimate court of appeal in justifying our behavior and our intermediate principles—a principle we work our way up to in the process of justifying these latter items—but its absence in everyday moral reasoning suggests, at least, that the structure of moral justification has a somewhat more complicated structure than is suggested by Sellars's simpler picture.

More seriously, though, even if people do have in mind a first principle of morality, it won't necessarily be the same one. On Sellars's deductive-nomological model, does this raise incommensurability worries? Do people endorsing different "founding intentions" even belong to the same moral community? A better picture of moral reasoning—and of the constitution of the moral community—emerges if we do not give the moral first principle the exalted status that Sellars does.

Finally—and related to the previous point—Sellars's formalist model offers a flawed account of what it is to belong to a moral community, an account subject to a number of worries (some of which I have just outlined). The contextualist view—aside from just being overall more consistent with Sellars's collateral commitments about justification—offers a better understanding of what it is to belong to a community and to intend as a member of that community. Understanding contextualism will help us better understand moderate collectivism.

4.1. Moderate collectivism and the constitution of the moral community

On Sellars's simplified model of moral reasoning, the moral community is constituted by an actually-shared "founding intention," and individually-held we-intentions are such because of their derivation from this founding intention. This model (as all models) idealizes along various dimensions; but I argue that the anti-foundationalist, contextualist view described above in Section 4 more accurately describes the structure of moral reasoning and justification for Sellars.

However, this leaves us with two questions:

1) On Sellars's simplified model, to be a member of the moral community is to be committed to the "founding intention" of this community, the supreme principle of morality. But on the con-

textualist view, the supreme principle of morality is a dialectical posit, not necessarily a psychological reality. What, then, is it to be a member of the moral community?

2) If, on the contextualist view, individual moral we-intentions are not in fact derived from the supreme principle of morality, then in virtue of what do they count as we-intentions?

In the remainder of Section 4, I will answer these questions and argue that Sellars's contextualism actually strengthens the collectivist dimension of his moral theory.

Let us take as our starting point Sellars's simplified, top-down model, in which all categorically valid imperatives are derived from the supreme principle of morality. Imagine embracing a parallel view about concept possession in general. The analogy is tenuous as an argumentative one; but it is perhaps useful to illustrate the present point. Suppose we identified concept possession with mastery of a single inferential rule—the "master rule" defining the concept in question. Thus, for example, the master rule for "red" might be "Respond to red objects with a belief token *red*." The details are not important. What matters is that Sellars will regard any such attempt as obviously misguided and bizarre.[15] To possess a concept, one must not have mastered a single rule (even if other rules logically follow from this rule). Rather, one must have mastered a battery of different transitions—language-entry, language-language, and language-exit. One must be able to respond to red objects with judgments that the object is red—but one must understand standard viewing conditions, and know when to withhold such judgments. One must be able to make appropriate inferences involving the word (e.g., from "x is red" to "x is colored"). One must understand how the concept combines with other concepts to create new meanings, and how this creates practical implications (e.g., "*red light* means 'stop'"). The list of capabilities one must possess in order to have mastery of this concept is vague and open-ended—different people have different capacities to discriminate, different inferential knowledge, and so on—but the point is that each person must possess some set of

[15]Sellars, of course, thought that the meaning of empirical terms was constituted by the material inferential rules governing their use, but also conceded that with respect to natural languages, "the vague, fluctuating, and ambiguous character of ordinary usage extends to these rules" (ITSA, 283/§13).

capacities to count as knowing what "red" is. Crucially, these capacities will significantly overlap. If we didn't, for example, largely coincide in our color judgments, then such judgments would have little point (and it is doubtful we would even be making *judgments*—with real conceptual content—in any full-fledged sense).

We should take a similar approach to commitment to the moral point of view. Commitment to this point of view does not involve embracing a single master principle—rather, it involves commitment to a variety of intermediate moral principles, practical proprieties—after all, morality in the first instance involves a variety of *practical* commitments—metacommitments regarding the point of morality, etc. Again, this list will be vague and open textured, allowing for disagreement.

The answers to our two questions from the beginning of Section 4.1 are implicit in the above, but let us make them explicit. Let us start with question 1: On the contextualist view, what is it to be a member of the moral community? What is shared on the more holistic view is not commitment to a single overarching principle, but commitment to a range of principles and norms. As I said above, this list is open-ended and allows for disagreement. But sharing in a "form of life" is what creates a community, whether it is a linguistic community, a community of whooping crane activists, or a moral community.

Let us turn to the second question: If, on the contextualist view, individual moral we-intentions are not in fact derived from the supreme principle of morality, then in virtue of what do they count as we-intentions? Again, the answer should be clear. As a conceptual matter, disagreement is only possible because of this shared agreement; the intersubjectivity of even idiosyncratic judgments depends—as a conceptual matter—on this background of shared attitudes. As Genia Schönbaumsfeld writes, "It is our shared linguistic practices that make meaning, and, hence, disagreements in opinion possible" (Schönbaumsfeld 2016, 71–72); and "without a background of shared concepts and responses. . . disagreement would be conceptually impossible" (Schönbaumsfeld 2016, 81–82).

Individual moral we-intentions display another dimension of dependence on our shared moral commitment—one that is particularly relevant to the second question. Although individual moral we-intentions are not derived from the supreme principle of morality, they are inferentially

dependent upon our shared moral commitments in a fundamental way. Sellars argues that "singular ought-statements imply universal ought-statements" in a special "dialectical" sense of "implies", in the sense that "one who makes a singular ought-statement is committed to support it with a universal ought-statement, [though] it is not necessary that he have the latter 'up his sleeve'" (IIOR, 370/§8). When we do have a moral dispute or disagreement, what one must ultimately appeal to in order to justify one's individual we-intention must eventually circle back to some element of the shared practice, some common ground. If you fail to appeal to some common ground, I will take you to have failed to discharge your burden of proof, to have failed to demonstrate your entitlement to the commitment in question. Thus, to be in the space of reasons—to engage in the game of giving and asking for reasons—requires that one have commitments that are inferentially dependent on our shared commitments, although in the special "dialectical" sense identified by Sellars.[16]

Sellars's late view is that membership in the moral community is determined by commitment to the supreme principle of morality. This represents an oversimplification, but also contains a key insight. Exploiting this insight allows a more nuanced view to emerge of what it is to be a moral agent—one of us. But this new view makes it even more clear that this involves agreement on a wide range of attitudes. Further, any disagreement takes place in the context of this more general agreement. Thus, this modified view of Sellars makes the collectivist element even more apparent.

[16] An anonymous referee for *JHAP* objects that on my account, there is no longer a distinction between I-intentions and we-intentions, since all intentions depend on a background of shared intentions. Here, we must distinguish between conceptual dependence and justificatory dependence. Intending as part of a group depends on the existence of something like what Tuomela calls an *ethos*, which is "the set of *constitutive* goals, values, beliefs, standards, norms, practices, and/or traditions that give the group motivating reasons for action" (Tuomela 2007, 16). I intend *qua* member of the community to the extent that my intention is *justified* (at least partially) in terms of this ethos. I offer a more formal account of this in my (2022). By contrast, my intention, say, to earn PADI Open Water Scuba Diving Certification *conceptually* depends on various institutions (such as PADI, obviously) which are (*per*, e.g., Searle 2009) created by we-intentions. But my intention isn't *derived* from those we-intentions, nor is it *justified* in terms of them. (It is more likely justified in terms of my preferences, etc.) Thus, although dependent on a set of we-intentions, this is an I-intention, since the dependence is conceptual rather than justificatory.

4.2. Moral disagreement

While I will address some objections against my moderate collectivism account in Section 5, I want in this section to address specifically two interrelated worries about moral agreement raised by the above contextualist account.

One might worry that between various disputants in a moral argument, there is actually more disagreement than agreement. Does this undermine the above argument? I would say a number of things at this point. First, the extent of this disagreement should not be overstated. To be sure, there are deep and persistent moral disagreements, even within relatively culturally homogenous societies. But focus on these disagreements makes it easy to overlook the fact that these take place against a background of agreement. Disputants to this agreement generally go about their lives manifesting various commitments (all of which hold *prima facie*, of course)—promises should be kept, one should tell the truth, one ought not harm others, the autonomy of others is to be respected, etc.—over which there is no legitimate disagreement.

However, I do not wish to argue that in the case of morality, there is more agreement than disagreement.[17] I don't think my argument hangs on this claim, and I'm sure many readers cannot be convinced of it. Nevertheless, my account requires that there be some significant agreement; and this is where a critic will argue that the danger arises. There are two interconnected worries. First, surely there are individuals—and indeed entire *communities*—with moral practices radically different from our own. Do we really share sufficient commitments with them to avoid the incommensurability worries that the critic argues are raised by the moderate collectivist account? Second (and related), doesn't the present account lead to a kind of parochialism? The moral community is constituted by shared intentions, and so if (per the first objection) we encounter an individual or community who doesn't share sufficient moral intentions with us, must we regard them as outside of our moral community?

I will start with the first objection. Sellars's original, formalist picture doesn't seem to require more agreement than disagreement: What constitutes, for example, the moral community (and presumably thereby enables specifically *moral* agreement and disagreement) is commitment

[17] An anonymous referee for *JHAP* pressed me on this point.

to the founding intention of this community. Presumably, agents committed to this could disagree substantially in their derivative commitments, but would still count as disagreeing—and being part of the same moral community—because of their commitment to the founding intention, and not because they agreed more often than they disagreed. To use a very un-Sellarsian (and anti-contextualist) analogy, the foundation holding up a structure doesn't have to be the biggest part of the structure, but it is that which allows the rest of the structure to stand.

While I have argued that Sellars's original model is flawed, it offers one helpful suggestion here: Perhaps when it comes to morality, the *type* of agreement can be as important as the *amount*. In morality, a large part of what we look for is *second-order agreement* about the role of morality in our practical reasoning and in the formation of intentions and volitions. These second-order commitments about morality will in most cases be implicit—but that is the usual pragmatist order of explanation anyhow. These commitments are somewhat difficult to capture, but consider Horgan and Timmons's "Moral Twin Earth" thought experiment. (I will adapt their example somewhat for my present purposes.) Imagine that our use of moral terms is purely consequentialist. Imagine further that we encounter a "Moral Twin Earth," whose followers practice a specific deontological theory. Now, the first-order content of the Twin Earthers' morality might be very different from our own, and they have a very different metaethics to boot. Should we conclude that we aren't both practicing morality, and can't engage in moral dispute as we will simply talk past each other? Horgan and Timmons argue that this is clearly not the case: Both communities have what is obviously a moral practice, as can be seen by their *second-order commitments* about this practice:

> [T]he uses of these terms on Moral Twin Earth bear all of the "formal" marks that we take to characterize moral vocabulary and moral practice. In particular, the terms are used to reason about considerations bearing on Moral Twin Earthling well-being; Moral Twin Earthlings are normally disposed to act in certain ways corresponding to judgments about what is 'good' and 'right;' they normally take considerations about what is 'good' and 'right' to be especially important, even of overriding importance in most cases, in deciding what to do, and so on. (Horgan and Timmons 1990–1991, 459)

Of course, in interpreting another's practice as a moral practice, we also no doubt will look for overlap in first-order commitments. But

as I already argued, I think that in practice, there will be significant overlap in first-order moral commitments as well, particularly of the everyday variety. If we think about even people with whom we have significant disagreements—and here it is important to imagine a person with actual moral commitments, not a sociopath or an amoralist—I think it is actually quite difficult to sincerely believe (other than on ideological grounds) that the person has no moral commitments at all that overlap with some of your own (even if the commitments are very basic—e.g., "Hey, that person over there—it would be wrong to murder them for no reason").

Even if we encounter an extremist who thinks that it is permissible to hold slaves and kill some people whom we regard as innocent, we will still regard ourselves as having a *moral* dispute with this agent: We will offer a moral critique of their behavior (which we can support with reasons) just as they will no doubt offer a moral critique of ours (similarly backed by reasons, although we will regard most of these as spurious). Again, a large part of what we share are (implicit) second-order commitments about the role of morality in our practical reasoning and in the formation of intentions and volitions.

Further, I suspect that even in the case of people with whom we have significant moral disagreements, we probably still share a number of first-order moral commitments. As I have already pointed out, if someone has something we recognize as a moral practice, it is difficult to imagine we don't share a number of routine commitments with them—the sort that we act on unthinkingly in our daily lives. Even if we were to imagine some sort of extremist—someone who endorses slave holding and the killing of certain innocents—there is still a certain amount of first-order overlap (as distasteful as it might be to project ourselves into their moral world). For example, they probably agree that it is wrong to kill the innocent, but disagree as to who is innocent. And I suspect that even with extremists, there is much quotidian morality on which we agree (ordinary promise keeping, and truth telling, etc.). There will be "patchy" areas of overlap (where their commitments might align with yours but not mine, or mine but not yours): They may believe loyalty is a moral value. They may believe in caring for certain vulnerable populations. You might share some of these moral commitments—indeed, many of them. Again, this isn't to downplay

the seriousness of the places at which they morally go off the rails, but simply to point out that we are unlikely to encounter a culture or population that has anything we could recognize as a moral practice, but which practice had literally *no* areas of overlap with ours—the sort of overlap which would enable us to at least enter into discussion with them (however fruitless such discussion might turn out to be).

Let us now turn to the second objection raised at the start of Section 4.2. Although we don't seriously think it is permissible to enslave and kidnap people (for a suitable value of "we"), obviously this view has not been shared throughout history. And surely there is no form of life shared by all rational beings! How then can the moral community encompass all rational beings?

The first part of the response is contained in my reply to the first objection: Even communities with significantly different moral commitments will share second-order commitments about morality, as well as a number of first-order commitments. So we shouldn't overstate the possible reality of the imaginative example of two communities with radically incommensurable moral practices.

The second part of my response is more concessive. I think Sellars acknowledges more or less explicitly that the moral community—as it stands—does not encompass the community of all rational beings. For example, Sellars of course thinks that (assuming the premises in OIM §144 can be established) "the concept of oneself as a rational being implies the concept of oneself as a member of an ethical community consisting of all rational beings" (OIM, 185/§144). But he also acknowledges that this would only mean that "the ethical community of rational being would have an 'implicit' existence" (OIM, 185/§144)—a clear acknowledgement, I think, that not all rational beings would endorse the founding intention of morality or recognize its bindingness. The lesson we should take away from such passages is that the construction of the moral community is an ongoing project, and the task of the Kantian is to put pressure on our conception of this community in the direction of ever-more inclusion to fill the space of rational agents. As Rorty writes,

> *We* have to start from where *we* are—that is part of the force of Sellars's claim that we are under no obligations other than the "we-intentions" of the communities with which we identify. What takes the curse off this ethnocentrism is not that the largest such group is "humanity" or "all rational beings"—no one, I have been claiming, *can* make *that*

> identification—but, rather, that it is the ethnocentrism of a "we"...which is dedicated to enlarging itself, to creating an ever larger and more variegated *ethnos*. It is the "we" of the people who have been brought up to distrust ethnocentrism. (Rorty 1989, 198)

Gerald Postema, whose theory of public reasoning is genealogically downstream from Sellars (*via* Jay Rosenberg) writes in a similar vein:

> Morality requires transcendence to a common or public point of view...It is a mistake, however, to assume that this is equivalent to a commitment to a common point of view with universal scope. This is a mistake because it makes universal assent a condition of validity of norms and universal intelligibility a condition of legitimacy. It turns an aspiration into a precondition, and we have no good reason to assume that the precondition is satisfied. (Postema 1995, 62)[18]

The moral community consisting of all rational beings—and also the consensus that helps constitute that community—will probably always be to some degree aspirational. This conclusion seems like a realistic conclusion from theorizing about morality rather than an objection to Sellars's theory.[19]

5. Objections and Replies

Various authors have offered arguments against my collectivist reading of Sellarsian we-intentions, or against collectivist readings in general. Below, I will discuss the most prominent recent discussions (by Kyle Ferguson and Stefanie Dach).

5.1. Ferguson

Ferguson succinctly states one of my central arguments against individualistic understandings of we-intentions as follows:

> According to the challenge, "as one of us" presupposes a community in the first place, which suggests that a community, its founding we-intentions, and the sharing of many other we-intentions are presupposed

[18] I am grateful to Matthew Chrisman for directing me to this neglected essay.

[19] See also Matthew Chrisman's (forthcoming) fascinating reconstruction of how we should understand this idealizing move. His account, in my view, successfully avoids relativism and parochialism without making implausible assumptions about sharing practices with radically different, unknown, or alien communities.

> by an individual's unshared we-intentions. The idea is that even if there are maverick we-intentions, they are exceptional, parasitic on the standard case of shared we-intentions, and the "as one of us" that signals their intersubjective form either assumes this community or is empty without one. Individualism might be true for *this* or *that* unshared we-intention, but one can have we-intentions *in general* only against a backdrop of shared we-intentions. If this is right, then it reverses the order of explanation I advocated. . .and restricts to a very narrow scope the relevance and utility of the individualist interpretation. It becomes a theory of platypuses rather than a theory of mammals. (Ferguson 2023, 55–56)

Ferguson offers three replies to this objection. His first reply is (at first glance) somewhat concessive. He "acknowledge[s] that the independence with which one we-intends is located in an irreducibly social context" and concedes that "in order to have the capacities to intend and to we-intend, one must be a social actor of some kind and one must have been brought up by a linguistic community" (Ferguson 2023, 56). However, he resumes the offensive in his next two replies. In his second reply, Ferguson argues, "even if 'as one of us' presupposes this community, and even if the community's very existence presupposes a massive set of shared we-intentions, a prior account of intentions' shareability is needed to explain sharedness. Sharedness strikes me as an actualized potentiality, where the potentiality is shareability just as brokenness actualizes breakability. In both cases, the potentiality precedes the actuality" (Ferguson 2023, 56). In his third reply, Ferguson simply quotes Sellars in support of his individualism:

> The group has shared intentions by virtue of the fact that its members intend in the mode "shall*w*." But that the members intend in this mode does not guarantee that in point of fact there are shared intentions. . .Yet the actual existence of shared universal intentions is no more an antecedent condition of participating in moral discourse than actual agreement on matters of fact is an antecedent condition of participating in factual discourse. (IIOR, 413/§76)

The first reply we can ignore, as it expresses agreement with much of what I have argued above. I shall also set aside the third reply, because I think that Sellars actually repudiated this view by the middle of his career.[20] "Intentions, Imperatives, and the Logic of 'Ought' "

[20] I think, though, that the individualism of the early Sellars is easily overstated. *IIOR* §76 is worth reading in its entirety, as he says a number of things that are difficult to square

(IIOR)—published in 1965—represents some of Sellars's earlier work on morality, and lacks much of the sophistication of his later writings on the topic. As argued above in Section 3, by OIM (and certainly by ORAV), Sellars does actually hold that a "shared universal intention" is an "antecedent condition on participating on moral discourse"—the moral community is created by a founding intention that is *in fact shared*, and would not otherwise exist (much as the WCS would not exist without some shared founding intention or set of shared intentions).

This leaves us with Ferguson's second reply. There is something right about Ferguson's reply; but there is also a way in which it gets things the wrong way around. No doubt, at the foundation of human shared intentionality lies the basic cognitive apparatus of *shareable* mental states. It would be a mistake, however, to retreat too far into individualism and to treat these states as wholly self-contained building blocks for a theory of shared intentionality—as though we begin with shareable intentions, and then only contingently are these sometimes shared. Such a view cannot be correct. Consider, by analogy, Hans Bernhard Schmid's criticism of what he calls the "association-model of the constitution of the plural or collective intentional subject" (Schmid 2023, 44) such as is common in the social contract tradition, whereby people (*via* some communicative act) agree to form a plural intentional subject—a *we*—such as a community or society. As Schmid points out, such accounts are circular; they can only generate shared intentionality by presupposing it:

> For you to be able to propose your idea for a joint intentional venture to me, and for me to be able to agree to the idea, we obviously need to *communicate* with each other in some linguistic or proto-linguistic form. Our conversation, however—of which your proposal and my endorsement thereof are the constituents—is already a joint action and is thus sharedly, jointly, or collectively intentional. Shared intention can't be a product of conversation because conversation—whether or not it issues in a particular shared intention—already is a collectively intentional activity; conversing with each other is something we do *together*. (Schmid 2023, 45)

As Schmid points out, "It has come to be a powerful idea to think of what we tell each other as something coming from an otherwise inaccessible

with a straightforward individualist interpretation. I will set this aside, however, as I am content to focus on Sellars's later works.

inner sphere that we transmit to each other in linguistic communication" (Schmid 2023, 206). However, this view misunderstands the way in which we are constituted as joint intentional subjects *simply in virtue of being practical rational agents*: "We are not primarily objects in each other's worlds: we share our subjectivity and are thus not closed off against each other like monads, but as plural minds with perceptions, evaluations, cognitions, and intentions the intentional subject of which is us, *together*, rather than you [and me] separately" (Schmid 2023, 206). One can hear in this echoes of Charles Taylor's (1992) Hegelian criticisms of atomism—the view (associated with thinkers like Hobbes and others in the liberal tradition) that we can exist as individuals prior to entry into society, and can make a rational decision as to whether to join society and be subject to its rules and mores. Atomism is the myth that participation in Spirit is something that we can choose—that we have the capacity for thought and choice prior to such participation.

From this point of view, we can see the danger of embracing an overly-individualistic account of we-intentions: It treats us first as individuals who are practical rational agents before sharing collective intentions—rational agents who then (contingently) can decide to join with others into community. But of course this is a pre-Hegelian myth: We are only able to be practical rational agents because we are *already* a part of Spirit, because we have internalized collective social structures (norms, functions, social roles, etc.) and think and live in terms of them. To be sure, this internalization is not all-engulfing, and we can think and intend *qua* individuals—but such thinking and intending presupposes collective thought and intending.

It is true that we need to make sense of the possibility of unshared we-intentions. But the individualist draws the wrong lesson from this. She takes unshared we-intentions as the paradigm or reference case—the starting point from which we construct our account of shared we-intentions. But this is wrong. Unshared we-intentions are only *we*-intentions because our starting point is shared we-intentions. That is the paradigm case; and we explain unshared we-intentions by reference to shared ones.

I have elsewhere offered a detailed account of what it is to have a we-commitment—and to be *entitled* to that we-commitment—that is

unshared by the larger group.[21] I shan't recapitulate the details here, but the crucial point is that the "we-ness" of such a commitment essentially depends on (a) the existence of a group ethos, and (b) the commitment being appropriately based on this ethos. The concept of an ethos is borrowed from Tuomela, who defines it as follows: "The *ethos* of a group g in its strict sense is defined as the set of *constitutive* goals, values, beliefs, standards, norms, practices, and/or traditions that give the group motivating reasons for action" (Tuomela 2007, 16). Thus, while it is possible to have an unshared we-attitude, such unshared we-attitudes are parasitic on a larger stock of common ground, and impossible without it. This is, I argue, the best way to understand the possibility of unshared we-attitudes. And, further, I think it is fairly clear that for late Sellars, individual moral intentions *in general*—both shared and idiosyncratic—presuppose shared intrinsically valid we-intentions, and are thus parasitic on the latter. For these reasons, we should embrace a moderate collectivist reading of Sellars on moral we-intentions.

5.2. Dach

Dach raises a somewhat different objection against my collectivist reading of Sellarsian we-intentions. Dach writes, "I disagree with Koons' collectivist interpretation of Sellars's we-intentions (Dach 2023, 96), according to which we-intentions which are not shared are deviant. . ..Sellars claims clearly, however, that intersubjective intentions 'need not be chorused' (CSDS, 734/§9). For Sellars, the norm against which we evaluate such intentions is not sharedness but objective reasonableness" (Dach 2023, 32–33, n 30).

I agree that we-intentions "need not be chorused," although I have urged the reader not to treat this fact as an indication that unshared we-intentions are first in the order of explanation. The serious objection, however, is that we evaluate we-intentions against the norm of objective reasonableness, not "sharedness." Collectivist accounts of we-intentions are (Dach holds) committed to the latter. According to Dach, there is nothing deficient about an unshared we-intention; the primary way in which a we-intention can be deficient in Sellars's framework is that it

[21] See Koons (2022, sec. V, especially V.3).

is not categorically reasonable, i.e., not derivable from the intrinsically reasonable intention.

I will argue that we can *both* (a) agree that objective reasonableness is the evaluative norm for we-intentions *and* (b) maintain a moderate collectivist view of we-intentions. The key to this reconciliation lies in remembering that in judging, we are always judging by *our* standards, using *our* conceptual scheme. One is reminded of Hilary Putnam's famous quip, "We should use somebody else's conceptual scheme?" Thus, the very standards we use to judge the reasonableness of an intention, or of a bit of conduct, are provided by our shared social practice. Thus, if I judge that a practice is immoral, I am ultimately judging that it falls short according to some shared standard or other.[22] As I noted earlier, to even count as making moral judgments—ones that could accord or conflict with other moral judgments—we must share in a common *form of life*. Wittgenstein's point is, I think, widely-accepted and largely uncontroversial.

This is true even if my judgment is not itself widely-shared. It is a platitude about moral claims that they have a tight connection with reasons, and we expect that one can (all things being equal) offer some kind of inferential justification for one's (non-basic) moral commitments. Thus, if asked to justify my moral commitment

Shall_{we} (I do not eat animals)

I might say, "Animals have interests" or "Animals feel pain" or even just "Animals have rights." In engaging in the game of giving and asking for reasons, we generally seek to establish *common ground*. And the most common strategy for doing so is to search for pre-existing common ground—to search for reasons we already share. For example, I may think you are already committed to—or can easily be convinced of the claim—that there is a strong connection between interests and rights.

[22]Stefanie Dach has made the plausible suggestion (personal correspondence) that such judgments are (or should be) accompanied by a kind of meta-awareness of these standards as "merely" our standards which we might improve upon at some time. Thus (and Dach may not be committed to all of the following), a future science might tell us better ways of promoting our welfare. Or an encounter with another culture—or simple moral experience, reflection, and the evolution of society—might lead us to realize certain of our moral principles need revision. For Sellars, nothing is *de jure* unrevisable, not even our standards of moral judgment. This is key to his account of the objectivity of such judgments, an issue I address at greater length in Section 6.

We may have to dig down a bit before we find this level of shared attitudes, but whether I know it or not, when I act on a commitment like the one above—one that is not shared by the rest of the group—I am in fact promoting what I take to be the group ethos, in that I am promoting what I take it other members of the group ought to be committed to (and, I think Sellars would say, are *in fact* committed to) by the reasons underpinning my commitment, reasons that we share. (In a similar connection Durkheim—whose work influenced Sellars's metaethical views[23]—wrote, "A rebellion against the traditional morality you conceive of as a revolt of the individual against the collective. . .However, what I am opposing to the collective is the collective itself, but more and better aware of itself" (Durkheim 1965, 66).)

To be sure, our moral attitudes are also answerable to the world—to facts about biology, human and animal flourishing, resource availability and distribution, the effects of poverty, and so on. But of course for Sellars there can be no moral Given—no fact can have moral import apart from a moral practice with a system of moral concepts and norms. Thus, when we appeal to these facts ("Animals feel pain," "Severe poverty can limit cognitive development," etc.) to justify a moral judgment (even an unshared one), we are doing so by appealing to some set of shared moral norms or standards.

6. Objectivity

Can my moderate collectivist theory offer a sufficient account of the objectivity of moral norms? There is a *prima facie* tension between understanding correctness in terms of sharedness, but also in terms of the norm of objective reasonableness. For Sellars, the objectivity of morality requires that there be a "decision procedure" (OIM, 182/§133) for deciding on categorically reasonable moral intentions. But, as Olen and Turner write, "for an adherent of the idea that ethical truths rest on collective intentions, there is a tension: the decision procedure cannot merely be the product of a particular community's standards and collective intentions, but it nevertheless must be grounded in the collective intentions of a community in order to be ethical at all" (Olen

[23]See Olen and Turner (2015, 2016).

and Turner 2016, 2067–68). Fully responding to Dach's objection from Section 5.2 requires resolving this apparent tension.

Sellars's answer (according to Olen and Turner) is that objectivity is fixed by consensus in the ideal community of rational beings. They write (2016, 2069) that in such a community, there would be a convergence between what is categorically reasonable and what each member of the community actually intends—hence, agreement and reasonableness would ultimately coincide.

This response is ultimately unsatisfying. While it is reminiscent of Sellars's neo-Peirceian conception of truth wherein truth is defined in terms of the CSP—the "Peirceish conceptual structure" (*SM*, Chapter 5), it seems like a departure both from Sellars's naturalism and from his grounding of morality in actually existing shared intentions. As Olen and Turner write, "the result is not of this world. It is a result in principle only, and not merely a result that erects a moral standard that is not of this world, but a standard in a community not of this world" (Olen and Turner 2016, 2069).

I think, nevertheless, a more plausible reading of Sellars's view is available; and it is a reading suggested by Sellars himself, when he writes,

> Alas, the ideal knowledge of our philosophical fiction is not even close at hand. Thus, the ideal 'consensus' of those who share the moral point of view is only 'in principle' there, and reasoning from the moral point of view proceeds in a context of ignorance and diversity of opinion. But, then, the same is true of consensus on matters of fact, scientific laws and theoretical principles. There, as here, the philosophical task must be to exhibit the conceptual structure within which this ignorance and this difference of opinion exist, and which, by making rational inquiry possible, provides the means by which (in principle, alas!) they can be overcome. (OIM, 183/§135)

"Coherence with the objectively reasonable intentions of an ideal ethical community" is—like many regulative ideals, including "objective truth"—the sort of thing that one can never be sure of having reached. If one is striving to reach this ideal, the only standards one has to measure one's progress are the current standards of the existing community of enquirers (barring the emendation of these standards). And when we judge that a particular belief, standard, or intention is or is not objectively true or objectively reasonable, we do so *qua* members of

the larger community, relying on commitments and entitlements that are communal as well. If I argue that your refusal to allow workers to organize violates basic tenets of justice and human rights, surely these are not concepts that I invented out of thin air; nor are these commitments I just happen to have *qua* individual rational agent (rather than *qua* member of an ongoing community with an underlying—if perhaps severely inadequately expressed—commitment to these values).

Such an account raises the specter of relativism—the worry that if we vest standards of correctness in the community, then the community is "*globally* privileged" and we therefore "cannot find room for the possibility of error."[24] But here we must be careful. Brandom, for example, is correct in cautioning against simply identifying normative truth "with what is taken true by all the members of a community, or by the experts in a community, or what will always be taken true by them, or what would be taken true by them under some ideal conditions for inquiry" (Brandom 1994, 600). For Brandom, objectivity is more of an orientation—it is a commitment among interlocutors that there is a difference between what is objectively correct and what is taken to be correct. (Brandom 1994, 600)

Treating objectivity as an orientation, essentially tied to fallibilism, doesn't change the essential fact that each interlocutor must reason from her own communally-grounded standards; and these standards of reason and rationality are tied to agreement and collective acceptance in the ways I have outlined under the heading of "moderate contextualism." Thus, standards of correctness are inherently tied to sharedness, in that all standards can only arise out of and be grounded in an ongoing social practice. And frankly, it's not obvious what the alternative is supposed to be. We are talking about standards of rationality—what is reasonable—and it is not clear how these could exist *in rerum natura*, apart from a functioning social practice. (I imagine it helps that for Sellars, of course, the relevant community is maximally inclusive.) But this tie to sharedness is elucidated in a way that preserves objectivity; such an account is not inherently relativistic, as it has an intrinsic tie with fallibilism: For Sellars, a necessary condition on the rationality of a practice is that it be open to revision. Thus, his famous dictum, "Empirical knowledge, like its sophisticated extension, science, is rational, not because it has a

[24] This is Brandom's criticism of I-we objectivity. See Brandom (1994, 599).

foundation but because it is a self-correcting enterprise which can put *any* claim in jeopardy, though not *all* at once" (*EPM*, 157/§113). No claim or standard of the community is immune to revision—that is the fallibilism that runs through virtually all contemporary pragmatism.

What about a social practice that did not embrace fallibilism, that did treat some claims as immune to revision?[25] Are those claims then "objectively rational" for that community? The community may think so—but *we* can recognize that they are wrong. Fallibilism is, as it were, a condition on the possibility of objective rationality—but it can only be seen as such from within a social practice that so recognizes it. There is no view from outside of all social practices. And so again, there is no divorcing judgments of objective rationality from certain basic shared attitudes constituting an actual social practice. Although the relationship between them is complex, sharedness and objective rationality are not entirely separate standards. Thus, worries about objectivity do not stand in the way of a moderate collectivist reading of we-intentions.[26]

This latter paragraph brings out a final point about social practice accounts and relativism. Although our practice might not recognize a claim or standard, it does not follow that another practice might not recognize (and be able to justify) this claim or standard, and therefore offer a critique of our practice, again undermining concerns about relativism and parochialism. While worries about incommensurability are in the offing, it strikes me as implausible (I don't want to say *in principle* impossible, although perhaps this too, for Davidsonian reasons) that the actions of another group of rational beings are so unintelligible that we cannot recognize *anything at all* they do as reason giving for us—that we cannot understand what they are doing and cannot carry on

[25] Stefanie Dach has suggested to me that some claims do seem immune to revision for Sellars—for example, our intrinsically categorically reasonable intentions. But I don't think Sellars does treat them as unrevisable. For example, as I argue in (forthcoming) Sellars's himself revises his supreme principle of morality: For most of his career it has a distinctly rule utilitarian flavor, but by *ORAV* it is much less utilitarian and much more Kantian. And in any case, unrevisability would directly contradict Sellars's stated commitment from *EPM* §113.

[26] Arguably, another element of objectivity is that our commitments are answerable to the way things really are, independent of our attitudes. Brandom takes this to be a condition on the objectivity of "specifically *representational* correctness" (Brandom 1994, 78). For Sellars, this might be considered a condition on the objectivity of linguistic practices that picture the world. It is not clear that our moral discourse pictures the world. However, a discussion of this issue would take us too far afield.

a dialogue with them and exchange reasons with them. Granted, there is no "view from nowhere" to banish the specter of relativism—but the whole point of a social practice account is to call off the search for such a view.

7. Conclusion

Recent commentary on Sellarsian we-intentions has often stressed Sellars's point that an agent's we-intention need not be shared. I have urged that this should be seen as a data point that must be accommodated within a theory of we-intentions, rather than an invitation to develop our entire theory around this claim. Our theory should of course be able to account for exceptional cases. But (to return to Ferguson's analogy), I argue that looking at non-shared cases of we-intentions and arguing for individualism is like looking at platypuses and echidnas and defining mammals in terms of egg-laying; it shows an improper choice of reference or paradigm case. I argue that at least from the publication of *Science and Metaphysics*, Sellars had clear collectivist commitments. Further, I argue that a moderate collectivist reading of we-intentions is philosophically superior to an individualist one. Crucially, the theory outlined above allows us to accommodate the exceptional cases (unshared we-intentions) and explain their exceptional character without explaining them away.

More needs to be said on a number of topics touched on here. For example, I gave somewhat short shrift to the topic of objectivity. Frankly, however, I see no plausible alternative to the theory that standards of rationality and reasonableness are rooted in actual social practice; questions of how to reconcile this with objectivity are not new, and have been a topic of philosophical conversation for decades. This doesn't, therefore, represent a particular Achilles' heel of Sellars's theory, if read in a moderate collectivist way.

Acknowledgements

I am grateful to Stefanie Dach for helpful comments on an early draft of this manuscript that led to refinement of the thesis and a number of

overall improvements in the argument. Two anonymous referees for *JHAP* also provided valuable feedback on a subsequent draft.

References

Brandom, Robert. 1994. *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Cambridge, Massachusetts: Harvard University Press.

Chrisman, Matthew. n.d. "Intersubjectivity in Practical and Empirical Cognition: A Discussion of Sellars' Moral Psychology." In *Collective Intentionality, Normativity, and the Space of Reasons: Engaging Wilfrid Sellars*, edited by Jeremy Randel Koons and Ronald Loeffler. New York: Springer.

Dach, Stefanie. 2023. "Sellars on the Intersubjectivity of 'We-Intentions'." In *Ethics, Practical Reasoning, Agency: Wilfrid Sellars's Practical Philosophy*, edited by Jeremy Randel Koons and Ronald Loeffler, 15–35. New York: Routledge.

Durkheim, Émile. 1965. *Sociology and Philosophy*. Translated by D. F. Pocock. London: Cohen & West Ltd.

Ferguson, Kyle. 2023. "We-Intentions and How One Reports Them." In *Ethics, Practical Reasoning, Agency: Wilfrid Sellars's Practical Philosophy*, edited by Jeremy Randel Koons and Ronald Loeffler, 36–61. New York: Routledge.

Hicks, Michael R. 2023. "Moral Motivation: The Practical Philosophy of Wilfrid Sellars. A Critical Notice of: Ethics, Practical Reasoning, Agency: Wilfrid Sellars's Practical Philosophy." *International Journal of Philosophical Studies* 31: 718–29.

Horgan, Terence, and Mark Timmons. 1990–1991. "New Wave Moral Realism Meets Moral Twin Earth." *Journal of Philosophical Research* 16: 447–65.

Koons, Jeremy Randel. 2019. *The Ethics of Wilfrid Sellars*. New York: Routledge.

———. 2022. "Knowledge as a Collective Status." *Analytic Philosophy* 63: 277–304.

Koons, Jeremy Randel, and Ronald Loeffler. 2023. *Ethics, Practical Reasoning, Agency: Wilfrid Sellars's Practical Philosophy*. New York: Routledge.

———, eds. n.d. "Taking One for the Team: Morality as a Species of Principled Team Reasoning." In *Collective Intentionality, Normativity, and the Space of Reasons: Engaging Wilfrid Sellars*. New York: Routledge.

Loeffler, Ronald. 2020. "Sellars' Theory of We-Intentions and Gilbert's Theory of Joint Commitment." *International Journal of Philosophical Studies* 28: 114–27.

———. 2023. "Moral We-Intentions as Individualistic We-Attitudes." In *Ethics, Practical Reasoning, Agency: Wilfrid Sellars's Practical Philosophy*, edited by Jeremy Randel Koons and Ronald Loeffler, 62–83. New York: Routledge.

Nguyen, C. Thi. 2018. "Escape the Echo Chamber." *Aeon*. https://aeon.co/essays/why-its-as-hard-to-escape-an-echo-chamber-as-it-is-to-flee-a-cult.

Olen, Peter, and Stephen P. Turner. 2015. "Durkheim, Sellars, and the Origins of Collective Intentionality." *British Journal for the History of Philosophy* 23: 954–75.

———. 2016. "Was Sellars an Error Theorist?" *Synthese* 193: 2053–75.

Postema, Gerald. 1995. "Public Practical Reason: An Archaeology." *Social Philosophy and Public Policy* 12: 43–86.

Rorty, Richard. 1989. *Contingency, Irony, and Solidarity*. Cambridge, MA: Cambridge University Press.

Schmid, Hans Bernhard. 2023. *We, Together: The Social Ontology of Us*. Oxford: Oxford University Press.

Schönbaumsfeld, Genia. 2016. *The Illusion of Doubt*. Oxford: Oxford University Press.

Searle, John. 2009. *Making the Social World: The Structure of Human Civilization*. New York: Oxford University Press.

Sellars, Wilfrid. 1960. "Science and Ethics." In *The Metaphysics of Practice: Writings on Action, Community, and Obligation*, 97–115. [Cited as SE].

———. 1963. *Science, Perception, and Reality*. Atascadero, California: Ridgeview Publishing Company. [Cited as *SPR*].

———. 1967. *Science and Metaphysics: Variations on Kantian Themes, The John Locke Lectures for 1965–6*. Atascadero, CA: Ridgeview Publishing Company. [Cited as *SM*].

———. 1991a. *Empiricism and the Philosophy of Mind*. In *Science, Perception, and Reality*, 127–96. [Cited as *EPM*].

———. 1991b. "Is There a Synthetic A Priori?" In *Science, Perception, and Reality*, 298–320. [Cited as ITSA].

———. 1991c. "Some Reflections on Language Games." In *Science, Perception, and Reality*, 321–58. [Cited as SRLG].

———. 2023a. "Imperatives, Intentions, and the Logic of 'Ought'." In *The Metaphysics of Practice: Writings on Action, Community, and Obligation*, 367–420. [Cited as IIOR].

———. 2023b. "Obligation, Intersubjectivity and the Moral Point of View." In *The Metaphysics of Practice: Writings on Action, Community, and Obligation*, 138–87. [Cited as OIM].

———. 2023c. "On Reasoning About Values." In *The Metaphysics of Practice: Writings on Action, Community, and Obligation*, 465–510. [Cited as ORAV].

———. 2023d. "Sellars to Solomon, 28 June 1976." In *The Metaphysics of Practice: Writings on Action, Community, and Obligation*, 732–36. [Cited as CSDS].

———. 2023e. *The Metaphysics of Practice: Writings on Action, Community, and Obligation*. Edited by Kyle Ferguson and Jeremy Randel Koons. Oxford: Oxford University Press.

Taylor, Charles. 1992. "Atomism." In *Communitarianism and Individualism*, edited by Shlomo Avineri and Avner de-Shalit, 29–50. Oxford: Oxford University Press.

Timmons, Mark. 1999. *Morality Without Foundations: A Defense of Ethical Contextualism*. Oxford: Oxford University Press.

Tuomela, Raimo. 2007. *The Philosophy of Sociality*. Oxford: Oxford University Press.

Journal for the History of Analytical Philosophy

VOLUME 15, NUMBER 6 (2026)

ISSN 2159-0303

jhaponline.org