Introduction: Situating Sellars's Ethical Theory in the Contemporary Landscape

In his early overview of Sellars's ethical theory, David Solomon situates Sellars's theory within what Solomon calls the tradition of 'classical metaethics' (CM). According to Solomon, the battle lines in CM are drawn roughly as follows: Cognitivists accepted that there is a logical relation between statements of fact (e.g., evidence) and moral judgments, but "denied, however, that any strong logical tie holds between" moral judgment and action (Solomon 1977, p. 153). Non-cognitivists, on the other hand, hold that there is a strong connection between moral judgment and action, but also hold that "a strong logical relation between [moral judgments and action would] preclude such a relation between [statements of fact and moral judgments]" (Solomon 1977, p. 153).

Sellars is grappling with these fundamental (and intransigent) issues in his earliest writings on ethical theory. In IILO, Sellars expresses the motivating element of moral judgment in the following terms: "to know that there are certain things that one ought to do *is* to have a sense of duty," (IILO, p. 160), a view which he claims "smacks of emotivism" (IILO, p. 160). Thus, Sellars attributes to Ayer the view that "The so-called thought that one ought to do A here and now is not, strictly speaking, a thought at all, but rather a specific way of being moved to do A" (IILO, p. 161). This, of course, captures what Sellars takes to be the chief weakness of emotivism: while it captures the connection between moral judgment and moral motivation, it fails to capture the cognitive element of moral judgment, holding instead "that ethical concepts are 'pseudo-concepts' and the logic of moral discourse is a 'pseudo-logic'" (IILO, p. 162). As Ayer himself happily concedes, if emotivism is true, then opposing ethical judgments cannot, strictly speaking, be incompatible with each other: "It is plain that the conclusion that it is impossible to dispute about questions of value follows form our theory…For as we hold that 'Thrift is a virtue' and 'Thrift is a vice' do not express propositions

at all, we clearly cannot hold that they express incompatible propositions" (Ayer 1946, p. 110).[1]

Sellars expresses the cognitive element of moral judgment by writing that "obligation, by its very nature, is intersubjective" (IILO, p. 160). And, of course—the Achilles heel of emotivism—Sellars is committed to the validity and soundness of certain inferences essentially involving moral premises.[2]

This intersubjective element of moral judgment, according to Sellars, "smacks…of intuitionism" (IILO, p. 160). However, Sellars is also unhappy with intuitionism, as the dominant forms rescue the cognitive element of moral judgment at the expense of explaining the connection between moral judgment and moral motivation:

> It was the signal merit of intuitionism…to have insisted on…the truly propositional character of prescriptive statements…But the epistemological and metaphysical commitments of ethical intuitionism, which precluded it from understanding the logical connection between 'thinking that one ought' and 'being moved to do,' thus forced it to make a mystery both of the conduct-guiding role of moral discourse, and of the uniqueness of prescriptive discourse which it had so happily emphasized. (IILO, p. 162).

For example, Prichard (who merits critical mention in the early pages if IILO), in rejecting the Kantian view that the *knowledge* that an action is right can influence us to act, writes, "we are, I think…bound to abandon this view. For one reason...if we face the purely general question, 'Can we really do anything whatever unless in some respect or other we desire to do it?' we have to answer 'No'" (Prichard 2002, p. 38).

The task for Sellars, then, is to reconcile these two seemingly incompatible elements of moral judgment. Thus, he must do justice to what the cognitivists recognize as important about moral judgments: they are intersubjective; they are in the space of reasons, and admit of (and require!) evidential support; they can stand in logical relations with other propositions; and so on. But he must also do justice to what the non-cognitivists want to preserve about moral judgments: moral judgments are fundamentally connected to action. *Ceteris paribus*, one who judges that ϕ is obligatory is motivated (to some degree) to do ϕ.

Sellars's solution to this problem is powerful and ingenious. In the first instance, he argues that moral judgments express intentions, and are thus conceptually connected to action. In arguing that moral judgments express a certain kind of intention, Sellars's theory is a very early model of the expressivism of plans and intentions perhaps most famously associated with Gibbard—although few philosophers working in this tradition know of Sellars's early contributions to this theory.

However, ordinary expressions of intention cannot, strictly speaking, contradict each other; nor can people share expressions of intention in ordinary cases. (If I intend to order vanilla ice cream, and you intend to order chocolate, we are not disagreeing; nor is there any conflict or contradiction between our intentions.) To solve this problem—to explain how moral judgments are genuine intersubjective—Sellars develops his account of we-intentions. Although the notion of we-intentions has been developed in sophisticated ways by Bratman, Gilbert, Searle, Tuomela, and others, Sellars's is (as far as I can determine) the first strictly philosophical account of the notion. [3]

So much for the connection to action. How does Sellars account for the cognitive aspect of moral judgment? Anticipating Bratman's work, Sellars demonstrates that we can and do reason among intentions, and develops a sophisticated logic for his expressivism that prefigures a similar logic developed—50 years later—by Mark Schroeder.

A moral theory must accomplish much more than this, of course, and Sellars is careful to elaborate his moral theory in a way that satisfies a number of constraints. Because Sellars is such a systematic philosopher, he is able to offer an ethical theory that is integrated with and motivated by his larger philosophical project. His larger philosophical goal is to reconcile— *via* a thorough-going nominalism—normativity (and the framework of persons more generally) with a strongly scientific world-view. Thus, Sellars's moral theory is constructed within a strictly naturalist philosophical framework.

The systematic nature of Sellars's work, combined with his deep familiarity with the perennial metaethical issues and problems that confront us just as they confronted Sellars's contemporaries (such as Stevens and Hare), enable him to offer a theory that offers powerful and convincing solutions to a number of outstanding philosophical problems. As already noted, his view of moral judgments as expressions of we-intentions (combined with his sophisticated logic of intentions) allows him to explain how moral judgments can be both cognitive and motivating. He offers a subtle account of reasoning—centered around material inference—that offers us insights on practical reasoning, and also on how judgment is connected to action. He offers a version of constitutivism—the theory that moral requirements stem from the nature of practical agency itself—that has a Hegelian/social dimension, allowing him to explain the necessity of the moral community, and the categorical validity of moral requirements. He offers a moral theory that reconciles consequentialist and deontological theories, filling in some of the problematic gaps in Kantian moral theory while managing to preserve key Kantian insights (such as the idea that the moral realm is the realm of autonomy as opposed to heteronomy). And this barely scratches the surface of the richness of Sellars's account, and what it has to offer. Further, all of this is developed from within a nominalism stricter even than Quine's, one that allows Sellars to articulate a vision of the moral community consistent with the strictest naturalism.

This isn't to say, of course, that Sellars's own ethical writings need no updating or supplementing. There are tensions between Sellars's writings in ethics and some of his other writings. Also, there are elements of Sellars's theory that can be enriched by contemporary philosophical work—work by later Sellarsian philosophers, by philosophers working on group intentions and cooperative rationality, and more. In resolving these tensions, and updating Sellars's theory with more recent work in the field, what emerges is a strikingly original and comprehensive theory that has much to contribute to contemporary debates. The goal of this

book is to explain Sellars's theory, and to develop it into one that can—and should—be taken seriously by contemporary theorists.

A note on methodology: There is a certain inevitable hubris inherent in trying to improve upon the work of one's philosophical betters, and so I have attempted to be as conservative as possible in the modifications I make to Sellars's own theory. The modifications I make fall into three classes, in ascending order of radicalness.

First, there are simple additions I make in light of more recent work on the subject matter which do not, in my view, contradict anything Sellars himself says. For example, although Sellars offers (as I noted above) the first philosophical analysis of we-intentions, a robust literature on the subject has developed since Sellars's death, and I supplement Sellars's account by drawing on more recent accounts.

Second, there are amendments I make in light of what I take to be inconsistencies in Sellars's own view. In such cases, I make a judgment (which I hope the reader shares) as to which element is more central to Sellars's overall project. For example, a deep and central commitment of Sellars's (and of later Sellarsians', like Brandom) is a rejection of formalism about reason. That is, Sellars argues that we have to understand materially valid inference (e.g., "Fido is a dog; therefore, Fido is a mammal") as prior in the order of explanation to deductively valid formal inferences. However, I will argue in Chapter 14 that Sellars embraces this formalism in certain aspects of his ethical theory. Given the absolute centrality of anti-formalism to the rest of his philosophical project, the formalism in his ethical theory must give way—but in a way that preserves the essential features of his ethical theory, because the formalism served a purpose there that cannot simply be abandoned. Thus, I attempt to make maximally conservative revisions.

Finally—and most seriously—there are elements of Sellars's theory that I revise because they seem unsustainable in light of external criticisms or more recent developments in

the literature. Pursuant to my conservatism, I try whenever possible to avoid these kinds of revisions, but there are places where they cannot be avoided. For example, I argue in Chapter 5 that certain elements of Sellars's logic of intentions must be revised because his logic, as he presents it, cannot construct some core deontic notions, like permissibility.

In every case, though, I first explain Sellars's original view, before going on to supplement or modify it. Thus, although this book is not merely exegetical, I always begin at the source before going beyond it. The reader may not agree with all of my decisions, but fruitful philosophical discussion usually begins in disagreement.

No work on a topic such as this can hope to be the last word on the subject, or to be in any way the definitive work on the matter; and it would be unforgivable hubris to have this as one's goal in writing about Wilfrid Sellars's ethical theory. Not every reader will agree across the board with my reading of Sellars. Nor will every reader agree with my decisions regarding how to reconcile what I take to be inconsistent elements in Sellars's theory, or how to supplement and extend Sellars's theory using the philosophical tools that were developed after his death. Further, because Sellars's ethical theory is so deep and rich, there are a number of topics that would welcome more in-depth treatment. Indeed, I'm sure that some of the chapters of this work could admit of book-length treatments, or at least of more detailed discussion than I am able to offer here. There is much more to be said, and much more work to be done.

But if I succeed in nothing else, I hope to convince the reader of at least this: Sellars's ethical theory is subtle and profound, and is worth serious consideration—not merely as an object of historical interest, but as a flourishing research project that can be brought into fruitful dialogue with contemporary accounts, that can enrich them and be enriched by them. In light of this, the recent surge of interest in Sellars's ethical writings can only be seen as a good thing.

Chapter Summaries

On the assumption that most philosophers interested in ethics are not conversant with Sellars's larger philosophical project, Chapter 1 ("Sellars's Synoptic Vision") gives a very brief overview of some of his chief aims and methods. Sellars's chief philosophical goal is to reconcile the framework of normativity (and of persons more generally) with a hard-nosed scientific realism. His method for doing this is to understand distinctly philosophical expressions—moral, epistemic, intentional, modal, semantic, and so on—not as attempting to describe the world, but rather as non-descriptive expressions in the metalanguage. Readers who are already conversant with Sellars's overall project may wish to skim this chapter, or skip ahead to Chapter 2.

In Chapter 2 I turn to Sellars's claim that the normative is causally reducible, but logically irreducible to the natural. This means that while (for example) the behavioral or neurological information *conveyed* by a belief-ascription can in principle be wholly captured by an (ideal) scientific theory, such a theory wouldn't *say* what intentional language said (in that it wouldn't ascribe beliefs per se). What normative expressions *say*, can only be *said* using normative language, even if what they *convey* can be reduced to empirical psychology, or neurology, or some other branch of science. Serious objections have been raised against the tenability of Sellars thesis of the "causal reducibility, but logical irreducibility" of the normative. Just as critically, Sellars's powerful argument for scientific realism can be turned against his argument for the causal reducibility of the normative, so that the normative turns out to be both logically *and* causally irreducible. I will demonstrate that this conclusion leads to a more complicated naturalism, but one that is still compatible with Sellars's austere nominalism. The thesis of causal irreducibility will turn out to be helpful in explaining, among other things, moral motivation (as we will see in chapter 10).

In Chapter 3, I turn a core element of Sellars's ethical theory, the key to his reconciliation of the cognitive and motivating aspects of moral judgment. Sellars argues that

moral judgments are expressions of intention. This not only aids in this reconciliation—intentions are conceptually tied to action, and are also involved in reasoning—but it is independently plausible. I try to motivate Sellars's position by considering the role of intentions in practical reasoning, as well as considering illuminating parallels with theoretical reasoning. Central to Sellars's account of moral judgment is a special kind of intention—a we-intention. I show how Sellars's account has the resources to make sense of phenomena like restricted-scope or loyalty-based obligations (i.e., special obligations I have qua member of some subset of rational humanity, such as qua member of a union or a family).

Sellars does too little to explain what, precisely, a we-intention is. In Chapter 4, I draw on recent accounts (like those of Gilbert and Tuomela) to enrich the Sellarsian account of we-intentions. The resulting account rejects *singularism*—it denies that group intentions are reducible to personal intentions—but argues that this move is consistent Sellars's strict naturalism. Various elements of these more recent accounts are adapted to comport with various other features of Sellars's project, such as his pragmatism and his view that an individual can express a we-intention not shared by the larger group. I conclude with some remarks on how Sellars's account, while lending itself easily to a social practice account of rules and norms, offers a defense against relativism.

Sellars's account of moral judgment is broadly expressivist; and such accounts face well-known problems such as the embedding problem. Sellars develops a very sophisticated logic of intentions, which foreshadows similar attempts by authors like Mark Schroeder by half a century. In Chapter 5, I explain the key features of Sellars's logic of intentions, and demonstrate how it provides us the tools to solve not only classical problems for expressivism (such as the embedding problem and the problem of negation), but also contemporary challenges to expressivist accounts (such as Schroeder's challenge that such accounts cannot handle tense and modal operators). I argue that while the basic insights of Sellars's account

are sound, various modifications and extensions are required to his account so that we can have a fully satisfying intentional logic, one that is able to reconstruct all of the required deontic notions and relations.

Sellars presents practical reasoning as based on *nomologicals* of the form, "Doing A_i if C_j is causally necessary to the realization of E_k" (SM 7.VIII.§51/p. 195). This, however, misrepresents practical reasoning, as there is often more than one way of achieving an end; and hence, the performance of no *one* means is in general a *necessary* condition on the achievement of one's end. In Chapter 6, I introduce the notion of a material inference, which is central to Sellars's philosophical project, and argue that the notion of a *material practical inference* gives us a more flexible model of practical reasoning. This, combined with Bratman's notion of intentions as elements of partial plans, allows us to develop an account of practical reasoning which (a) more accurately represents practical reasoning (and the role of intention in practical reasoning) than does Sellars's original, extremely rigid nomological-based model; and (b) is consistent with other elements of Sellars's philosophy (such as his views on theoretical or causal-explanatory material inference, and on the role and status of desires).

Many discussions of rationality start with the assumption that all rationality must be conceived of as individual: individual agents are acting on intentions with contents that might at most be parallel (in Sellars' terminology), but cannot be shared. Drawing on the work of Gilbert, Sugden, Tuomela, and others, I argue in Chapter 7 that there is a notion of cooperative rationality, which is not reducible to individual rationality; and that the moral point of view can only be justified from the standpoint of cooperative rationality—not individual rationality. Further, Sellars's account of we-intentions goes hand-in-hand with an account of cooperative rationality to present a cohesive picture of the moral point of view. I conclude by discussing what rationality requires when individual and team rationality conflict, and make some brief comments on relativism and supererogation.

Rules play a central role in Sellars's philosophy. As Sellars writes in LRB, "When God created Adam, he whispered in his ear, 'In all contexts of action you will recognize rules, if only the rule to grope for rules to recognize. When you cease to recognize rules, you will walk on four feet" (p. 298). However, radical particularists in ethics argue (often on the basis of holism about reasons) that rules play no role in ethical judgment or moral justification. I argue in Chapter 8 that while radical particularism is not very plausible, it is incumbent upon the Sellarsian to develop a notion of a rule that has genuine force, but admits of exceptions—that can serve the role of Sellars's important *material inferential rules*. Fortunately, recent years have seen the emergence of well-developed accounts of *defeasible rules of inference*—both less formal accounts (like Lance and Little's) and formal ones (like Horty's). I show how these accounts can be pressed into service for Sellars's philosophical project.

I turn in Chapter 9 to address a second concern about rules, specifically, Wittgensteinian concerns that we cannot understand intelligent behavior in terms of following rules. The argument of this chapter attempts to establish two points. First, Sellars's account of *pattern-governed behavior* (behavior that is not rule-following, but which must be understood in relation to rules) responds to Wittgensteinian worries, while preserving an essential role for rules, a role to be explored at length in this chapter. Second, I demonstrate that the social practices at the bottom of Sellars's account, and the rules that govern them, must be understood in terms of collective attitudes—both we-intentions and collective beliefs—and that therefore individual rational agency strongly presupposes collective attitudes. This latter point is particularly important, because it allows us to build toward the ultimate Sellarsian goal of proving the reality—and indeed, the *necessity*—of the community of rational beings, which will bring us (in Chapter 13) to the final goal of establishing the *categorical reasonableness* of the formal end of morality.

As I noted at the beginning of Chapter 3, a persistent problem in metaethics is the question of how to reconcile the cognitive and motivational elements of moral judgment. The lynchpin to the 'moral problem' is the Humean philosophy of mind, which holds that belief and desire are 'distinct existences.' With the help of McDowell, I try in Chapter 10 to undermine allegiance to the Humean model, arguing that moral judgment doesn't involve sterile cognition of a realm of facts, but is a type of judgment which already embodies our concerns, attitudes, and normative commitments. Again, though, this doesn't mean we are cognizing 'queer' facts, because moral judgment isn't in the first instance (for Sellars) a type of *belief*, and normative utterances aren't *descriptive*—even though they are very much in the space of reasons.

Building on the argument from the previous chapter, and on the work of Sellars, Brandom, and McDowell, I present in Chapter 11 a Kantian account of moral motivation. I distinguish two cases: in the first case, we are motivated by explicitly moral judgments (e.g., "Kindness is good") because such judgments are expressions of intention, and hence conceptually connected to action. In the second case, we do not make an explicitly moral judgment, but judge that some feature of the world is salient in some way (e.g., "My friend is in trouble and open to being comforted"). Extending the argument from the previous chapter, I argue that to see certain features as salient is already to make a judgment embodying various concerns, attitudes, and commitments; and hence, it is to make a judgment that is also essentially *practical* in nature (i.e., tied to intention and action).

In Chapter 12, we begin our approach to categorical validity. Sellars rejects the idea that the moral 'ought' can reside in the consequent of a hypothetical imperative, such as "If you intend E, then you ought to do A." Following Jean Hampton, Joseph Heath, Christine Korsgaard, and others, I will argue that all reasoning cannot be reduced to instrumental reasoning; Sellars is correct that we need a substantive account of what ends are intrinsically

reasonable. However, we can see that many attempts to 'ground' morality are motivated by moral foundationalism—by the 'Given' transposed to the practical sphere—particularly when these attempts demand that morality be grounded in non-moral reasons, and given a 'non-circular' justification. Our anti-foundationalist, anti-Given argument will demonstrate that the latter demand simply cannot be met in any justificatory enterprise—be it epistemology, morality, or what have you—and we should not hold practical reason up to this impossible standard. Just as Sellars rejects the Given in epistemology, so should the Sellarsian reject the Given in practical reasoning.

Sellars was ambivalent as to the prospects of deriving the "reality… of an ethical community consisting of all rational beings" (SM 7.XX.§144/p. 225), and hence the intrinsic validity of the formal end of morality. I think the prospects are rosier than he anticipated, and set out to prove in Chapter 13 the reality of this community—and hence the categorical reasonableness of the formal end of morality. Our dialectical opponent is the rational egoist; this is the point of view that Sellars saw as competing against the moral point of view for our allegiance. I argue that rational egoism (RE) cannot be established just by considering the nature of practical reason. To establish RE on teleological grounds would require proving that an agent's welfare is the only good (for her). But I will demonstrate that this position cannot be consistently held. I already argued in Chapter 9 that the very norms that are expressed in an agent's pattern-governed behavior embody a range of we-mode attitudes, and that pattern-governed behavior must be understood in the we-mode. Thus, rational agency itself cannot be understood apart from the community. Nor can the agent's welfare, autonomy, or other essential goods or capabilities—indeed, various essential elements of her identity. Thus, there is no meaningful way to argue that an agent's well-being matters, but the welfare of the community—which (partially) constitutes her agency, her welfare, and so on—does *not* matter.

Further, and more radically, to be a rational agent in the first place is already to occupy the perspective of the 'we', of the community of rational (and hence, as I will argue, *moral*) agents.

In Chapter 14, I try to reconcile an inconsistency between Sellars's ethical theory and his larger philosophical project. Sellars rejects formalism in theoretical reasoning, arguing that

This dry, well-made match is struck
Therefore, the match will light

is a good inference, and doesn't depend for its validity on the insertion of a major premise. In ethics, though, Sellars seems to embrace this formalism, arguing that

I am in C_i
Therefore, Shall$_{we}$ [I do A_i]

is good *qua* instantiation of the intermediate moral principle

Shall$_{we}$ [I do A_i, if in C_i]

which in turn is good only if derived from the supreme principle of morality

Shall$_{we}$ be [Our welfare is maximized]

This formalism is inconsistent with Sellars's other commitments. We should recognize that intermediate moral principles have original (but not foundational) authority; they are 'auxiliary moves' as described in SRLG. But there is still a role for the supreme principle of morality: It is a theoretical postulate which serves a dialectical role to resolve disputes among mid-level intermediate moral principles and to bring consistency into our moral practice. Thus, we can preserve the formal structure of Sellars's ethical theory, while preserving his opposition to formalism *per se*.

I began this book with a discussion of Sellars's strict naturalism—his *scientia mensura* principle—and the nominalism he developed to accommodate normative (and, more generally, philosophical) concepts within this principle. I conclude in Chapter 15 by tying together various elements in the book to demonstrate how Sellars's ethical theory comports with his stringent naturalism. The key move for Sellars, of course, is his understanding of normative

claims as non-descriptive. This obviates the need to 'place' ontologically-problematic normative facts into a scientific world-view; it avoids 'queerness' worries as to how such facts can be intrinsically motivating; and it avoids is-ought gap concerns about how natural facts could generate normativity. By drawing on social practice theorists like Joseph Rouse, I will demonstrate that while Sellars's theory is non-descriptive, it paints a picture of our moral practice as ineliminably tied to the natural world through essentially world-involving practices. We are thus left with a picture of our moral practice as robustly objective, involving binding normativity, yet fully consistent with Sellars's strict nominalism and his *scientia mensura*.

---

[1] Compare the emotivism of C.L. Stevenson, a contemporary of Ayer's: "We must distinguish between 'disagreement in belief' (typical in the sciences) and 'disagreement in interest.'…Let me give an example of disagreement in interest. A. 'Let's go to a cinema tonight.' B. 'I don't want to do that. Let's go to the symphony.' A continues to insist on the cinema, B on the symphony. This is a disagreement in a perfectly conventional sense…It is disagreement in *interest* that takes place in ethics. When C says 'This is good,' and D says 'No, it's bad,' we have a case of suggestion and counter-suggestion. Each man is trying to redirect the other's interest" (Stevenson 1937, p. 27).

[2] Fundamentally, as Steven A. Miller notes, "Ayer 'failed to distinguish between emotions and attitudes' [SE, pp. 226-7] insofar as he needlessly ran together kneejerk reactions and carefully-considered, thoughtfully-habituated positions about the world" (Miller 2018, p. 20).

[3] Although his account clearly did not appear in a vacuum, as he was influenced by sociologists like Durkheim and Bouglé (see Olen and Turner 2015, 2016) and, I strongly suspect, by philosophers like George Herbert Mead.