Kant and Sellars on the unity of apperception

David Landy

Abstract: That Wilfrid Sellars claims that the framework of persons is not a descriptive framework, but a normative one is about as well known as any claim that he makes. This claim is at the core of the famous demand for a synoptic image that closes, "Philosophy and the Scientific Image of Man," makes its appearance at key moments in the grand argument of, "Empiricism and the Philosophy of Mind," and is the capstone of Sellars' engagement with Kant in *Science and Metaphysics*. Whereas mere things can be subject to ought-to-be rules – e.g. a clock ought to chime on the hour – to be a person, as Sellars understands it, is to be subject to ought-to-do rules – e.g. one ought to wind one's clocks to chime on the hour. *Prima facie*, though, there is more to being a person than just being subject to ought-to-do rules. For example, on at least some common ways of using 'person' to be a person is to have a unified consciousness, i.e. to be a single subject of a manifold of experience persisting through time. Arguably, that is what Kant takes a person to be. What I hope to show here is that it is what Sellars takes a person to be too. I.e. the exciting twist here is that as Sellars sees it being a single subject of experience persisting through time is being subject to a particular kind of ought-to-do rules, namely, those concepts-qua-inferential-rules that are the means by which we represent the world of causally-related objects existing in space and persisting through time.

Keywords: personal identity; practical reason; theoretical reason; transcendental deduction; synoptic image.

> Thus, in the union of pure speculative with pure practical reason in one cognition, the latter has primacy, assuming that this union is not *contingent* and discretionary but based a priori on reason itself and therefore *necessary*.
>
> (Kant 2015: 101)[1]

[1] My thanks to Samantha Matherne, Willem deVries, Tim Jankowiak, Micah Dubreuil, Luz Seiberth, Danilo Manca, and Giacomo Turbanti for reading and commenting on earlier versions of this paper. Thanks also to the participants of my Fall 2020 graduate seminar on Sellars at San Francisco State University for all of our discussions on this and related topics. Finally, thanks to the participants of the North American Kant Society Pacific Study Group 2021 meeting for their helpful questions and comments.

philinq X, 1-2022, pp. 49-72
ISSN (print) 2281-8618-ETS doi: 10.4454/philinq.v10i1.393
Submitted: January 2021
Accepted: October 2021

That Wilfrid Sellars claims that the framework of persons is not a descriptive framework, but a normative one is about as well known as any claim that he makes. This claim is at the core of the famous demand for a synoptic image that closes "Philosophy and the Scientific Image of Man",[2] makes its appearance at key moments in the grand argument of "Empiricism and the Philosophy of Mind",[3] and is the capstone of Sellars' engagement with Kant in, *Science and Metaphysics*.[4] Whereas mere things can be subject to ought-to-be rules – e.g. a clock ought to chime on the hour – to be a person, as Sellars understands it, is to be subject to ought-to-do rules – e.g. one ought to wind one's clocks to chime on the hour.

Prima facie, though, there is more to being a person than *just* being subject to ought-to-do rules. For example, on at least some common ways of using 'person' to be a person is to have a unified consciousness, i.e. to be a single subject of a manifold of experience persisting through time. Arguably, that is what Kant takes a person to be.[5] What I hope to show here is that it is what Sellars takes a person to be too. I.e. the exciting twist here is that as Sellars sees it being a single subject of experience persisting through time *is* being subject to a particular kind of ought-to-do rules, namely, those concepts-qua-inferential-rules that are the means by which we represent the world of causally-related objects existing in space and persisting through time.

I take Sellars' reasons for holding this set of theses to be essentially Kantian, and so my procedure for explicating them will be to trace a single philosophical thread through both Kant's and Sellars' thinking surrounding these issues. I begin with the historical problematic to which Kant's Transcendental Deduction is intended as an answer. By what right does one apply the pure *a priori* concepts of the understanding? As the necessary means for representing the analytic unity of apperception, i.e. for representing oneself as the single subject of experience persisting through time. That leads to a consideration of the question of what the temporally-discursive experiences are of which one is supposed to be the single subject, and what the nature of the relation is of these experiences to such a subject. Here Sellars provides the answer. The question is ill formed. There is no *relation* of experiences to a subject because experiences are not themselves *things*. Rather, 'an experience' is a nominalization of the verb 'experiencing', which is itself a description of the act of a person. That thesis, then, brings us squarely to the question of what the framework

[2] Sellars 1963c: §114.

[3] Sellars 1963a: §12.

[4] Sellars 1967: Chapter VII, §1.

[5] E.g. A361. All citations from the *Critique* are to Kant 1998. I use the standard convention of citing the pages numbers of the original A and B editions.

of persons is, and why Sellars is so confident that it is an ineliminable feature of any future iteration of the synoptic image of the world. The answer to the latter question is that descriptive images themselves (scientific or manifest) are constituted by the rules that govern them, and it is only persons that can be subject to such rules.[6] So, in the end, we return to Kant's claim in the Transcendental Deduction that our representation of a world of causally-related objects existing in space and persisting through time is the means by which one represents oneself as the single subject of experience persisting through time. The descriptions that the scientific image provides are only possible, and necessary, because of what Kant would call their "ultimate principle": the framework of persons.

1. *The historical problematic*

To begin, consider the historical problematic surrounding the concept of the self with which Kant finds himself confronted.[7] Descartes seems to regard as valid an inference that moves from a premise of the form

(D1) [I think x] and [I think y] and [I think z],

to a conclusion of the form

(D2) [The I that thinks x] = [The I that thinks y] = [The I that thinks z].

That is, Descartes takes the fact that he can introspectively observe that he thinks x, and that he can introspectively observe that he thinks y, and that he can introspectively observe that he thinks z, to imply that it is one and the same thing, he, the thinking thing, that is the single subject of all of those thoughts.

> Is it not one and the same "I" who is now doubting almost everything, who nonetheless understands some things, denies everything else, desires to know more, is unwilling to be deceived, imagines many things even involuntarily, and is aware of thigs

[6] Willem DeVries points out in correspondence that there might also be more primitive representational systems that are likewise constituted by the rules that govern them, but which do not require persons to be subject to these rules. For example, the cries of monkeys that distinguish between predators above and predators below. What Sellars says about such systems in Sellars 1963b: §14-§16 is that while it is true that these particular cries are only explicable via appeal to the larger pattern of which they are a part, and that this is something that they share with the objects of ought-to-be rules, because ought-to-be rules are logically connected to ought-to-do rules, such cries are not genuinely rule governed. More below.

[7] This way of framing this problematic derives from Sellars' lectures on Kant at the University of Pittsburgh (Sellars 2002c) by way of Jay Rosenberg's lectures on Kant at the University of North Carolina (Rosenberg 2005). I have put it to use elsewhere as well (Landy 2015).

just as true as the fact that I exist, even if I am asleep all the time, and even if he who created me is doing all he can to deceive me? [. . .] The fact that it is I who am doubting and understanding and willing is so evident that I see no way of making it any clearer. (Descartes 1984: 19)

Of course, Hume finds the matter to be significantly less clear. He denies that this inference – from the introspective availability of certain experiences or perceptions to the identity of the subject of these experiences – is valid at all.

For my part, when I enter most intimately into what I call *myself*, I always stumble on some particular perception or other, of heat or cold, light or shade, love or hatred, pain or pleasure. I never catch *myself* at any time without a perception, and never can observe any thing but the perception. (T 1.4.6.3-4; SBN 252)[8]

What Hume points out here is that, when we introspect, we find exactly the matter that Descartes does – this or that perception – but that this is not sufficient to yield an experience of the self – something that endures through time and is the subject of these perceptions. Lacking such an experience of the self, Hume turns his attention to a different question. Given that I am able to introspect and find such-and-such a manifold of perceptions, what makes these perceptions, but not others distinctly *mine*? Hume's first answer to this question is,

> that the true idea of the human mind, is to consider it as a system of different perceptions or different existences, which are link'd together by the relation of cause and effect, and mutually produce, destroy, influence, and modify each other. (T 1.4.6.19; SBN 261)

The mind is a bundle of perceptions united by certain relations of cause and effect. In reconsidering his view in the appendix to the *Treatise*, Hume notoriously expresses his dissatisfaction with that account, but does not specify what the grounds of that dissatisfaction are.[9]

Kant, by contrast, is more than happy to express the grounds of *his* dissatisfaction with Hume's account. Most importantly for current purposes, Kant takes (D2) to be *analytic*, and dubs this representation of oneself the analytical unity of apperception. As Kant sees it, Hume is exactly right that merely adding together the manifold of representations represented in (D1) is not suffi-

[8] References to the *Treatise* are to Hume, *A Treatise of Human Nature*, ed. Norton and Norton, hereafter cited in the text as "T" followed by Book, part, section, and paragraph number, and to Hume, *A Treatise of Human Nature*, ed. Selby-Bigge, rev. by Nidditch, cited in the text as "SBN" followed by the page number.

[9] There are at least two dozen extant distinct interpretations of what bothers Hume in the Appendix.

cient for representing the subject of these representations as identical through time (D2). Where he goes wrong, however, is in inferring from this insight that such a subject cannot be represented at all.

> Namely, this thoroughgoing identity of the apperception of a manifold given in intuition contains a synthesis of the representations, and is possible only through the consciousness of this synthesis. For the empirical consciousness that accompanies different representations is by itself dispersed and without relation to the identity of the subject. The latter relation therefore does not yet come about by my accompanying each representation with consciousness, but rather by my adding one representation to the other and being conscious of their synthesis. Therefore it is only because I can combine a manifold of given representations in one consciousness that it is possible for me to represent the identity of the consciousness in these representations itself, i.e., the analytical unity of apperception is only possible under the presupposition of some synthetic one. (B133)

Kant holds that one can (and must) represent oneself as a single subject of experience persisting through time, and that the means for doing so is forming a single complex representation the components of which are the manifold of representations of which one can become introspectively aware.[10] That is, Kant holds that representing oneself as in (D2) is made possible by,

(K) I think [x + y +z].

Since [x + y +z] is a single unified representation, it is necessarily had by a single unified thinker. Since the components of [x + y + z] are the very representations that one finds via introspection, e.g. in (D1), (D1) and (D2) alike follow from (K): it is one and the same thinker that is the subject of x, y, and z. The representation of the self of the form presented in (K) is what Kant calls the synthetic or transcendental unity of apperception. We will return to this representation at the close of the current study.

In the meantime, notice that the success of Kant's strategy for resisting Hume's conclusion that one cannot so much as represent oneself as a single subject of experience persisting through time hangs on (among other things) Kant's claim that (D2) is analytic. It is analytic that the subjects of each of the representations that one finds in introspecting are all identical. Here is a famous passage to that effect.

[10] Strictly speaking, in the Transcendental Deduction, the unity of apperception under consideration is purely formal and so it does not yet include persistence through time. That condition is added to the concept of a person later in the *Critique*, once Kant has reintroduced our particular forms of intuition, Space and more fundamentally Time, in the Schematism.

> The I think must be able to accompany all my representations; for otherwise something would be represented in me that could not be thought at all, which is as much to say that the representation would either be impossible or else at least would be nothing for me. (B131-2)

By "The I think" here, Kant means a univocal representation, one that represents the same subject of experience as the subject of every representation that it accompanies. In light of Hume's rejection of the possibility of any such representation, however, it is worth asking why exactly Kant takes this thesis to be not only possible, but also true, and *analytic*. Hume appears to be able coherently to hold that there is no single subject of all of "my" representations persisting through time. If Kant is right, though, this appearance belies an underlying incoherence. What is that incoherence? Here, in answering this question, we will turn for the first time to Sellars.

2. *Sellars on sensations and thoughts*

To see why Sellars follows Kant in taking,

(D2) [The I that thinks x] = [The I that thinks y] = [The I that thinks z]

to be analytic, it will be helpful to begin again with Hume. Recall that Hume takes a distinctly *ontological* approach to accounting for the human mind. Hume holds that the mind is nothing other than the perceptions that compose it. Here is how he summarizes his own position in his anonymously-published abstract of the *Treatise*.

> He asserts, that the soul, as far as we can conceive it, is nothing but a system or train of different perceptions, those of heat and cold, love and anger, thoughts and sensations; all united together, but without any perfect simplicity or identity. *Des Cartes* maintained that thought was the essence of the mind; not this thought or that thought, but thought in general. This seems to be absolutely unintelligible, since every thing, that exists, is particular: And therefore it must be our several particular perceptions, that compose the mind. I say, *compose* the mind, not *belong* to it. The mind is not a substance, in which the perceptions inhere. (A 28; SBN 657-658)

Hume's portrayal of Descartes here is almost certainly inaccurate, but putting that matter aside, notice again that Hume holds that the mind is a complex composed entirely of its perceptions. I.e. he holds that our primary understanding of perceptions categorizes them as *things* and the question of the unity of the subject of experience hinges on the further question of what

the *relation* of these things is to that subject. In the end, Hume denies that there can be any such relation, and so settles (temporarily, at least) on his view that the mind just is a bundle of perceptions. I.e. the mind is *composed* of its perceptions, which stand in certain relations only to each other.

What is significant about this way of understanding Hume's account is that it rests on what Sellars takes to be a false premise: that our primary understanding of mental representations categorizes them as *things*. In place of such an understanding, Sellars urges that we interpret the idioms of mental representa*tions* as nominalizations of verbs of represent*ing*, which apply first-and-foremost to *persons*.

> [I]t is surely implausible to take such statements as
>
> Tom has a feeling

To be anything but a derivative (but legitimate) way of saying what is said adequately and non-relationally by such statements as

> Tom feels…

Thus, in general,

> Tom has a V-tion,

where 'V-tion' is a verbal noun for a kind of "experience", would be a derivative (but legitimate) way of saying what is said adequately and non-relationally by

> Tom Vs. (Sellars 2002a: 313-314)

As Sellars sees it, the vocabulary of ideas, perceptions, etc., while it *appears* to categorize mental representations as things, is actually a derivative idiom. Primarily, thinking, perceiving, representing, etc. are *acts* of *persons*. Just as we might say,

> Dave wore a smirk

and thereby invite the question of what the relation is between Dave and the smirk that he wore, such a question is easily avoided by noticing that 'a smirk' is a nominalization of the verb 'smirking'. Smirks do not exist without the persons that wear them, not because the existence of smirks depends in some metaphysical way on the existence of persons, but rather because 'a smirk' is a derivative way of representing the act of smirking, which is something that persons do. The above is more perspicuously put as,

> Dave smirked.

Similarly, Sellars suggests that 'I think x' is not a statement of a relation between me and *the thought* x, but rather a representation of me as think*ing*. More specifically, it is the representation of me as thinking *x*, where 'x' is also not a *thing* (say, a content) to which my thought relates, but is rather a functional classification of my thinking.

> The thought that snow is white occurred to Jones,

which is doubly relational in appearance, turns out to have as its foundation the non-relation state of affairs expressed by

> Jones ·snow is white·ed. (Sellars 2002a: 318)

Just as single quotes mention the word between them qua a word in a particular language – e.g. 'dog' has three letters – Sellars' dot quotes individuate words according to their inferential role – e.g. German 'rot's are ·red·s.[11] I.e. 'rot' plays the same inferential role in German as 'red' does in English. Sellars' first point here is that in attributing the thought 'snow is white' to Jones, firstly, we should conceive of that thought not as an entity distinct from Jones himself, but rather an act of thinking that Jones performs. His second point is that in in specifying that Jones is thinking that snow is white, we are classifying that act of thinking as playing a certain inferential role: it is a thinking of a snow-is-white kind, or is a snow-is-white thinking.

To return to our analogy, rather than interpret,

> Dave wore a smirk that was wry,

as a relation between Dave, the smirk that he wore, and its wryness, we should understand it as a derivative way of expressing an adverbial classification more perspicuously represented by,

> Dave smirked wryly.

Dave is the only *thing* represented in that proposition; he is represented as having acted, he smirked; and he acted in a particular way, he smirked wryly.

Analogously, then,

(D1) [I think x] and [I think y] and [I think z],

[11] I use "inferential role" here as shorthand for the entire complex of language-entry, language-language, and language-exit transitions that Sellars takes to constitute a language. It is important to note, however, that this shorthand belies the fact, about which Sellars is explicit and emphatic, that language-entry and language-exit moves are not inferences.

which attributes a manifold of representations to me is not the statement of a relation between a thing, me, to some other things, the thoughts, x, y, and z. Rather, it is a functional classification of the various acts of a person, me, as an act of functional kind x, an act of functional kind y, and an act of functional kind z. The important point here is that the notion of a person is logically prior to the notion of that person's thoughts in just the same way that the notion of a person is logically prior to the notion of a smirk. 'A thought,' just like 'a smirk' is a nominalization of its corresponding verb, 'thinking', and no account is needed of the relation of a person to their thoughts.

Of course, Sellars famously insists on a distinction between thoughts (conceptual representations the content of which can be described functionally) and sense impressions (which do not have a content *per se*, but which are characterized firstly according to their causal role, and then also by their intrinsic characteristics). What I have been concerned with above is thoughts, but Sellars offers an analogous treatment of sense impressions.

In this perspicuous language we would not say,

> Tom senses a red triangle

but

> Tom a-red-triangles

where the verb 'a-red-triangle' stands for that kind of sensing which is brought about in standard conditions, and in standard perceivers, by the presence of a literally red and triangular object. (Sellars 2002a: 317)

To say that a subject has a certain sensation is not to assert that a relation holds between that subject and that sensation, but is rather to classify an act of that person's as being of the kind that is typical of persons in such-and-such circumstances. In the case of both thoughts and sensations, the idiom of mental *representations* is derivative of the idiom of *persons* and their acts of mental *representing*, and is used primarily to classify such acts, either according to their inferential role in the case of thoughts, or according to their causal role in the case of sensations. So, our concepts of persons are logically prior to our concepts of their representations, and the question of the relation of former to the latter ought not to arise. Thinking x and thinking y (or sensing x and sensing y) are ways that *I* act just as smirking or winking are ways that I can arrange my facial features. My thought x and my thought y depend on me in the same way that my smirk or my wink do.

Of course, this interpretation of the deep logical grammar of representation is not by itself sufficient for accounting for the analyticity of the analytical

unity of apperception. The question of whether the person who winked was the same as the person that smirked can be a perfectly coherent one in certain circumstances. Analogously, the question of whether the person who thought "Snow is white" was the same person as that who thought "Grass is green" can be a perfectly coherent one. There is more work to be done in telling Kant's and Sellars' story. For the moment, however, another task has it pushed its way to the top of our agenda.

Readers familiar with Sellars might note that the account just offered of the idiom of representation, both that of sensation and of conceptual representation, as being derivative of the idiom of persons and their acts of representing, is one that has thus far concerned only the manifest image. That is, this account has cast representing as something that persons do, and 'persons' is very much a manifest-image category. One might wonder, then, what happens to this account when the descriptive component of the manifest image is replaced by that of the scientific image. Does the category of persons persist? Is it guaranteed to persist across all such changes in images?

3. *The persistence of 'persons'*

In answering these questions, it is important to recall that the manifest image is itself a synoptic image: it has both a descriptive component and a normative one. I.e. one aspect of the manifest image is the picture of the world that it contains. This picture represents the world as being composed of tables and chairs, elephants and mice, etc. Another aspect of it is non-descriptive, and consists of the vocabulary of reasons, entitlements, commitments, etc. Demands for explanation and prediction (more on the source of which soon) reveal the inadequacy of the *descriptive* component of the manifest image, and thus the scientific image is created as its replacement, but there are no such parallel pressures put on its normative aspect. While certain philosophical considerations might put explanatory pressures on our account of norms – e.g. Sellars take very seriously the need to account for the motivational power of normative beliefs – the pressure to produce a picture of the world of increasing accuracy and predictive and explanatory success does not, as far as Sellars is concerned, put a corresponding pressure on the normative idiom of the manifest image. That idiom is never intended as descriptive, and so the need for an improved description of the world is largely orthogonal to its functioning.[12]

[12] Sellars famously comments on this confusion of these two aspects of the manifest image: "Now the idea that epistemic facts can be analysed without remainder – even 'in principle' – into non-epistemic facts […] is, I believe, a radical mistake – a mistake of a piece with the so-called 'naturalistic fallacy' in ethics." (Sellars 1963a: §12)

This difference between the descriptive aspect of the manifest image and its normative aspect is what requires that we pay careful attention to the distinction between sensory representing and conceptual representing. As we noted above, to classify a mental representing as a sensing, and as a sensing of a particular kind, is to give (or issue a promissory note for giving) an intrinsic characterization of it. In "Tom a-red-triangles" the verb 'a-red-triangle' stands for that kind of sensing which is brought about in standard conditions, and in standard perceivers, by the presence of a literally red and triangular object. As such, sensings are themselves "mere" states of the experiencing subject, and their ultimate representation is the province of the scientific image.[13] By contrast, in "Jones ·snow is white·ed" the verb '·snow is white·ed' classifies Jones' thought as playing a certain inferential role, or as being subject to certain *rules* of inference. For example, it takes Jones to be subject to criticism if he also thinks that snow is blue, or if he does not also think that snow is the same color as clouds, etc.

According to the manifest image, it is persons that are the logical subjects of ascriptions of mental states, both sensory mental states and conceptual ones. In the case of sensory states, the scientific image replaces the manifest image, and so the logical subjects of the scientific image replace the logical subjects of the manifest image. Persons qua the logical subjects of sensory states become persons qua collections of atoms in the void, and as Sellars sees it, eventually persons qua absolute processes. In the case of conceptual representings or thoughts or thinkings, however, since these are not *per se* ontological, but rather normative, their logical subject persists through changes in our conceptions of its material constitution. It is *a person* that is subject to rules of performance should they fail to satisfy their inferential commitments, and our concept of a person is not the concept of a particular kind of matter-of-factually describable substance, but rather is the concept of that which is subject to such norms. To use a version of one of Sellars' preferred analogies, the rules that govern playable moves in a game of chess apply equally to human, computer, and alien players alike, regardless of whether they are made primarily of carbon, silicon, or more exotic kinds of matter. What makes one a player of chess, is that one is subject to the rules of chess. Analogously, what makes one a thinking person, is that one is subject to the (linguistic and/or conceptual) inferential rules that constitute the content of our thoughts.

For as was pointed out, not only are concepts pertaining to conceptual representations analogical counterparts of concepts pertaining to verbal be-

[13] Thereby hangs a tale. I have attempted to relate that epic yarn in Landy (2019). Others have told their own version of it as well, e.g. Rosenberg (1982), Rosenthal (2015), Seibt (2015), and Seibt (2000).

haviour but, which is more important, the latter concepts are concerned with correctnesses and uniformities of linguistic configurations, extra-linguistic objects and non-linguistic behaviour. As for the "qualitative content" of these configurations, it must, we have said, be such as to be capable of taking part in these configurations. (Sellars 1967: 173)

To classify a mental state as a conceptual representing is to classify it according to what makes it correct or incorrect, what one is obliged or forbidden to represent in virtue of endorsing such a representing, what warrants such representing and what one is required to do in virtue of having so represented. As Sellars notes here, such classifications are not descriptions, and entail descriptions of the world only insofar as they require that the world be such as to make possible such classifications. To return to the chess analogy, classification of some worldly object as a chess piece leaves enough latitude for the classification to apply to such diverse objects as pieces of wood, luxury automobiles, pixels on a screen, and human beings. Thus, it seems as though while the replacement of the descriptive component of the manifest image might entail a recategorization of our concepts of acts of *sensing* as distributed across logically complex subjects, there is no reason to think that such a recategorization will be required for our *conceptual* representings.

Sellars, however, argues for an even stronger conclusion. Not only do we have no reason to think that conceptual representings and the normative structure of persons of which they are a part will require replacement, but we can also know that they will not. I.e. the concept of persons is guaranteed to persist through all subsequent replacements of descriptive content by the scientific image.[14]

[14] In response to a question about whether 'person' would be a category in Sellars' "ultimate ontology" following one of his lectures at Notre Dame, Sellars proclaims:: "No, I mean it will. On the contrary. Ontology is functioning, again, in an ambiguous way here. Of course there are persons. The question is: could a person have the kind of features that in practical reasoning we essentially conceive of them as having and still, in some sense, be pluralities? That's why, in my essay, 'Towards a Metaphysics of the Person,' I discussed Kant's Paralogisms. In the paper, 'This I or He or It that thinks…' the presidential address, at the end of it I, again go into Kant's Paralogisms. Because couldn't a person have exactly those features which are required by a normative view and still be a plurality, be a system, as opposed to a Cartesian simple? See, this is the old question going back to Plato of the simplicity of the soul, roughly. And that's what Kant was arguing about in the Paralogisms. And that's what I was arguing about. Kant, in fact, said that a person *could* be a system and still have those features which his ethics required him to have, and that's exactly what I say, and that's the only sense in which persons would not be ultimate simples. If you mean by 'what your ontology contains' the ultimate simples that it contains, you see, then persons might not be ultimate simples. But, as I said, that's the old classical issue of the simplicity of the soul" (Sellars 2018: 300). Notice that Sellars' response begins by pointing out that "ontology" is functioning in an ambiguous way in the question. This is because he does not take 'person' to be an ontological, or descriptive, category at all. Thus, whatever changes we make to the descriptive component of our image of the world, the category of 'persons' remains in tact, and as we are about to see, must remain in tact.

But this seems to leave the door wide open, for, as was pointed out, almost anything can be used to play the game of chess. Thus we might be inclined to say that almost anything could be the material cause or "matter" (in the Aristotelian sense) of the configurations which are conceptual representations. Yet things are not quite so simple. For though we have been emphasizing that the candid thinkings-out-loud which are the models for mental acts are not *actions*, and that the mental acts for which they are the models are not *actions* but rather acts in the Aristotelian sense, nevertheless, though we have not been emphasizing the point, there *are* mental actions. And, indeed, if there were no *actions* pertaining to thinking […] there would be no thinking […] but at best processes which, however sophisticated would be simulations of thought. (Sellars 1967: 173-174)

For all that we have said about conceptual acts of representing to this point, such acts could be governed entirely by what Sellars calls ought-to-be rules. Ought-to-be rules are rules that are applied without the object of those rules having to be in any sense aware of those rules. E.g. a clock ought to be wound so that it chimes on the hour. That rule does not require any *action* from the clock, even if we describe the chiming of a clock as an act in Sellars' Aristotelian sense. Sellars' point here, however, is that the clock's being subject to that ought-to-be rule does require something of someone. Whoever is responsible for the winding of the clock is subject to a corresponding ought-to-*do* rule: he or she ought to wind the clock so that it chimes on the hour.

Now ought-to-be's (or rules of criticism as I shall also call them), though categorical in form, point beyond themselves in two ways. In the first place they imply (in some sense of this protean term) a reason, a because clause. The exploration of this theme would seem to take us back to the excluded topic of hypothetical imperatives. In the second place, though ought-to-be's are carefully to be distinguished from ought-to-do's they have an essential connection with them. The connection is, roughly, that ought-to-be's imply ought-to-do's. Thus the ought-to-be about clock chimes implies, roughly,

> (Other things being equal and where possible) one ought to bring it about that clock chimes strike on the quarter hour.

This rule belongs in our previous category, and is a rule of action. As such it requires that the item to which it applies (persons rather than chimes) have the appropriate concepts or recognitional capacities. (Sellars 1969: 508)

In contrast to the ought-to-be rule that governs the clock, the ought-to-do rule that governs the person responsible for winding the clock does require that the person that is its subject is able, in some suitable sense, to represent that rule as governing his or her behavior. Paradigmatically, the subject of an

ought-to-do rule is able to represent their actions as being governed by rules of practical reasoning. For example, they can employ the classic form of a practical syllogism.

1. I shall bring about E.
2. Bringing about E implies doing A.
3. Therefore, I shall do A.[15]

In the example of the person responsible for winding the clocks, such a syllogism might be constructed as follows.

1'. I, a clock winder, shall bring it about that clocks chime on the hour.
2'. Bringing it about that clocks chime on the hour implies winding the clocks daily.
3'. I shall wind the clocks daily.

Ceteris paribus, forming the intention expressed in 3' results in my subsequently forming what Sellars calls a here-and-now volition to wind a clock, which in turn results in my then and there winding a clock. In the case of a conceptual agent constructing a descriptive image of the world, as we saw Sellars point out above, the inferences in question, "are concerned with correctnesses and uniformities of linguistic configurations, extra-linguistic objects and non-linguistic behaviour." An example of a practical syllogism with these concerns would be something like the following.

1''. I, a subject of experience, shall bring it about that my thinking is consistent.
2''. Bringing it about that my thoughts are consistent implies rescinding either my belief that it has rained (because the streets are wet) or my belief that it has not rained (because I have not heard rain in some time).
3''. Therefore, I shall rescind either my belief that it has rained (because the streets are wet) or my belief that it has not rained (because I have not hear rain in some time).

Ceteris paribus, forming the intention expressed in 3'' results in my subsequently forming a here-and-now volition to rescind one of these beliefs (perhaps after checking the weather, seeing if the neighbor has been running his sprinkler, etc.), which in turn results in my then and there rescinding one of those beliefs.

Notice that each of the syllogisms that we have considered concludes with

[15] Cf. Sellars 2002b: 308.

the formation of an intention (indicated by the 'shall' that appears in each of them), but does not reach its fruition until that intention itself results in a volition, which in turn results in the action intended. While this terminus is in no sense guaranteed by the mere fact of the reasoning having occurred, it is part of the logic of practical reasoning, that such actions are its *ceteris-paribus* results. I.e. one important difference between theoretical and practical reasoning is that practical reasoning aims at, and *ceteris paribus* results in, action. As Sellars sees it, that is because practical reason begins and ends with intentions, and intentions would not be intentions if they did not bear this relation to volition and in turn to actions. This connection of practical reasoning to action brings us back to the connection between the subject of experience and practical reasoning.

Recall that the dialectic we have been tracing ran thusly. We noticed that Sellars holds that the logical subjects of *thoughts*, persons, are guaranteed to persist through all replacements of the descriptive content of the manifest image by the progressively more explanatory scientific image. We wondered what the grounds for this claim were, and began investigating them by noticing that Sellars' takes it to be an essential feature of conceptual thinking that it is subject to certain norms, rules of inference. This, in turn, led us to distinguish ought-to-be-rules from ought-to-do rules, and to further note that something's being subject to the former implies that something, a *person*, is also subject to the latter. The outlines of our guarantee are now coming into focus. Conceptual thinking is rule-governed representation; it is persons that are the subject of such rules; thus conceptual thinking guarantees the persistence of persons.[16]

The question remains, however, whether anything here further guarantees the unity of *a person across a variety of representations* (especially through time), which is the essential feature the analytic unity of apperception. Sellars' answer is that it does. Specifically, it is the essentially first-personal role of *intentions* in the pieces of practical reasoning that constitute conceptual thinking that does so.

These considerations highlight the fact that the intention expressed by a 'shall' statement is invariably the speaker's intention. Thus,

Tom shall do A

Expresses the speaker's intention that Tom do A. This 'first person' feature of intentions consists in part in their relation to the

I shall do

[16] In fact, it guarantees the persistence not just of persons but also the *concept* of persons because, as we are about to see, a person just is that which conceives of itself using the concept person.

Which can become the commitment to do something *here* and *now* which is volition. (Sellars 1967: 184-185)

It is part of the logic of practical reasoning that the subject of the intention that serves as a premise must be the same as the subject of the intention that is the conclusion (and that eventually, *ceteris paribus*, performs the action so intended). For example, such arguments would not be valid if it were one person that intended to achieve some end, but another that took up the means to that end. E.g. the following inference is obviously invalid.

1. Dave shall achieve end E.
2. M is the means to E.
3. Therefore, Sherlock shall pursue M.

The very idea of practical reasoning depends on the univocality of the subject of such reasoning, that it is the very same person that adopts both the ends and means at issue. Furthermore, though, this form of reasoning supposes not just that it is *some* one person that pursues both these ends and means, but that it is some *I* that does so. Notice the difference between the following two arguments.

1. Dave has end E.	1'. I shall achieve end E.
2. M is the means to E.	2'. M is the means to E.
3. Dave pursues M.	3'. I shall pursue M.

1 and 3 are third-personal reports on Dave's intention and do not themselves result in the formation of any new intentions or volitions, and do not directly lead to any actions. 1' and 3', however, are each the expression of an intention, and 3' *is* importantly a new intention that is formed in the course of this reasoning, and does, *ceteris paribus*, lead to the appropriate volition and action.[17] So, effective practical reasoning presupposes the ability to represent *myself* as the *single* subject of practical norms.

Combining this conclusion with our earlier observation that Sellars understands conceptual representation as representation according to conceptual norms, we arrive at the conclusion that for there to be any conceptual repre-

[17] Notice that if in the following argument, 1'' and 3'' *are* intentions, but they are the speaker's intentions to shape Dave's behavior.

Dave shall achieve end E.
M is the means to E.
Dave shall pursue M.

I.e. the intentions expressed here are the speaker's intentions *that Dave achieve end E* and *that Dave pursue M*. My thanks to Willem DeVries for his guidance in sorting out these three arguments.

sentation at all presupposes that those representations are the representations of a single subject of experience persisting through time, i.e. the "existence" of persons, i.e. the analytic unity of apperception. If I am subject to ought-to-do rules because I subject my actions to practical reasoning, then it must be one and the same 'I' that is the subject of the intentions that are the premises of those syllogisms as it is the subject of the intentions that are their conclusions. If, however, it is *by* subjecting my actions to such bits of reasoning that I represent myself as a person at all, then it does turn out to be analytic that the I that is subject of any one of my representations must be identical to the I that is the subject of any other. That is, if what it is to be the subject of a conceptual representing is to be the subject of a piece of practical reasoning concerning that representing, and all practical reasoning requires that the subject of each of its steps is one and the same person, then being the subject of a conceptual representing requires a unity of the subject of piece of practical reasoning through all of its steps.

Of course, if we are to conceive of persons as being the subjects of practical means-end reasoning, and of their conceptual representing as actions subject to criticism in light of the ought-to-do rules adopted via such reasoning, then we must also ask what the *ends* being pursued in such reasoning are. If the unity of the conceptualizing subject is the unity of a person subject to norms of practical reasoning, then this raises the question of what the *ends* are of such a reasoner.[18] To what end does one engage in conceptual representation at all?

4. *The end of theoretical reasoning*

Before we can delve into Kant's and Sellars' answer to this question, it will be helpful to say a little bit more about the kind of representation that they both take conceptual representation to be. Both Kant and Sellars begin their treatments of concepts by noting the different roles played by that which serves as the cognitive analogue of the subjects and predicates of linguistic propositions. Here is Kant.

> [T]he cognition of every, at least human, understanding is a cognition through concepts, not intuitive but discursive. All intuitions, as sensible, rest on affections, concepts therefore on functions. By a function, however, I understand the unity of the action of ordering different representations under a common one. Concepts are therefore grounded on the spontaneity of thinking, as sensible intuitions are grounded on

[18] Certainly, there is more than just one end of all practical reasoning, but what we are pursuing here is what Kant would call its supreme principle.

the receptivity of impressions. Now the understanding can make no other use of these concepts than that of judging by means of them. Since no representation pertains to the object immediately except intuition alone, a concept is thus never immediately related to an object, but is always related to some other representation of it (whether that be an intuition or itself already a concept). Judgment is therefore the mediate cognition of an object, hence the representation of a representation of it. In every judgment, there is a concept that holds of many, and that among this many also comprehends a given representation, which is then related immediately to the object. (A68/B93)

Whereas an intuition is a determinate singular representation that pertains immediately to the object (roughly, refers to it), a concept is a kind of meta-representation. It serves as a function that takes certain intuitions as its inputs and outputs other intuitions. More specifically, as I have argued elsewhere, Kant takes concepts to be inferential rules (Landy 2015). Intuitions related to one another via such rules collectively form a picture of the world of objects existing in space and persisting through time as necessarily connected to each other. To understand how this account of mental representation is supposed to work, consider the following inference:

1. *x* is to the north of *y*.

---

2. *y* is to the south of *x*.

At first blush, this inference appears to be an enthymeme. It is not an example of *modus ponens*, *modus tollens*, conjunction elimination, or any other formal rule of inference. Since it is not valid in virtue of its logical form, the thinking goes, the ring of validity that it has can only be due to a suppressed premise. The valid argument that this one stands in for is really the following:

3. *x* is to the north of *y*.
4. If *x* is to the north of *y*, then *y* is to the south of *x*.

---

5. *y* is to the south of *x*.

The problem now is that while 4 does make for a formally valid inference when paired with 3 and 5, there is a closely related inference for which 4 is of no help.

6. Suppose *x were* to the north of *y*.

---

7. Then, *y would be* to the south of *x*.

4 is of no help here because it concerns only the actual relation of *x* to *y*. It concerns only what is the case if *x* actually is to the north of *y*, not what would be the case were *x* to be north of *y*. In order to validate the inference from 6 to 7, what is needed is a proposition that applies to these counterfactual situations as well. To this end, one might be tempted to offer,

8. If *x* is to the north of *y*, then *necessarily*, *y* is to the south of *x*.

Notice, however, that if this premise is accepted, so can the meta-level rule of inference,

9. '*x* is to the north of *y*' implies '*y* is to the south of *x*'.

That is, if 8 is true, then 1 can never be true where 2 is false. Thus, the inference from 1 to 2 is valid. Thus, in any system of representation robust enough to encompass the truth of some subjunctive conditionals, there will be principles of inference corresponding to these conditionals that are valid in virtue of something other than merely their logical form. Kant's proposal as I understand it is that it is *by* licensing such inferences, in the form of deploying concepts-as-inferential-rules, that we represent spatiotemporal objects as necessarily connected to one another. In the case above, it is by licensing the inference from 1 to 2 that we represent x and y as standing in a certain lawful spatial relation. Sellars offers an example of using a similar logical mechanism to represent the world's causal structure.

> Law-like statements, therefore, are empirically based principles of inference which authorize such inferences as, to use a crude example, 'Lightning now, therefore thunder shortly.' It also authorizes such conditionals as 'If there had been lightning then, there would have been thunder shortly' and such statements as 'There was thunder then because there had been lightning shortly before' and 'That there was lightning shortly before made it necessary that there be thunder then. (Sellars 2002b: 313)

Consider again the inferences at issue.

1'. There was just a flash of lightning.

2'. There will be a clap of thunder soon.

On its face, this argument is an enthymeme, in need of a supporting premise such as:

3'. If there was just a flash of lightning, there will be a clap of thunder soon.

Of course, while 3' might be sufficient to complete this argument, the closely related argument,

4'. Suppose there had been a flash of lightning just now.

5'. Then there would be a clap of thunder soon.

requires instead,

6'. If there were a flash of lightning, then, there would be a clap of thunder,

which implies that meta-level rule of inference,

7'. 'There was just a flash of lightning' implies "There will be a clap of thunder soon".

Thus, Sellars concludes with Kant that the way that we represent the world of spatiotemporal objects necessarily connected to one another by causal laws is by relating representations of these objects to one another (intuitions in Kant's case, names in Sellars') via inferential rules. To put it in the idiom of Wittgenstein's *Tractatus*, that we relate "*x*" and "*y*" to each other via material-inferential rules pictures that *x* and *y* are necessarily connected to each other.[19]

With that said, the question that we posed at the close of the previous section can now be put into a more specific form: what is the end for which representing the world of necessarily-connected spatiotemporal objects as such is the means? Kant's answer to this question brings us full circle. Recall that I began this study by noticing Kant follow Hume in rejecting the inference they both find in Descartes from a premise of the form,

(D1) [I think x] and [I think y] and [I think z]

to a conclusion of the form,

(D2) [The I that thinks x] = [The I that thinks y] = [The I that thinks z].

We further noted at the time that while Kant rejects this *inference*, he

[19] Nevertheless, as we shall see, the full flavor of actual modal discourse involves the way in which sentences in the first level language game containing modal words parallel sentences containing rule words ('may', 'ought', 'permitted', etc.) in the syntactical metalanguage. This parallelism is quite intelligible once one notes that the moves which are signalized in the object language by sentences containing modal words, are enjoined (permitted, etc.) by sentences containing rule words in the syntactical metalanguage (Sellars 1963b: 209).

nonetheless takes (D2) to be analytic, and so sets out to discover how it is that we can represent ourselves as single subjects of experience persisting through time (D2, the analytic unity of apperception), if not by merely representing the manifold of perceptions that constitute such a subject (D1). What Kant discovers is,

(K) I think [x + y +z].

Since [x + y +z] is a single unified representation, it is necessarily had by a single unified thinker. Since the components of [x + y + z] are the very representations that one finds via introspection, e.g. in (D1), (D2) follows: it is one and the same thinker that is the subject of x, y, and z. The representation of the self of the form presented in (K) is what Kant calls the synthetic or transcendental unity of apperception. Kant reports his result as being the surprising claim that the analytic unity of apperception depends on the synthetic one. There is another surprising claim implicit here as well, though. The representation [x + y + z] is the representation of *an object*, which as Kant reveals later is itself an instance of the representation of *the world*, i.e. the necessary connection of all spatiotemporal objects. So, what is required for representing oneself as the single subject of a manifold of representations (the analytic unity of apperception) is to represent oneself as the single subject of a single complex representation (the synthetic unity of apperception) of the world of necessarily-connected spatiotemporal objects.

> The supreme principle of all intuition in relation to the understanding is that all the manifold of intuition stand under conditions of the original synthetic unity of apperception. All the manifold representations of intuition stand under [this principle] insofar as they must be capable of being combined in one consciousness; for without that nothing could be thought or cognized through them, since the given representations would not have in common the act of apperception, I think, and thereby would not be grasped together in a self consciousness.
>
> Understanding is, generally speaking, the faculty of cognitions. These consist in the determinate relation of given representations to an object. An object, however, is that in the concept of which the manifold of a given intuition is united. (B136-B137)

What Kant is saying here is that the justification (the answer to his famous *quid juris*) for representing the necessarily-connected world of spatiotemporal objects is that it makes possible the analytic unity of apperception! That is, the end of theoretical reasoning, as Kant understands it, is the representation of oneself as the single subject of experience persisting through time. Since we have seen that and why Kant and Sellars take that representation to be

analytic, we can now also see why Kant takes the theoretical reasoning to be a non-optional form of our cognition: it is the necessary means to the analytic end of representing ourselves.

The only question remaining in this study, then, is whether Sellars agrees. I have suggested that Sellars agrees with Kant about the nature of the analytic unity of apperception. I have also suggested that he agrees with Kant about the nature of theoretical reasoning. Does he also draw the justificatory connection between these two pieces of representation that Kant does? The first thing to note in answering this question is that there may not be just one answer to it. For example, James O'Shea has argued that Sellars' position on this issue changed over the course of his career (2007: 129-136). As O'Shea sees it, Sellars gives one answer to the question of what justifies theoretical reasoning in his early essays including "Empiricism and the Philosophy of Mind" (1956), and a different one in later essays such as "On Accepting First Principles" (1988). For current purposes, I will focus on the view that O'Shea takes Sellars to articulate only in the later essays. To that end, here is Sellars in 1975 articulating how he would respond to a demand to justify employing the kind of cognitive framework that we have just been discussing.

> It must, indeed, be granted that principles pertaining to the epistemic authority of perceptual and memory beliefs are not the sort of thing which *could* be arrived at by inductive reasoning from perceptual belief. But the best way to make this point is positive. *We have to be in this framework to be thinking and perceiving beings at all.* (Sellars 1975: §45)

Any particular set of concepts-qua-material-inferential rules will be adopted on the grounds that they best explain observed empirical generalizations. But what justifies our inductive and explanatory practices themselves? Sellars' answer is that what justifies the use of this framework is that it is necessary for *being a thinking and perceiving being at all*. He elaborates.

> I pointed out a moment ago that we have to be in the framework of these (and other) principles to be thinking, perceiving, and now I add, acting beings at all. But surely this makes it clear that the exploration of these principles is but part and parcel of the task of explicating the concept of a rational animal or, in *VB* terms, of a language-using organism whose language is *about* the world in which it is *used*. (Sellars 1975: §46)

Notice that Sellars makes special mention of the fact that employing a conceptual framework is a necessary condition not only for thinking and perceiving, but also for acting. That addition fits with the general outline of the dialectic we have been sketching. The unity of the experiencing subject is the unity

of the subject of inferential ought-to-do rules, it is the unity of a conceptual *agent*. What we can now see is the point that Sellars makes in the second sentence here: that the unity of such conceptual agents is achieved via representing oneself as the single subject of experience *of a causally-structured world* (of objects existing in space and persisting through time). So, indeed, Sellars does share Kant's views not just of the experiencing subject, and of the experienced world, but also of the justificatory connection between these.

David Landy
San Francisco State University
landy@sfsu.edu

References

Descartes, Rene, 1984, *The Philosophical Writings of Descartes*, vol. 2, Eng. transl. by J. Cottingham, R. Stoothoff and D. Murdoch, Cambridge University Press, Cambridge.

Hume, David, 1974, *A Treatise of Human Nature*, ed. by L.A. Selby-Bigge, Oxford University Press, Oxford.

Hume, David, 2000, *A Treatise of Human Nature*, ed. by D. Fate Norton and M.J. Norton, Oxford University Press, New York.

Kant, Immanuel, 2015, *Critique of Practical Reason*, Eng. transl. by Mary Gregor, Cambridge University Press, Cambridge.

Kant, Immanuel, 1998, *Critique of Pure Reason*, Eng. transl. by Paul Guyer and Allen W. Wood, Cambridge University Press, Cambridge.

Landy, David, 2019, "Sellars' Argument for an Ontology of Absolute Processes", in *Journal for the History of Analytical Philosophy*, 7, 1: 1-25.

Landy, David, 2017, *Hume's Science of Human Nature*, Routledge, New York.

Landy, David, 2015, *Kant's Inferentialism*, Routledge, New York.

O'Shea, James, 2007, *Wilfrid Sellars: Naturalism with a Normative Turn*, Polity Press, Cambridge.

Rosenberg, Jay F., 1982, "The Place of Color in the Scheme of Things: A Roadmap to Sellars' Carus Lectures", in *The Monist*, 65: 315-335.

Rosenthal, David, 2015, "Quality Spaces, Relocation, and Grain," in James O'Shea, ed. *Sellars and His Legacy*, Oxford University Press, Oxford: 149-185.

Seibt, Johanna, 2015, "How to Naturalize Sensory Consciousness and Intentionality within a Process Monism with Normativity Gradient: A Reading of Sellars", in James O'Shea, ed. *Sellars and His Legacy*, Oxford University Press, Oxford: 186-223.

Seibt, Johanna, 2000, "Pure Processes and Projective Metaphysics", in *Philosophical Studies*, 101: 253-289.

Sellars, Wilfrid, 2018, *Wilfrid Sellars Notre Dame Lectures 1969-1986*, ed. by P.Amaral and D. Landy, Ridgeview Publishing Co., Atascadero.

Sellars, Wilfrid, 2002a, "Metaphysics and the Concept of a Person", in Jeffrey F. Sicha, ed. *Kant's Transcendental Metaphysics*, Ridgeview Publishing, Atascadero: 304-329.

Sellars, Wilfrid, 2002b, "On Accepting First Principles", in Jeffrey F. Sicha, ed. *Kant's Transcendental Metaphysics*, Ridgeview Publishing, Atascadero: 461-571.

Sellars, Wilfrid, 1975, "The Structure of Knowledge", in Hector-Neri Castañeda, ed. *Action Knoweldge and Reality*, The Bobbs-Merrill Company, Indianapolis: 295-348.

Sellars, Wilfrid, 1969, "Language as Thought and as Communication", in *Philosophy and Phenomenological Research*, 29, 4: 506-527.

Sellars, Wilfrid, 1967, *Science and Metaphysics: Variations on Kantian Themes*, Routledge and Kegan Paul, London.

Sellars, Wilfrid, 1963a, "Empiricism and the Philosophy of Mind", in *Science, Perception and Reality*, Ridgeview Publishing, Atascadero: 119-179.

Sellars, Wilfrid, 1963b, "Some Reflections on Language Games", in *Science, Perception and Reality*, Ridgeview Publishing, Atascadero: 300-333.

Sellars, Wilfrid, 1963c, "Philosophy and the Scientific Image of Man", in *Science, Perception and Reality*, Ridgeview Publishing, Atascadero: 7-40.