Explication in the Space of Reasons: What Sellars and Carnap Could Offer to Each Other

Krisztián Pete[1] · Adam Tamas Tuboly[2,3]

Abstract

In this paper, we reconsider the highly underrated Carnap–Sellars relationship, arguing that Sellars might be able to provide an interesting resolution to some of Carnap's finest problems around explication by offering a grand-scale picture of science/common-sense or manifest interactions. The narrative developed here points toward the need for some stratification and re-evaluation of a field of scholarship that all too often still engages in challenging and contradictory dichotomies, undermining the genuine intentions of scholars who were collaborating with, as well as living and researching in parallel with each other. We would like to explore what these figures can offer to each other, and whether, for example, a truly Sellarsian perspective could illuminate any problematic points in Carnap that have not yet been problematized by Carnapians but could still improve the original system.

Keywords Rudolf Carnap · Wilfrid Sellars · History of analytic philosophy · Manifest imagine · Scientific image · Explication · Common sense

> *"[T]he process of revision must be compared to repairing a ship at sea, rather than to reconstructing a building on the same old solid foundation." (PR 32).*

1 Introduction

While both Wilfrid Sellars and Rudolf Carnap have been drawn into various scientific and epistemic realism-antirealism debates, they—and especially the images that are attached to their names and legacies—are very often conceived exclusively as rivals in the history of analytic philosophy.[1] In these accounts, Carnap is the ahistorical, dogmatist, and formalist prototype of the fallen, simple-minded empiricist positivist, while Sellars is the hero of historically minded, sophisticated philosophers who brought back important Kantian insights and perspectives into analytic philosophy (a general narrative, moving from simple-minded Carnapian empiricists, to sophisticated Kantian Sellarsians and Hegelian Brandomians, which was very much upheld by Richard Rorty (1997) in his introductory story to Sellars's EPM). While both had a broad interest in logic, epistemology, and the scientific enterprise in general (knowing and respecting each other's work and opinion, see Gabbani 2018), according to textbook historiographies, Sellars takes off where Carnap broke down.

There seems to be a fundamental difference between the way Carnap envisages the modification and development of conceptual frameworks along the lines of explication—i.e., the concepts that make up a conceptual framework are replaced piece by piece, one concept at a time, or the entire family of concepts directly connected to a certain concept is replaced—and the way Sellars seems to conceive the change of conceptual frameworks, whereby the manifest image of

---

✉ Adam Tamas Tuboly
tubolyadamtamas@gmail.com

1 Institute of Philosophy, University of Pécs, Pécs, Hungary

2 MTA Lendület Values and Science Research Group, HUN-REN, Research Centre for the Humanities, Budapest, Hungary

3 Institute of Transdisciplinary Discoveries, Medical School, University of Pécs, Pécs, Hungary

our everyday knowledge is replaced eventually by the scientific image—that is, in one step:

> [T]he most fruitful way of approaching the problem of integrating theoretical science with the framework of sophisticated common sense into one comprehensive synoptic vision is to view it not as a piecemeal task [...] but rather as a matter of articulating two whole ways of seeing the sum of things, two images of man-in-the-world and attempting to bring them together in a 'stereoscopic' view. (PSIM, 55)

In this paper, by moving back and forth between details and the bigger picture, we aim to show what was right in Carnap and what was right in Sellars; or to put things into less evaluative terms, what each could legitimately offer the other.[2]

After reconstructing Sellars' ideas on the manifest and scientific images, we identify a problem which could be solved, or at least elucidated, through recourse to Carnap's procedure of explication. There is a difficulty regarding the criterion of success that was noted already by Sellars, and we argue that he might be able to provide an interesting resolution to Carnap's problem, by offering a grand-scale picture of science/common-sense or manifest interactions. Carnap was interested in a fine-grained, detailed, transitory picture of how one moves from the manifest image to the scientific image via explication, while Sellars' project was rather a grand narrative about the relation between these two images. But viewed in these terms, their visions and levels of inquiry are not competitive, but rather *supplementary*. As Carnap (1963, p. 940) replied to Strawson, "the naturalist and the constructionist methods are not necessarily competitive, but rather mutually complementary, since each of them fulfils a certain purpose."

While this conclusion is obviously not radical and won't change the historiography of analytic philosophy, it still points toward the need for some stratification and re-evaluation of a field of scholarship that all too often still engages in challenging and contradictory dichotomies, undermining the genuine intentions of scholars who were collaborating with, as well as living and researching in parallel with each other. We would like to explore what these figures can offer

to each other, and whether, for example, a truly Sellarsian perspective could illuminate any problematic points in Carnap that have not yet been problematized by Carnapians but could still improve the original system.

2 The Manifest and Scientific Images: Fusing Two Frameworks

Sellars is perhaps best known for his attack on the "Myth of the Given," but there is another important conception at the heart of his philosophy, namely the distinction between the manifest and the scientific image. The two images represent two fundamentally different human activities and conceptions of the world. One gives meaning to our everyday activities—it is our life-world, so to speak, which makes the world (and ourselves in it) intelligible and comfortable for us. Sellars often calls this image the world of "common sense."[3] The other image aims to describe and explain the world (the same as the world of the manifest image) by a different method, namely the postulation of imperceptible entities. While the objects of the manifest image are directly accessible to us (directly known under the right circumstances), the scientific is built on theoretical entities that are postulated, not perceived. As far as his views on the former image are concerned, Sellars (PH 61) can clearly be considered a direct realist, since in the manifest image, "seeing that a leaf is green is not a matter of seeing it look green and inferring from this ... that it is green." The entities or objects of the manifest image are normally as we perceive them. The phenomenal world of public physical objects, the manifest image, exhibits correlations between the directly perceptible qualities and relations of its objects, and the generalizations based on these correlations are intended to be explained (in a non-direct way) by the postulates of the scientific image.

The Sellarsian distinction between the manifest and the scientific image is by no means a simple one, and different interpreters understand it differently. The most common and widely accepted conception is that these "images" are ideal types of different conceptions of the nature of the human person within the natural world ("man-in-the-world," as Sellars refers to it in PSIM). According to this conception, the two images are two different answers to the question of what constitutes the world (and humans within it): the manifest world is made up of middle-sized physical objects and their perceptible properties, and—most importantly—persons,

---

[2] The Carnap–Sellars, or more generally, the positivist-Sellars literature, is anything but vast, usually representing the preferences of the respective authors; specifically, Carnap scholars often argue that Carnap was more nuanced than Sellars took him to be, and Sellars scholars try to show that despite an admitted and legitimate influence of the positivists, Sellars was right in his moves and arguments against them. All these works are highly important for future research; what we try to contribute here is an analysis drawing on both Carnap's and Sellars' perspectives at the same time. See Carus (2004), Olen (2016, 2017), and Gabbani (2018).

[3] However, describing these two images as identical is somewhat misleading, since the manifest image also contains the inductive statistical methods of the sciences. It is therefore scientific in this correlational sense – more sophisticated and refined than our everyday worldview.

while the scientific image is made up of imperceptible entities. Of these two answers (or images), only one can be true. Sellars' *scientia mensural*—"that in the dimension of describing and explaining the world, science is the measure of all things, of what is that it is, and of what is not that it is not" (EPM §41)—supports this reading, with the conclusion that "the framework of common sense is radically false" (SRI 354). So much so that Brandom (2015) adopts this interpretation, in which the manifest image is phenomenal (in the Kantian sense), only an appearance that our senses present to us, and the scientific image is the reality that lies behind all appearances.

This interpretation of the two images has been criticized by many authors, ranging from the view that such a "clear dichotomy" of images is either untenable (van Fraassen 1999) or an oversimplification of a much more complex relationship (deVries 2016), to the position that the opposition of images should be treated more flexibly, as a kind of philosophical tool, rather than as clearly distinct descriptions of the world (Dach 2018). Whether we take an instrumental approach to these images, as proposed by Dach, or a more traditional substantive one, both are driven by a common assumption, namely that the images represent two different conceptual frameworks.

Sellars considers conceptual frameworks to be systems of categories[4] (or concepts) that give meaning to our experiences and help us to know our way around the world. As such, conceptual frameworks always play both an explanatory and a practical role. This is also evident from the fact that conceptual frameworks are rule-governed activities, much like language games, with rules that define the permissible moves and their meaningful sequences. In a game, each move entails commitments, and each move also entitles to some kind of conclusion. In this sense, conceptual frameworks are ultimately uses of language integrated into non-linguistic activities. Their workings are governed by language-entry, intralinguistic, and language-exit rules (SRLG, SM IV, §61).[5]

There is a descriptive ontology built into every conceptual framework, since the basic objects of the framework constitute an ontology; in this sense, the categories that make up a conceptual framework also constitute its ontology. If we engage—purposefully, in an intelligible way—with the world equipped with a certain conceptual framework, we can only engage with things whose categories make up the conceptual framework in question. If the basic categories of a conceptual framework are persons and things, then we have to deal with persons and things in that framework. If the basic categories are subatomic particles in the framework, then we must know our way around subatomic particles. Each conceptual framework has its own categories, and each conceptual framework accordingly determines the kinds of things we may encounter. A conceptual framework is not a new interpretation or description of something that is already given, but a way to set up the world. As Sellars says in EPM (§45): *we now recognize that instead of coming to have a concept of something because we have noticed that sort of thing, to have the ability to notice a sort of thing is already to have the concept of that sort of thing and cannot account for it.*"

And it is precisely this ontological feature that is problematic in the case of the manifest and the scientific conceptual frameworks, since these two frameworks are about the same thing, the world, and the human beings in it. According to Sellars (PSIM 57), both the manifest and the scientific images "purport to be a complete image, i.e., to define a framework which could be the whole truth about that which belongs to the image." There are problems with this formulation: on the one hand, even Sellars does not say that science (in its present state) is complete, and on the other, the sameness of "that which belongs to the image" is either trivial or hopelessly vague. If frameworks define "what there is," then frameworks always constitute "the whole truth" about their subject; and if "that which belongs to the image" has a framework-independent reference, then we cannot identify it, since the identification of "noticing" is always framework-dependent. Frameworks are comparable only if the basic categories of one are reducible to the basic categories of the other. However, this is not the case with the manifest and the scientific image, even if scientific theories are based on models derived from our everyday experience[6]:

> Thus although methodologically a development within the manifest image, the scientific image presents itself as a rival image. From its point of view the manifest image on which it rests is an 'inadequate' but pragmatically useful likeness of a reality which first finds its adequate (in principle) likeness in the scientific image. (PSIM 57)

---

[4] Sellars uses both terms (conceptual framework and category) extensively in his essays – not only in relation to Kant, but also in relation to abstract entities (e.g., EAE) and his general philosophical stance. In PSIM, these very terms appear in the introduction of the images, and the description of the 'clash' between them.

[5] Language-entry transitions or rules determine how the speaker should respond to perceptible (or introspectible) objects, events, or situations with linguistic activity; intralinguistic rules are linguistic inference rules (either material ones, such as "this is green" → "this is colored," or formal ones, such as the *modus ponens*); and language-exit rules determine how the speaker should respond (through a non-linguistic action) to certain kinds of linguistic episodes.

[6] Instead of the logical method of theory construction of the logical empiricists, Sellars proposes a model-based theory construction, in which parts of an already familiar conceptual framework are taken as models. See section XIII of the EPM.

It is evident from this brief statement that Sellars regards these images as incompatible conceptions of the world. They are incompatible because of their completely different basic ontologies. Their incompatibility, according to Sellars, cannot even be resolved, as he demonstrates with the infamous *grain argument* (PSIM 63.) We thus either side with the Platonic tradition, the *philosophy perennis*, and say that "manifest objects are what really exists," thereby giving an instrumentalist reading to the scientific image, or—as implied in the quote above—we say, "manifest objects are 'appearances' to the human mind of a reality constituted by systems of imperceptible particles" (PSIM 62–63). The problem is that there are elements and features of the manifest framework that we do not want to give up, even though they are not reducible, or do not fit into the scientific conceptual framework: persons as *unities*, as fundamental entities of the world, who are in a kind of direct causal relationship with it. None of this can be maintained if we replace the manifest image with the scientific one.

What is important for the present discourse is not the primacy of the scientific framework, but the replacement of one conceptual framework with another—how this can happen—and the process by which elements of conceptual frameworks penetrate and infiltrate each other. No such framework is an unchangeable, "stable, objective order of concepts," as stated by André Carus (2004, p 327). This is obviously true of the scientific image, which is constantly evolving. Scientists are continually constructing conceptual frameworks for describing and explaining the world that do not need to retain even the basic categorial scheme of the manifest image. These conceptual frameworks of science could develop in a piecemeal fashion, and even apparently radically different ones could be related to each other. We can say that Einsteinian 'simultaneity' is different but similar to Newtonian 'simultaneity' (SM V, §45, CC 53–54)—the concepts have changed, but the conceptual role they play within each theory has remained the "same" or at least "functionally similar" (CC 52). The same is true of the manifest image, which Sellars describes as an empirical and categorial refinement of the "original image" in PSIM. Gradual depersonalization is an essential part of the development of this image. In this respect, Sellars (CC 10) argues for change, rather than replacement.[7]

But what can we say about the relationship between the manifest and the scientific framework? The accepted view, as suggested by Sellars in PSIM, is that the conceptual framework of the presently incomplete scientific image must be progressively refined (step by step) until it acquires sufficient explanatory coherence to replace the manifest image. Then it will be perfectly rational to switch to the conceptual frameworks of the scientific image regarding all human conduct in and vis-à-vis the world. To do this, we, of course, need to resolve the incompatibilities of the grain argument,[8] and be able to incorporate persons as basic *unitary* entities (not groups of cells or organs) into the scientific picture. However, there are indications in Sellars's work that would suggest otherwise, given that the images interact with each other. As Dach (2018, p 572) notes, "The relation between the MI and the SI is thus neither one-dimensional nor a one-way street. Of course, their fusion is accompanied by some kind of reduction of MI objects (the SI is, after all, ontologically prior), but the process is one where the conceptual frameworks of the MI and the SI need to be mutually adapted." On the one hand, the construction of scientific theories usually starts from the everyday (manifest) conceptual framework, using its phenomena as a model for the theory (EPM sect. XIII); and on the other hand, the scientific image penetrates the manifest in some areas, while leaving other areas untouched. For Sellars, one such intrusion is the case of mental states, beliefs, desires, perceptions, and impressions (Jonsean mental entities[9]), which are undoubtedly part of the manifest image (EPM §59, SK II, §55), but are postulates in origin, and hence elements of the scientific image.

It is, therefore, a mistake to interpret Sellars as having "an essentially static view of the manifest image," just as it is to reduce his alleged misinterpretations of Carnap to the thesis that an intuitive "manifest image *constrains* the choice of language," and therefore we are not free to choose a language that fits better to our scientific purposes (Carus 2004, pp 325–6). What Sellars has in mind is that conceptual

---

[7] We are grateful to an anonymous referee for drawing our attention to this aspect of conceptual change that takes place within the images. We ourselves have focused – narrowly – only on the conceptual interactions between the two images.

[8] The so-called "grain argument" challenges reductive materialism on the basis of the discrepancy between the manifest and scientific conceptions of the color of a pink ice cube. It originated in PSIM, and Sellars later reformulated it in SSOP and SSIS. The argument concerns the irreducibility of the ultimate homogeneity of ordinary color concepts to the particulate, aggregative concepts of modern science (properties or relations of basic physical particles that lack the macro properties). Sellars, however, advances this argument not to advocate for dualism, but to introduce his proposed counterparts to homogeneous properties, the *sensa*, into the scientific image. For a comprehensive and concise discussion of the argument, refer to Lycan (1995, 93–97).

[9] Sellars (SK II.55) sees them not as new categories or new entities of the framework, but as states of the already existing fundamental category of the manifest framework, namely that of the person. Rosenberg (1982) and Sicha (2002) handle this issue differently by essentially introducing a two-step notion of "theoretical postulation" (one that does not transgress MI, and one that does).

frameworks have an essentially normative character,[10] which in turn presupposes certain values, goals, and circumstances, as well as their recognition. The conceptualization of these is only made possible by some conceptual framework. There is no unchangeable and essentially given framework—Sellars argues against this in all his work—and we cannot go outside all frameworks. Both images are open to reconceptualization, both are self-correcting. Which is not a bad thing, because, as Sellars (SK I, §3) argues, we pursue philosophy "to become *reflectively* at home in the full complexity of the multidimensional conceptual system in terms of which we suffer, think, and act."

Although Sellars gives examples of piecemeal conceptual change in both images (within themselves), and envisages that the manifest image will eventually be replaced by the scientific image, he does not explicitly discuss how the radically different conceptual frameworks interact, i.e., how the scientific image blends into or infiltrates the manifest image. It is evident that our everyday conceptual practices do integrate certain components or processes of scientific frameworks. Today, atomic theory and microphysics are not unfamiliar concepts to the manifest image. Accepting O'Shea's (2007) suggestion about the categorial given and Brandom's interpretation of the non-inferentiality of the manifest framework (1997, pp 164–5), one could say that subatomic particles are part of the manifest image of a skilled atomic physicist, since they can be directly detected by their traces in cloud chambers.

3 Explication as the Middle Road

We anticipate that Carnap's *explication* could be viewed as a helpful and enlightening process for solving this problem. Carnap introduced explication as a method in some detail in *Meaning and Necessity* in 1947:

> The task of making more exact a vague or not quite exact concept used in everyday life or in an earlier stage of scientific or logical development, or rather of replacing it by a newly constructed, more exact concept, belongs among the most important tasks of logical analysis and logical construction. (Carnap 1947, pp 7–8)

The old, inexact concept is called explicandum, the new and supposedly more exact concept is the explicatum.

Explication involves three major steps: (a) clarifying the explicandum, (b) defining the explicatum, and (c) using the explicatum to determine its range of application.

From the late 1940s, Carnap considered this to be *the most important task of scientific philosophy* and offered multiply examples from the history of logic and philosophy (for a discussion as to why this notion is still not unambiguous, what it is and how it is supposed to work, see Reck 2024). For Carnap, Tarski's semantic theory of truth, the Frege-Russell idea for dealing with numbers in terms of sets and attempts at approaching definite descriptions are examples of explication. His own philosophical projects between the 1940s and 1970 were examples of explicating such notions as meaning and necessity, probability, analyticity, confirmation, and entropy.

Another often-cited example of Carnap is the explication of the comparative notion of "warmer." This seems to be a somewhat vague, and subjectively determined, phenomenological concept, which leaves a lot of room for error. However, when we introduce the notion of "temperature," we can say that "x is warmer than y" if "x's temperature is higher than y's." This is a more exact, numerical notion, which fits into Carnap's (1950, p 7) "well-connected system of scientific concepts."

While explication could be seen as a familiar process, it is, in fact, not identical to either explanation, definition, or regular (conceptual) analysis. While explanation is itself a notion that has to be explicated (à la Hempel), definition and analysis to some extent require that the concept we start with and the concept we arrive at have the same meaning, if this is to be considered a valid process at all. Preservation of meaning and application (co-intensionality) and co-extensionality are requirements for analysis, but not for explication. Moreover, while we can talk about the rightness of definitions and the soundness of analysis, these are unimportant, or even unfeasible, notions in the case of explication. As Carnap (1950, p 4) says, "since the datum [the explicandum] is inexact, the problem itself is not stated in exact terms; and yet we are asked to give an exact solution. […] It follows that […] we cannot decide in an exact way whether [a proposed solution] is right or wrong."

But how, then, are we to judge or evaluate an explication? Carnap famously lays down four requirements:

1. The explicatum has to be *similar* to the explicandum.
2. The characterization of the explicatum has to be *exact*.
3. The explicatum has to be *fruitful*.
4. The explicatum has to be as *simple* as possible.

(1) *Similarity* amounts to the idea that in *most* cases where the explicandum has been used, the explicatum must be used as well, and in *most* cases where the explicandum has not been used, the explicatum shall not be used either.

---

[10] In DeVries' (2005, 272) succinct formulation: "Every conceptual framework necessarily has a prescriptive or normative dimension. Indeed, conceptual frameworks are *essentially* and *principally* normative in their being, for a conceptual framework is *constituted* by the valid inferences, both formal and material, and by the proprieties of response and behaviour that are licensed by it.".

As Carnap (1950, 7) notes, "close similarity is not required, and considerable differences are permitted." To reiterate, explication is not the same as analysis.

(2) *Exactness* concerns the description or characterization of the new concept in such a way that the explicatum can be introduced into "a well-connected system of scientific concepts," or as Carnap (1963, 936) says elsewhere, "exactness and clarity are best achieved by a certain degree of systematization." Thus, embeddedness in a system has the virtue that the correct use of the explicatum in ambiguous cases can be determined. Carnap does not offer much detail about how to conceive of exactness, but if the problem with the explicandum is its broad, perhaps even inconsistent usage, then by restricting its range we are getting a more exact concept, in the sense that we can always decide whether *x* falls under the explicatum or not.

(3) The most Carnap (1950, p 7) has to say about fruitfulness is that the new explicatum should be used "for the formulation of many universal statements (empirical laws in the case of a nonlogical concept, logical theorems in the case of a logical concept)."

(4) Simplicity is taken to be relative against the background of (1), (2), and (3).

Most of these requirements are not at all unproblematic. Quite obviously, most of Carnap's examples and his own practice come down to logico-mathematical problems, where exactness amounts to a type of formalization. However, other examples (like that of fish and Pisces, salt and sodium chloride, "warmer" and "having a higher temperature") emerge from the sciences and everyday contexts that are less sensitive and open to logico-mathematical formalization (on this, see Reck 2024). This causes a kind of general uncertainty as to the scope of exactness, and raises the possibility of different forms and versions of exactness. Carnap (1950, p 7) also states explicitly the idea that *exactness* comes with the integration into a "well-connected system of scientific concepts." Systematicity thus plays a major role here, while he was quite obviously less concerned with what this would look like, what kind of system this would be, how to choose between systems, and so on. These seem to be broader issues that need to be answered from a different perspective. *Simplicity* is also a relative concept, marred by the same problems as exactness.

Similarity and fruitfulness are even more interesting. "Similarity" is a tricky notion without any explicit or exact definition and rules. How could we define it and until what point two *distinct* concepts are similar? What we can do, however, is to define clearly what should be explicated at all. Carnap was aware of this and acted cautiously. He saw the problem that much traditional and contemporary philosophy is not particularly concerned about origins. "There is a temptation to think that, since the explicandum cannot be given in exact terms anyway, it does not matter much

how we formulate the problem" (Carnap 1950, p 4). Nothing could be more wrong for Carnap. He thought that in order not to lead "entirely futile" discussions, we must do "all we can to make at least *practically clear* what is meant as the explicandum" Otherwise, we would be researching "What is *x*?" type questions at such a general and abstract level that this would not have any bearing either on science or everyday life.

What, then, can we achieve in this regard? To avoid wishful thinking, what Carnap envisions is to make "at least practically clear" what we mean by a certain concept. "Practically clear" amounts to the idea that one should be able to correctly predict the usage of the concept in simple cases. As Carnap (1950, p 4) says, "what *X* means by a certain term in contexts of a certain kind is at least practically clear to *Y* if *Y* is able to predict correctly *X*'s interpretation for most of the simple, ordinary cases of the use of the term in those contexts." Dealing with inexact terms makes the whole business shaky, of course, but Carnap thinks that there are countermeasures we can take; after all, what he aims for is "a relatively good mutual understanding as to th[e] intended meaning" of the concepts. One option is to provide "some examples for its intended use and other examples for uses not now intended," with some informal explanations. All this circling around helps to state the problem, but not to solve it in any way. *Consequently, there is an obvious relevance of our everyday life and understanding when we embark upon explication.*

Carnap's example is the word "true." If we are interested in the term "true," we first need to categorize and taxonomize its various uses in simple cases. We talk about a "true friend" or a "true democracy," meaning something like "genuine," "supportive," or "faithful." But there are other meanings of "true" that amount to "correct," "accurate," "not false," or "veridical" if applied to statements, assertions, reports, and stories. This is not explication yet, Carnap says, just a *necessary and obligatory* housekeeping task to make *practically clear* what to do where. Explication only starts when we take on Tarski's semantic approach to the concept of "true."

This process of clarifying the initial map of concepts is extremely important for identifying the starting point of explication and acknowledging any possible *similarity* with the outcome. Carnap devotes an entire section of *Logical Foundations of Probability* (§2) to this problem, and emphasizes repeatedly that philosophers err when they, from the very first moment, offer their own definitions of the concepts in question. Instead, they should be checking at the beginning whether a problematic concept is "practically clear enough to serve as a basis for an investigation, for an analysis or explication" (Carnap 1950, p 4). Similarity is thus both a problematic and highly fruitful concept and criterion for Carnap; problematic in the context of formalization

and exactness (a formalized, more exact concept will have a more restricted usage than the previous one it replaces). But as Carnap notes, the whole discussion becomes "entirely futile" if we do not anchor inexact concepts so that they can be explicated within everyday usages and concerns. For to do so could be of relevance in comparing them, especially with such concepts as "use" and "function," taken by Sellars from Wittgenstein's vocabulary. Even if Carnap seemingly downplays the importance of similarity for theoretical reasons, he still leaves open the door to practical and ordinary considerations within his philosophy (this has basically been done by Carus (2007) regarding the Enlightenment aspect of Carnapian explication).

Despite all the difficulties, what makes things a little easier for Carnap is that explication is an open-ended process, whereas exactness, simplicity, and even fruitfulness are comparative concepts that come in degrees, meaning the process can be iterated, and the former explicatum could become the new explicandum. It is a *transformation*, as Carnap (1950, p 1) says. We are thus not creating the explicatum out of nothing, but changing, transforming an *already given* concept into a new one, always against a sophisticated background. In his answer to Strawson, Carnap (1963, pp 934) even claims that while Strawson draws a rigid boundary between everyday and scientific concepts, he himself sees "here no sharp boundary line but a continuous transition." The everyday background thus functions like a constant guide and judge, keeping track of the process to prevent us from going astray. As Carnap (1963, pp 936) says, during explication, the explicatum is "intended to take the place of the explicandum, and that means, of course, that it is to be used for the same purpose as the explicandum." Speaking of the "same purposes," "ordinary concepts," and comparisons between different issues all points to the practice of ordinary language philosophy—broadly speaking, the field of the manifest image.

Carnap does not have, of course, any concrete theory or approach towards the ordinary, problematic context from which an explication originates. He was not much interested in that work, and plausibly hoped that others would do it. In *Logical Foundations of Probability*, Carnap points to Arne Naess' then forthcoming book on *Interpretation and Preciseness* (1953), where Naess defines such concepts as "preciseness" and "formulation," and develops methods to measure ambiguity, vagueness, and similar notions that are vital to any housecleaning prior to explication, to the decisions taken during explication, and to evaluating the process after explication. But Naess was well-known also for his work on the term "true": he interviewed hundreds of people on the street, inquiring about their intuitions and definitions of "true," "truth," and similar notions (Naess 1938; Chapman 2011). Naess' work was empirical, focusing on how to make practically clear what ordinary people mean

by those concepts. While we are skipping over the details here, suffice it to say that Carnap was not unreasonable to presume that such pre-explication work could be done, and had good examples for it at hand (especially given that Naess had some role in the semantics debates between Carnap and Neurath, hence Carnap knew his work firsthand from conversations and conference roundtables, even referring to Naess' systematic "questioning of people" in *Introduction to Semantics*, 1942, p 29). One could also cite Carnap's own work on the concepts of probability (especially in Carnap (1947), intended, with some practical application in mind, as "a guide in life") and meaning (Carnap 1955), or even Hans Reichenbach's (1947) then recent attempt to formalize natural languages with logical tools, by fine-tuning and making them more exact.

But what about fruitfulness? Reck and Dutilh Novaes (2017, p 202) identify "fruitfulness" as "ultimately the most significant requirement for an explication overall." Speaking about justifying decisions in Carnap's philosophy, Sellars (1963b, pp 433–434) notes that "as to the nature of such justification, however, [Carnap] gives no more than a few obscure hints. One looks in vain for an unpacking of 'expediency,' 'fruitfulness,' and 'conduciveness to the aims for which the language is intended.'" We hope to show that perhaps Sellars has something to offer on this and many other perplexing questions, for example concerning the relation of frameworks to explication (Carnap rarely discusses them explicitly together, but see below), or how to engage with explication within frameworks. Obviously, as each framework is defined by the rules governing the usage of its concepts, explicating these *concepts* would change the rules that govern the concept, and many other concepts that are related to it by the rules. Thus, in the end, we are not just explicating concepts atomistically, but always have to deal with some part of the framework as well—something which is again not discussed by Carnap, but is important for Sellars.

4 Conceptual Change: Internal and External Questions of "Language Engineering"

For Carnap, of course, explication is not identical with the choice between frameworks, and as he seldom links the two notions, the connection between them is somewhat ambiguous. He takes the *choice* between conceptual frameworks to be a purely external question, and consequently the question "Shall we introduce such and such forms into our language?" turns out to be a practical question, "a matter of decision rather than assertion" (ESO, 29). In contrast, as explication is not a question of introducing or replacing an entire *conceptual framework*, it is tempting to treat it as an internal question within a framework, for example when a scientific concept is being made more exact (when moving, for

example, from temperature to mean kinetic energy and molecules). Nonetheless, in graduating from an entirely ordinary concept to a scientific one, explication may have a more external and pragmatic flavor, for instance, when deciding to explicate "warmer" in terms of "having higher temperature." This seems to be due to the pragmatic choice (favoring certain values, like systematicity, quantitative issues, or replicability) that an ordinary concept should be explicated via a scientific framework. Indeed, according to André Carus, it must be somewhat external, otherwise it could not be called *conceptual engineering*. As he writes, "This choice among precise explications cannot, therefore, be settled internally within the language of any particular precise explication. It is what Carnap called an 'external' problem (and Plato a 'dialectical' one)" (Carus 2004, p 329). In other words, "explication, in contrast, is a process external to the precise target language" (Carus 2007, p 279). According to the criteria set out in the previous section, we can suspect that it is not even possible to take this to be an internal question, as Carus and Howard Stein, a student of Carnap, acknowledges:

> The explicatum, as an exactly characterized concept, belongs to some formalized discourse—some 'framework'. The explicandum … belongs ipso facto to a mode of discourse outside that framework. Therefore *any* question about the relation of the explicatum to the explicandum is an 'external' question; this holds, in particular, of the question whether an explication is adequate—that is, whether the explicatum does in some appropriate sense fully represent, within the framework, the function performed (let us say) 'pre-systematically' by the explicandum. (Stein 1992, p 280)

And this is problematic because the criteria of explication—at least as Carnap introduces them—seem to be internal rather than external requirements. The *exactness* of explication presupposes a "well-connected system of scientific concepts," i.e., an explication is successful if we can replace an old concept with a new one *within* a given conceptual system or a transitory one, by encapsulating different stages according to specified rules. Carnap's formulation of *fruitfulness* as the "formulation of many universal statements" also suggests an internal interpretation (not to mention *simplicity*). The initial meaning identification also seems to be an internal issue (although in our proposal it can be read as an external one). The real problem for an internal reading is *similarity*, but we believe that an external—Sellarsian—reading of explication—which is also consistent with the discussion above—makes this and the other criteria less problematic. We also believe that Carnap's explication is well suited to the conceptual change that Sellars talks about in the context of the successive conceptual frames of the images (although explicitly only of the scientific image).

Explication, in Carus' (2004, p 329) interpretation, is "neither purely cognitive nor purely practical," but "an engineering matter that requires the mutual adjustment of cognitive and practical. Our knowledge shapes our values, and our values shape our knowledge," where knowledge belongs to the cognitive domain and values to the practical. Regarding inductive methods, for example, Carnap (1952, p 7) writes that "the choice of an inductive method [from the multiplicity of alternatives] is merely a matter of whim," but can be "more or less adequate" (just like an explication), so that the concept in turn "would then not be purely theoretical but rather of a pragmatic nature." Sellars would agree with this assessment—namely that the choice between conceptual frameworks cannot be purely cognitive, but must also be an (even essentially) pragmatic one. He argues that in addition to our cognitive-explanatory interests, accepting a theory, or adopting a conceptual framework, is always a pragmatic choice—whether it is worthwhile or reasonable to accept (IV). That is, Carnap and Sellars seem to agree on this point of securing a role both for cognitive and practical considerations in conceptual engineering and linguistic improvement. In fact, depending on where the emphasis is placed, their respective discussions might be helpful for understanding how a more robust Sellarsian pragmatism could reinforce the decision-based character of Carnapian engineering, and how Carnapian technical formalism (aimed at codifying cognitive issues instead of pragmatical components) might fortify Sellarsian pragmatism.[11]

Formalism can be useful when we are analyzing, when we are working within a conceptual framework, but it is not particularly useful when we have to choose between radically different conceptual frameworks. What should motivate our choice? Even if the choice is a "whim," and even if Carnap acknowledges that it is a practical one, he does not say much about it:

> The introduction of the new ways of speaking does not need any theoretical justification, because it does not imply any assertion of reality. ... To be sure we have to face ... an important question; but it is a practical, not a theoretical question; it is the question whether or not to accept the new linguistic forms. (ESO 31)

For Carnap, normative issues are apparently intractable because they are inaccessible to exact and formal methods from an empiricist point of view.[12] However, we must realize

---

[11] Note that "formalism" could mean various things in general, and even within the Carnapian corpus. Thus we use the term here only as a guiding idea stemming from general historiographic accounts. On Carnapian formalisms, see Schiemer (2024).

[12] This does not mean, however, that normativity and rationality are entirely bogus to Carnap. Especially in his late (unpublished) writings of the 1950s and 60s, he devoted some energy to spelling out a framework of value functions, inductive intuitions, and similar notions to

that formal and definitional activities are incapable of reaching outside the given conceptual framework. We may say that it is reasonable and rational to accept a new framework, but presumably we mean only that we should accept it (see Sellars in IV). The "normative force" at play in such a decision is definitely not the *internal conception of normativity* that Olen (2016, Ch. 6) associates with Carnap. It is here that Sellars and his later conception of rules—what Olen (2016) calls the *external conception of normativity*—can be of assistance to Carnap. According to this conception, rules, whether syntactic, semantic, or inferential, do not function simply as definitions, but require reference to a doing or action.[13] Nevertheless, the choice of a framework is not an innocent "choice of instrument." The chosen framework also carries an ontology within it. While the choice itself is not motivated by an ontology or an inescapable, unchangeable "given" framework, an external conception of normativity requires external reference points, practical values, and goals. If we take the *fruitfulness* criterion of Carnapian explication to be non-internal, we should be able to say why one framework is more fruitful than another.

But what happens if we nonetheless attempt to give an internal reading to the criteria of explication? In this case, insofar as accepting the framework is tantamount to accepting an ontology,[14] we do indeed arrive at an analytic assertion within the framework according to which the framework itself has no empirical content (EAE 432–33). By explicating a new concept, we incorporate it into the existing rules of our language; we define it by these rules. Explication will then have become essentially analytic (its truth doesn't depend on empirical facts), as far as the basic objects of the framework are concerned. To say that the basic entities of a framework exist does not contribute much to our knowledge (which is already framed by these very concepts). These assertions, taken as statements within the framework they are about, cannot be interpreted in any other way than

tautologically. To say—within the framework of atomic theory—that there are atoms is merely a restatement of the basic categories of atomic theory, and does not constitute useful knowledge. They become significant and meaningful only when they are understood as the rejection of the framework itself, that is, as statements external to the framework. In this sense, Sellars (EPM §41) is prepared to say, as a "philosopher," that there are no coloured physical objects. This rejection of the framework of the manifest image is constructed from another, scientific and realist, framework. Normally, it is the other way round: scientific frameworks operate, develop, and are evaluated within the context of everyday manifest frameworks. The manifest framework (which is fundamentally based on perceptible objects and properties) determines the (in the case of science, explanatory) goals that serve as an external reference point for the choice of framework. "It is the rock bottom concepts and principles of common sense which are binding" the scientist (SRI 355).[15] The manifest framework is therefore methodologically prior to the scientific. It is the framework into which "direct noticing" is built. Although it is in no sense unchangeable or replaceable—meaning it is not given—it serves as a metaconceptual framework for the others.[16] And a scientific framework "can achieve first-class status only if a proper subset of its expressions acquires a direct role in observation" (SRI 363). In other words, reconceptualization is not an internal question; it necessarily appeals to a meta-framework.

To elaborate on this, let us recall the following distinction advanced by Sellars:

[A] statement is analytic₁—analytic in the broad sense—if it is 'true (or false) *ex vi terminorum*,' if, that is to say, *given that the reasonableness of using the language to which it belongs is not in question*, the statement does not require, indeed it would be a mistake to give, a justification in terms of observation. In this sense both "2 + 2 = 4" and the fundamental

---

track decisions across frameworks and practical actions. Although there may be important issues to consider in the context of induction, notably regarding its justification and background, this has to be reserved for a later occasion. For now, see Carus (2017).

[13] Although Carus (2004) extensively criticizes this conception of rules, arguing that it is alien to the logical empiricist tradition, and especially to Carnap, it is rather similar to the conception of Morris (1938), who states just that in the *Encyclopedia of Unified Science*, namely that semantic and syntactic rules are dependent on pragmatic ones.

[14] This seems to be supported by Carnap (ESO, 23) as well, when he notes that in certain frameworks, it would be true to say that numbers exist, while in others it would not, or that "if someone decides to accept the thing language, there is no objection against saying that he has accepted the world of things. […] To accept the thing world means nothing more than to accept a certain form of language […].".

[15] With this metaphor, Sellars (SRI 353) does not provide a remedy for givenists, as he writes: "to reject the myth of the given is not to commit oneself to the idea that empirical knowledge as it is now constituted has no rock bottom level of observation predicates proper. It is to commit one self rather to the idea that even if it does have a rock bottom level, it is *still* in principle replaceable by another conceptual framework in which these predicates do not, *strictly speaking*, occur. It is in this sense, and in this sense *only*, that I have rejected the dogma of givenness with respect to observation predicates.".

[16] The common-sense conceptual framework (the manifest image) is prior *in the order of knowing* to concepts pertaining to the scientific image, though *in the order of being* it is not. "[T]he conceptual space of common sense physical objects is underived, their content qualities must be directly rather than analogically conceived, for it is only in terms of perceived, and therefore conceptualized, qualitative difference that form and structure can be distinguished" (SRI 356).

principles of, say, molecular theory are analytic. Correspondingly, a statement is synthetic$_1$ if—again given that the reasonableness of using the language to which it belongs is not being challenged—it is appropriate to justify the statement by an appeal to observational evidence. In this sense neither arithmetical statements nor the fundamental principles of molecular theory are synthetic. … A statement is analytic$_2$ if it is analytic$_1$ and if the non-logical or descriptive terms it contains either occur vacuously, or if they occur vacuously in the statement one gets by replacing definable terms by their definitions. In this second sense of "analytic," "2 + 2 =4" is analytic, but the fundamental principles of molecular theory are not. Indeed, these principles (which are analytic$_1$) and "There are 10^6 molecules on the point of this pin" are alike synthetic$_2$. And surely the coincidence of the empirical with the synthetic$_2$ yields a sense of closure. (EAE 438–9)

According to this formulation, a fruitful and meaningful explication must be a process that aims at an analytic$_1$ (and synthetic$_2$) result, but which obviously cannot be achieved in an analytic$_2$ fashion when empirical questions and explanations are on the table. For Carnap, who is predominantly interested in artificially constructed logico-mathematical languages, this is not a very interesting meaning of explication. Everything he writes explicitly about explication is consistent with this internal interpretation. Sellars, on the other hand, is more interested in analytic$_1$ and synthetic$_2$ terms, because they are emblematic of the "changes of concepts" that take place in the course of a scientific development:

> The point stands out most clearly in the case of the evolution of a scientific theory. Here it makes obvious sense to say that a certain concept belonging to the theory at one stage is a development of a concept belonging to the theory at an earlier stage. (SM 133)

Such conceptual changes must be fundamentally normative. When we revise the categories of a conceptual framework, the appropriateness, usefulness and, of course, necessity of the revision is a perfectly legitimate question. But these questions must be formulated and answered from outside this conceptual framework itself—and naturally in a meta-framework. Conceptual frameworks are constituted by rules (analytic$_1$ and synthetic$_2$, as well as analytic$_2$ rules). These rules do not merely regulate an agent's behavior; they are the prerequisites for identifying problems (by establishing elementary categorial entities and relations) and the procedures for resolving those problems. In a sense, these rules provide us with the means to explain and justify our choice of framework. However, this justification must be "framework external", in that it must be grounded in aims and objectives external to the chosen framework.

As can be seen from the above, for Sellars, conceptual frameworks and theories are not simply instruments to be used or not to be used. Their modification or choice is not entirely arbitrary—it is always motivated by and can only be done within a meta-framework. Even mathematical-logical frameworks have to fit into non-mathematical, non-logical frameworks that are connected to the everyday manifest framework. This framework has no meta-framework, meaning we cannot retreat to safer ground, since the basic objects and predicates of this framework are directly given (SRI 339). And this doesn't make it an easy task to distinguish between analytic$_1$ (but synthetic$_2$) and analytic$_2$ in the common-sense framework, since both are distinct from synthetic$_1$, which "appeals to observational evidence." Therefore, the reasonableness of the basic principles of the framework cannot be easily addressed.[17]

This aspect (the meta-framework from which the choice can be made) is completely missing from Carnap's discussion. By adding it to explication, we can fill the gaps he left. Bearing in mind the difference between the rules that guide the moves within the framework and the rules that set the framework, we can divide the criteria for explication into two groups. Some are *internal*, since they concern the internal structure of a framework, such as *simplicity* or *exactness* (as Carnap conceives it); and others seem to be external, such as *similarity* (if identity is excluded) and *fruitfulness*—at least this is what Dutilh Novaes and Reck (2017, p 199) propose.

What are the criteria for similarity? Carnap doesn't have much to say about this, except that close similarity is not needed for explication. But how can we ensure that the explicatum is similar to the explicandum, without being identical? How does the degree of similarity remain high enough without there being a one-to-one correspondence between the two? An easy suggestion would be that they talk about roughly the same phenomena, or that they are interchangeable in the relevant cases. Pragmatically speaking, the question is whether the explicatum can indeed be used for the same (or at least sufficiently similar) purposes as the explicandum. However, these formulations do not define a formally exact criterion, which is bad news for Carnap. And although Dutilh Novaes and Reck (2017, p 202) classify *exactness* as an internal criterion, it actually leads to very similar problems as similarity: "It is clear that, with each iteration of the explication process, we obtain an increase in exactness. But at each step, something is 'transformed' as well, resulting both in gains and in losses." And this is

---

[17] This distinction can be approached from a different angle by appealing to Sellars' (LTC 508) related distinction between *ought-to-be* rules (a.k.a. *rules of criticism*) and *ought-to-do* rules (a.k.a. *rules of action*).

precisely what makes the notion of exactness so difficult: it cannot be the same, it must exclude something from a previous, imprecise notion, i.e., it must classify certain applications as illegitimate. The freedom of choice we have, the principle of tolerance, applies only to the extent that the choice of framework is not strictly limited by cognitive requirements.

The issue at hand is that if the explicatum and explicandum lie in different frameworks, they become almost impossible to relate to each other. It is therefore not feasible to establish a proper connection between different and incommensurable relata. Nevertheless, we believe that Sellars' conception may offer a viable solution to this predicament. According to Sellars, meaning is not a relation, and meaning statements are not relational statements. Thus, statements, such as "'Und' (in German) means *and*," are not relational statements. They contain *distributive singular terms* on the left, which denote some—or all—utterance tokens of "und" (not abstract singular terms, which denote abstract entities), and *illustrating sortals* on the right, which illustrate the function of "und" in German by presenting an example that—being familiar to the listener and speaker—has a similar function. This illustrating sortal is meant to rehearse a set of linguistic rules that govern the use of an expression (concept) with which the hearer is already familiar. To designate this set of rules of inference, Sellars introduces the dot quotation. In his account, both sides of the meaning statement refer to natural linguistic objects,[18] "means" becomes nothing more than a "specialized form of the copula" (CC §32). Thus, understanding the meaning of an expression is not a *knowing what* but a *knowing how* type of knowledge. As such, the meaning statement mentioned above will be identical to the following statement: "'Und's (in German) are ·and·."[19] In this way, the identity of concepts belonging to different conceptual frameworks does not require a reference point external to both frameworks. Their "identity means sameness of functions, and belongs in a continuum with similarity of function" (CC §43). The explicandum and the explicatum are similar because their functional and inferential roles are similar. This is an easy workaround when the two frameworks are similar in some relevant sense, like Riemannian and Euclidean geometry (CC §51) or classical

and relativistic kinematics (SM V. 45). But what can we say about radically different frameworks? First, similarity can be established by appealing to *structural similarity*, for instance when Sellars talks about moments in a temporal series being similar to points in a line because they "have *second-order* properties in common, e.g., transitivity (SRI 346). Although the first-order properties of moments and points are rather different, their second-order properties, or at least some of them, are identical. The same can be said of mass, where certain second-order properties have been preserved from Aristotle to Einstein. We can apply this structural similarity to dissimilar concepts as well. Although Eddington's two tables differ considerably in their first-order properties, they share certain second-order properties. Some or even most of the spatio-temporal relations to other objects or aggregates in their respective environment may be identical, and it is possible that the same holds true for their causal connections. Second, even if the change is not constrained epistemically, it must be constrained by practical factors, usually external to the chosen framework, but conceptualized in another, predominantly pragmatic framework. When we consider the exactness of concepts, or the exactness of the theories which these concepts (explicatum) belong to, we always have to evaluate them in a broader framework. Theories which are essentially concerned with description and explanation can only be assessed in a framework they are embedded (their "external" meta-framework) which itself is always predominantly pragmatic, i.e. imbued with values and ends. So, the comparison can be based on a common and familiar framework, either because the frameworks in question are related to the common-sense conceptual structure, or because one of the frameworks in question is itself the common-sense framework. Since this common-sense framework is "the rock bottom," in this particular case of conceptual change or explication (call it the situational-bedrock framework), it can always serve as a meta-framework for the comparison. Its constitutive rules (language entry/inference/ exit uniformities) provide a solid background for this decision. It is not possible to adopt a position outside all frameworks (the view from nowhere), for then we would have no concepts, no goals, and, of course, no basis for appealing to notions of correctness or adequacy.

Essentially, the similarity required by explication should be the similarity between the functional roles identified by the semantic rules. Clearly, such semantic rules will only be meaningful and significant if they are analytic$_1$, i.e., if they do not pertain to particular actions or situations (otherwise, they would not be general enough to form a bridge between conceptual frameworks); at the same time, they cannot be empty, meaning they must not be analytic$_2$, otherwise they could not be applied externally. Analytic$_2$ rules merely govern formal inferences; they do not determine the use or application of concepts themselves. Analytic$_1$ and

---

[18] Natural linguistic objects are essentially linguistic tokens viewed as physical structures, as "objects in the natural order." It is also worth noting that while we can rightfully say that dot-quotation refers to natural linguistic objects, it does not simply pick them out as physical entities, but rather relates to them in terms of their functioning in the wider system of linguistic objects (through their functional properties).

[19] Sellars' conception of non-relational meaning can be found in the vast majority of his writings, but for a detailed summary, refer to DeVries (2005 Ch. 2) or O'Shea (2007 Ch. 3 & 4).

synthetic$_2$ rules, however, often refer to or determine the structural roles and second-order properties, thus allowing the establishment of similarity.

The real test, however, is fruitfulness, which Carnap himself sees as the most fundamental criterion. Carus (2007) also raises the possibility that exactness should be subordinated to fruitfulness, since explication must ultimately be a purely pragmatic matter.) We believe that what was said about similarity and exactness above could also be stated in relation to fruitfulness: a new concept must be fruitful for a weighted majority of the purposes for which the concept being replaced was intended, but it must perform more functions and be better at them—as assessed from a meta-framework. In the case of scientific theories, these tasks and purposes will typically be predominantly descriptive and explanatory ones, which ultimately require feedback to observations—as parts of the manifest meta-framework—along the correlation rules of the theory. And this is precisely why Sellars (EPM §42) grants science the right to have the final word "in the dimension of describing and explaining the world."

Here, fruitfulness stems from the prospect of a greater explanatory power. What makes an explication fruitful is nothing other than the "explicatum's ability to connect with other concepts on the basis of observed facts" (Dutilh Novaes and Reck 2017, p 205). However, these other concepts must be "about" the "observed facts", which seems to conform to Sellars's position that every conceptual system or process of altering a conceptual system must be performed within another conceptual system. After all, the goal is to produce *new knowledge* about the phenomena to which the explicandum pertains, as is clearly visible from the "Pisces" example (ESO 6), where both "Pisces" and "fish" belong to a conceptual framework, albeit a different one. The latter is a concept of our most fundamental common-sense framework (the manifest image), and the only way to tell whether Pisces is about fish is to find out whether a link, a strong relation, is established *between* these concepts. In some instances, even Carnap is permissive about this embeddedness, not to mention Carnap scholars.[20] What Carnap opposes is the idea that natural language can serve as a universal and objective metalanguage for every explication. Sellars would profoundly agree with this, as evidenced by his rejection of the 'myth of the *categorial* given' (O'Shea 2007, 2021). However, this position is not incompatible with the idea that methodologically or situationally, there must always be a conceptual

framework in which the goals and problems that govern the alterations of conceptual frameworks are formulated.

To put it differently, there must be a *mismatch* between the explicatum and the explicandum. While this mismatch harbors the very possibility of gaining new information from an explication, it is also the source of the 'inherent paradoxity of the Carnapian explication':

> On the one hand, a particular formalization has to be sufficiently similar to its target phenomenon to be rightly described as a formalization *of that target phenomenon*, and also to be applicable to the *same, or at least closely related, purposes*. On the other hand, the formalization will be more useful insofar as it says something about the target phenomenon which prior, informal conceptualizations of it *did not reveal*. In other words, an adequate formalization is one that is faithful to the target phenomenon *and* reveals something new about it; there is an obvious tension between these two desiderata. We call this the *paradox of adequate formalization*. (Dutilh Novaes and Reck 2017, p 211)

The air of paradox can be cleared, however, if we abandon the relational conception of meaning (so vehemently defended by Carnap (1963)[21] against Sellars), and the idea that explication is a strict formal analysis, and instead adopt the proposition we briefly sketched out for similarity above; by appealing to the structure (the rules and the second-order properties) of conceptual frameworks, we are thus able to compare them (thereby ensuring that the concepts involved in explication roughly 'talk about' the same thing), even when the frameworks are significantly different. This appeal to the structural similarity of frameworks and to their embeddedness[22] in the manifest image, together with the directness of this—in a sense default, methodologically prior—framework, provides us with a tool for judging whether the criteria of explication are met or not. And this augmentation will be helpful not only in cases of science and empirical knowledge, but in many different conceptual

---

[20] Carnap (1963, 933–40) explicitly admits that constructed languages *are* embedded in evolved languages, although he regards this to be a contingent fact. Carus (2007, 292) shares this attitude when he speaks about the indispensability of some 'folk' categories "to human emotional and practical needs, to the progress of science, or to something else.".

[21] He writes: "I would not reject, as Sellars seems to do, all factual or descriptive relations between material objects and abstract entities, at least not if 'relation' is understood in the wide sense which is customary in modern logic. … [I]t seems to me that some psychological concepts may be regarded or reconstructed as relations (in the wide sense of the logical terminology, not in the causal sense) between a person and an abstract entity; e.g., believing may be taken as a relation between a person and a proposition (as is done by Church, comp. §9 VII), and thinking-of as a relation between a person and a concept (intension or sense) and the like" (Carnap 1963, 924–925).

[22] On the one hand, the *models* of scientific theory formation typically originate from the manifest image. On the other hand, the non-descriptive, non-purely theoretical explanatory objectives of scientific activity also trace back to the manifest image.

systems, since explication and fruitfulness need not be limited to scientific enterprises; even the nature of explanations may be open to debate in cases where the default framework (usually a version of the manifest image) has very different characteristics. Willem deVries (2005 Ch. 10, 2012) argues convincingly that the images do not only possess a cognitive reality but obviously also a practical one. The same idea can be brought into operation in the case of affective frameworks, in which the ultimate purpose of the framework is not to bring about an explanation but rather manipulative success. To illustrate this, let us imagine that in the past—or in the distant future, for that matter—some transcendent (religious) conceptual framework is integrated into the manifest image, and that no postulational science of any kind has emerged. In such a fictitious world, even the nature of the explanation might be different from ours, and any increase in explanatory power would be understood in a correspondingly different way. While we would not call this a scientific progress, it is nonetheless a conceptual change, and explication could therefore play a role in it.[23]

What would the initial Carnapian concept of explication appear to be under the suggested approach? Carnap was convinced that explication is concerned with individual concepts or, at the very least, with isolable parts of a conceptual system that can be replaced one by one, without affecting the rest of the framework. His examples seem to suggest this. Based on our Sellarsian approach, we can conclude the following: If the rules that belong to a concept, or the broader context of the concept in question, can be translated one by one into another—for example a part (or the whole) of the atomic-theoretical framework that treats physical objects as a set of atoms and molecules into a manifest framework that treats physical objects as fundamental entities—then there is no problem at all. It is a clear-cut case of explication in the formally correct fashion. What makes this possible, and to what extent it can be carried out, is determined solely by its fruitfulness, i.e., how much it can contribute to the fulfillment of the practical aims of the receiving framework (which in the case of introducing atomic theory into the manifest framework is presumably prediction). It is important to note, however, that for other practical purposes, the rules of the manifest framework still apply, i.e., the concept of physical objects is not completely replaced in all its contexts by the concept of molecular aggregate, the latter being present in the manifest framework only "instrumentally," as long as (some) scientific conceptual framework "could replace the common-sense framework in *all* its roles" (SM

V. 90), which is a corollary of Sellars' scientific realism. In other words, this approach to explication has limited utility, only serving to refine or improve a conceptual system from within, without altering the entire system. As a result, this type of "translation" is unsuitable for more extensive conceptual changes, such as cases where fundamentally distinct conceptual frameworks are linked through a relationship that Sellars (SM V) terms "conceptual successorship." These types of changes cannot be implemented internally, as they appeal to objectives and aims external to the framework in question.

The Sellarsized explication that we propose, while manifestly external, can never be applied from an absolute, omnipotent or locationless point of view. There must always be a more directly accessible fundamental (*in the order of knowing*) framework in operation throughout the process of explication, but it does not need to be irreversible or of unchallengeable epistemic authority. Its immediacy is ensured by the functional roles encapsulated in its language entry/inference/exit rules, which may be understood as "entirely natural, customary, instinctive, or 'second nature' to us, and so as typically taking place without any explicit higher-order cogitation concerning rules or justifications" (O'Shea 2007, pp 85–6). They are the situational-bedrock uniformities for all human activity.[24] Thus, we can also comply with the principle of tolerance, insofar as no framework is ontologically or metaphysically superior to any other, but there must always be a framework that is methodologically necessary to formulate the semantic-pragmatic statements that carry out the explication itself, even if that framework may change multiple times. There is nothing wrong with this, however, since this priority is only a methodological one. Take Carnap's discussion, again, of inductive methods. In his booklet on the topic, he draws a continuum of possible inductive methods for scientific inquiry, and argues for the view that the decision between them cannot be made from a God's eye point of view, but has to take into account the most diverse values and motivations:

> The adoption of an inductive method is neither an expression of belief nor an act of faith, though either or both may come in as motivating factors. An inductive method is rather an instrument for the task of constructing a picture of the world on the basis of observational data and especially of forming expectations of future events as a guidance for practical conduct. (Carnap 1952, p 55)

---

[23] There is an indispensable human or social aspect to explication acknowledged by Carnapians: "In a Carnapian 'language-engineering convention', we can imagine ourselves participating as human beings of flesh and blood—the people we actually are, rather than some abstraction in a hypothetical state of nature" (Carus 2007, p. 304).

[24] Nevertheless, these rules themselves and the uniformities they set up are always subject to "*normative assessment and rational appraisal*" (O'Shea 2007, p. 86). This is the essence of rejecting the "given.".

An instrument, similarly to a "saw or an automobile," can be changed at any time during its application. Whenever we are no longer satisfied with it, we can replace it, and choose from another available method (or framework, for that matter) that has been used and suggested by others. There is an abundant multiplicity of methods, concepts, and frameworks, that is also subject to explication—after all, "life is a process of *never ending adjustment*; there are no absolutes, neither absolutely certain knowledge about the world nor absolutely perfect methods of working in the world" (Carnap 1952, p 55, our emphasis).

In this stereoscopic vision, we see at once the explication proposed by Carnap, and the Sellarsian change of conceptual frameworks. During the process, one concept of a framework is replaced by another, even though these concepts—explicatum and explicandum—are not interchangeable and cannot be substituted for each other *salva veritate*. In other words, the explication of each concept must also entail the explication of a relevant part of the framework—or the whole framework—which has the explicandum at its core, since the explication does not only concern the concept, but also the rules (be they syntactic, semantic, or pragmatic) that bind the object of explication to the conceptual framework (in this context, we can mention the other two criteria of the explication as internally binding both the explicandum and the explicatum). Carnap demonstrates how explication is done, but he does not say much about the pragmatic considerations at work and the inevitable embeddedness of the explication process in a situational-bedrock framework. Sellars, however, does. For him, explication is a purposeful human activity that must follow certain rules. The criteria for explication should not be purely descriptive, as Carnap seems to imply. Sellars actually suggests that the aims and purposes of explication are pragmatic; for example, explanation is not sought for its own sake, but in order to make the world intelligible. If it is rational to adopt a theory, this always has something to do with our human activities. For instance, the scientific image has clear practicality, as argued by deVries (2005). Nonetheless, the pragmatic ends and goals need to be comprehensible to us, and this can only be achieved through a conceptual framework. Consequently, the process of explication and conceptual change is contingent on an additional conceptual framework, one in which the explication is embedded in a pragmatic way. And the conceptual framework, which is the situational-bedrock framework for most explications and conceptual transformations, is the manifest image of persons.

The two philosophers, whose views are often presented as opposing one another, have a great deal in common in terms of conceptual change or engineering—and even on issues where they do not agree, they are still very much able to lend each other a hand. We hope to have contributed to a convergence of their views by highlighting and analyzing some of the relevant aspects of their philosophies.

Declarations

References

Brandom RB (1997) Study guide. Wilfrid Sellars, and the Empiricism and the philosophy of mind. Harvard University Press, Cambridge

Brandom RB (2015) From empiricism to expressivism: brandom reads Sellars. Harvard University Press, Cambridge

Carnap R (1942) Introduction to semantics. Harvard University Press, Cambridge

Carnap R (1947) Probability as a guide in life. J Philos 44(6):141–148

Carnap R (1950) Empiricism, semantics, and ontology. Rev Int Philos 4(11):20–40 (**ESO**)

Carnap R (1952) The continuum of inductive methods. The University of Chicago Press, Chicago

Carnap R (1955) Meaning and synonymy in natural languages. Philos Stud 6(3):33–47

Carnap R (1963) The philosopher replies. In: Schilpp P (ed) The philosophy of Rudolf Carnap. Open Court, LaSalle, pp 859–1013

Carus AW (2004) Sellars, Carnap, and the logical space of reasons. In: Awodey S, Klein C (eds) Carnap brought home: the view from Jena. Open Court, Chicago, pp 317–355

Carus AW (2007) Carnap and the twentieth-century thought: explication as enlightenment. Cambridge University Press, Cambridge

Carus AW (2017) Carnapian rationality. Synthese 194:163–184

Chapman S (2011) Arne Naess and empirical semantics. Inquiry: Interdiscip J Philos 54(1):18–30

Dach S (2018) Sellars's two images as a philosopher's tool. Metaphilosophy 49(4):568–588

deVries WA (2005) Wilfrid Sellars. Acumen, Chesham

deVries WA (2012) Ontology and the completeness of Sellars's two images. Humana Mente 5(21):1–18

deVries WA (2016) Just what is the relation between the manifest and the scientific images? Int J Philos Stud. https://doi.org/10.1080/09672559.2015.1129740

Gabbani C (2018) Sellars and Carnap: science and/or metaphysics. In: Corti L, Nunziante A-M (eds) Sellars and the history of modern philosophy. Routledge, New York and London, pp 197–215

Lycan W (1995) Consciousness. MIT Press, Cambridge

Morris CW (1938) Foundations of the theory of signs. International Encyclopedia of Unified Science I (2). University of Chicago Press, Chicago

Naess A (1938) 'Truth' as conceived by those who are not professional philosophers. Jacob Dybwad, Oslo

Naess A (1953) Interpretation and preciseness: a contribution to the theory of communication. Jacob Dybwad, Oslo

Novaes CD, Reck E (2017) Carnapian explication, formalisms as cognitive tools, and the paradox of adequate formalization. Synthese 194:195–215

O'Shea J (2007) Wilfrid Sellars. Naturalism with a normative turn. Polity Press, Cambridge

O'Shea J (2021) What *is* the myth of the given? Synthese 199:10543–10567

Olen P (2016) Wilfrid Sellars and the foundations of normativity. Palgrave Macmillan, New York

Olen P (2017) A forgotten strand of reception history: understanding pure semantics. Synthese 194(1):121–141

Reck EH (2024) Carnapian explication: origins and shifting goals. In: Richardson A, Tuboly AT (eds) Interpreting Carnap: critical essays. Cambridge University Press, Cambridge, pp. 127–149

Reichenbach H (1947) Elements of symbolic logic. The Free Press, New York

Rorty R (1997) Introduction. In: Sellars (1956/1997), pp 1–12

Rosenberg JF (1982) The place of color in the scheme of things: a roadmap to Sellars's Carus lectures. Monist 65(3):315–335

Schiemer G (2024) Carnap on the formality of logic and mathematic. In: Richardson A, Tuboly AT (eds) Interpreting Carnap: critical essays. Cambridge University Press, Cambridge, pp. 171–191

Sellars W (1951) Some reflections on language games. Sellars, Wilfrid, science perception and reality. Routledge and Kegan Paul, London, pp 321–358 (**SRLG**)

Sellars W (1954) Physical realism. Philos Phenomenol Res 15(1):13–32 (**PR**)

Sellars W (1962) Philosophy and the scientific image of man. In: Colodny RG (ed) Frontiers of science and philosophy. University of Pittsburgh Press, Pittsburgh, pp 35–78 (**PSIM**)

Sellars W (1963a) Phenomenalism. Sellars, Wilfrid, Science Perception and Reality. Routledge and Kegan Paul, London, pp 60–105 (**PH**)

Sellars W (1963b) Empiricism and abstract entities. In: Schilpp P (ed) The philosophy of Rudolf Carnap. Open Court, La Salle, pp 431–468 (**EAE**)

Sellars W (1964) Induction as vindication. Philos Sci 31(3):197–231 (**IV**)

Sellars W (1965) Scientific realism or irenic instrumentalism: a critique of Nagel and Feyerabend on theoretical explanation. Sellars W Philosophical Perspectives. Charles C. Thomas Publisher, Springfield, pp 337–369 (**SRI**)

Sellars W (1968) Science and metaphysics: variations on Kantian themes. Routledge and Kegan Paul, London

Sellars W (1969) Language as thought and as communication. Philos Phenomenol Res 29(4):506–527

Sellars W (1970) Towards a theory of the categories. In: Reidel D (ed) Sellars, Wilfrid: Essays in philosophy and its history. Dordrecht, pp 318–339 (**TTC**)

Sellars W (1971) Science, sense impressions, and sensa: a reply to Cornman. Rev Metaphys 24(3):391–447 (**SSIS**)

Sellars W (1973) Conceptual change. Sellars, Wilfrid: essays in philosophy and its history. D. Reidel, Dordrecht, pp 77–93 (**CC**)

Sellars W (1975) The structure of knowledge. In: Castañeda H-N (ed) Action, knowledge, and reality. The Bobbs-Merrill Company, pp 295–347 (**SK**)

Sellars W (1982) Sensa or sensings: reflections on the ontology of perception. Philos Stud 41(1):83–111 (**SSOP**)

Sellars W (1956/1997) Empiricism and the philosophy of mind. Harvard University Press, Cambridge (**EPM**)

Sicha J (2002) Introduction. In: Sicha J (ed) Kant's transcendental metaphysics. Atascadero, Ridgeview, pp 1–260

Stein H (1992) Was Carnap entirely wrong, after all? Synthese 93(1/2):275–295

van Fraassen BC (1999) The manifest image and the scientific image. In: Aerts D (ed) Einstein meets magritte: the white book—an interdisciplinary reflection. Kluwer, Dordrecht, pp 29–52