Stathis Psillos & Demetra Christopoulou

THE A PRIORI: BETWEEN CONVENTIONS AND IMPLICIT DEFINITIONS*

1. Introduction

A thumbnail sketch of the philosophical thinking about the a priori would surely include that it has been dominated by two major approaches: the Kantian absolute conception of it and the Millian-Quinean absolute rejection of it (section 2). Yet, one can find in the literature claims about the existence of a ›functional a priori‹, a ›relative a priori‹, a ›relativised a priori‹ and suchlike. They are all meant to carve a space between the two extremes. An important thought behind the search for a middle ground is that the supposed coincidence between the constitutive and the unrevisable is wrong. The entitlement to accept a principle as being constitutive of experience prior to any empirical justification of it is compatible with an entitlement to revise or abandon such a principle on empirical grounds.

If a priori principles are meant to be independent of experience, how should this claim of independence be understood so that room is left for the possibility that a principle is *both* independent of experience and revisable on empirical grounds (section 3)? A straightforward and natural way to approach this issue is to think of constitutive principles along the lines of Poincaréan conventions, which can be seen as delineating a new sense of the a priori – the *conventional* a priori principles. These are substantive principles that are constitutive of theoretical frameworks – in the sense that they define (or constitute) the object of knowledge – without being either synthetic a priori or empirical generalisations. Still, their negation is conceivable and they are revisable on empirical grounds (section 4).

* Earlier versions of this paper have been presented at: the workshop ›Implicit Definitions and A Priori Knowledge‹ in the Free University of Berlin; and seminars in the Dept of Philosophy, London School of Economics, the Dept of philosophy, Bilkent University (Ankara) and the Dept of Philosophy and History of Science, University of Athens. Many participants in the above have helped us with useful comments and criticism. We want to thank them all – but especially: Samet Bagce, Andreas Berg-Hildebrand, Nikola Kompa, Christian Nimtz, David Papineau, Yiannis Stephanou, Christian Suhm, Stelios Virvidakis, Tim Williamson and John Worrall. Demetra Christopoulou's research has been funded by a postdoctoral fellowship from the Greek State Scholarships Foundation.

This conception of conventions (or ›definitions in disguise‹) is developed along the lines of implicit definitions (section 5). But there is a drawback, namely that for implicit definitions to capture a sense of apriority they must be non-arrogant: they should be such that they are not subject to empirical confirmation. There has been an ingenious way to avoid arrogance that goes back to Rudolf Carnap's early work on reduction sentences and his claim that implicit definitions should be conditional. Carnap's approach, based on Carnap-sentences as implicit definitions, is superior to Poincaré's in that it offers a clear way to decompose a theory into two parts, one a priori the other a posteriori. Carnap's conditional implicit definitions secure a place for the constitutive a priori in science, but it turns out that it is a rather anaemic conception of it compared to Poincaré's more substantive one (section 7).

2. The Absolute Conception vs. the Absolute Rejection

According to the Kantian conception of the a priori, the possibility of human knowledge requires placing a priori restrictions on the admissible models of the experienced world – only those models are admissible that conform to a set of synthetic a priori principles. This captures a sense of *constitutive a priori*: some principles are necessary presuppositions for knowledge (and for doing science) – necessary in the sense of being *sine qua non* for understanding the world. Since those principles that are necessary for experience precede experience, they cannot be defeated by it; they are permanent and *unrevisable*; they are *necessarily true*.

Kant thought that these two senses of being necessary – necessary presuppositions for doing science and necessary as permanent and unrevisable – ought to *coincide* if some principles properly were taken to be independent of experience. This coincidence is the kernel of the absolute conception. It presupposes the incompatibility of two kinds of stance that can be said to capture the entitlement to justifiably hold a principle *independently of experience*. The *first* is an entitlement to accept a principle as being constitutive of experience prior to any empirical justification of it (e.g., before it is being inductively established, verified etc.). The *second* is an entitlement to revise or abandon such a principle.

According to the Millian-Quinean absolute rejection of the a priori, there cannot be any justification independently of experience. Mill's chief point was that *all* justification, even justification of the laws of arithmetic, is inductive. Quine's chief point was that *everything* can be revised or abandoned in light of experience. If a theory is confirmed, everything it says is confirmed; conversely, if a theory is refuted, any part of it can be revised (or abandoned) in order for harmony with experience to be restored. Since, according to the absolute conception, statements that are supposed to be a priori are unrevisable, Quine drew the conclusion that there are simply no a priori principles. Besides, the absolute rejection capitalised on the fact that the absolute conception was meant to offer a deep explanation of why a priori principles are independent of experience, and hence unrevisable. This explanation was in terms of some trait *X* that a priori principles share; some trait that *explains* why there is entitlement to some principles independently of experience.

Apriority is an epistemic condition but has been backed up by a modal trait (a priori principles are necessary truths) or a semantic one (a priori principles are analytic truths) that is supposed to ground and explain it. The candidates, however, that have been offered as explanatory of apriority have been found wanting. As Kripke (1972) made famous, there are cases of propositions such that if they are true they are necessarily true, but whose truth can be known only a posteriori. Conversely, there are propositions whose truth is knowable a priori, even though the truth they express is contingent. It's now fair to say that the concepts of apriority and of being necessarily true do not coincide.

Analyticity cannot be equated with apriority in the strong sense that (all and) only analytic truths can be a priori. Even though many empiricists – notably the logical empiricists – thought that there are analytic truths and that they are (the only) a priori truths, Quine's arguments against analyticity have conclusively shown that there is no non-circular way to characterise analyticity. This, of course, does not show that there are no analytic truths – but it does question that we have a coherent idea of what we attribute to them when we call them analytic.

It's been part of the absolute rejection that there is no such missing trait *X* to be found – shared by all and only the supposed a priori principles. There is no uncontroversial candidate for a feature that a priori principles are supposed to share and in virtue of which there is entitlement to hold them independently of experience. So, if we think (as we should), that to call a principle a priori is to mark an entitlement to accept it *independently of experience*, we face the following dilemma: either there can be no such entitlement or there must be a way to characterise it that does not explain this entitlement by reference to some extra trait (necessity, analyticity etc.) attributed to the objects of this entitlement.

3. A Middle Way

What exactly does the claim of independence of experience amount to?[1] One way to approach this question is to look at the sources of justification for a certain belief. Several sources (e.g. sensory experience, memory, perhaps also introspection) are excluded from providing a priori entitlement to a subject. More positively, it is taken that *pure* reason is the single source of a priori entitlement.

[1] Casullo's (2003) is perhaps the best discussion of these issues.

It's not clear, however, what the boundaries of pure reason are; nor should it be taken for granted that there is anything like rational insight as a direct, irreducible intellectual faculty of the human mind that confers justification on some principles independently of experience. But even if we were to grant pure reason and rational insight as sources of justification, it would be questionable that experience does not intrude at all in their operation. In the most trivial cases, and irrespective of the sources of justification, to believe a proposition presupposes the acquisition of certain concepts – and this acquisition rests on experience. Let us, however, agree with Field (2005) and others that cases such as the above do not count. Let us take it that independence of experience kicks in after the concepts have been, somehow, acquired. The issue is still with us: how exactly is experience (and its limits) to be understood?

Boghossian and Peacocke (2000) have rightly noted that experience can be construed in a number of ways. In its strictest sense, it means *perceptual* experience. In its most liberal construal, it includes any conscious state. If we adopted the strict interpretation, whatever has introspection as its source could not count as an experience, since introspection is bounded by the subject's body. Not so, if we took the more liberal view. One could know a priori (independently of experience) that one is, for instance, in pain. Not so, on the strict sense. The very idea of independence of experience is, then, ambiguous – or, better, malleable. Consider, for example, various mathematical beliefs. It may well be that a person's entitlement to a mathematical belief relies on several sources that implicate experience in both of the above senses, e.g. mental arithmetic, counting collections, using calculators or consulting mathematical texts. Or consider cases such as checking long mathematical proofs, where the checking cannot be done with ›the mind's eye‹. Is then the entitlement to a mathematical belief independent of experience? Viewed under the rubric of the sources of belief, the answer is indeterminate. There is simply no robust way to draw the line between sources of belief that depend on experience and sources of belief that do not.

Field (2005) has suggested a useful distinction according to which a person's entitlement is *independent of experience* in a *weak* sense if the experiences she already had are not *responsible* for this entitlement and in a *strong* sense if the experiences she might have could not *defeat* this entitlement. He stresses that the philosophically important sense is the strong one. This is an interesting move since it leaves behind as philosophically uninteresting the sources of a subject's belief vis-à-vis the issue of her a priori entitlement. It follows from Field's move that independence of experience accrues to propositions that cannot be defeated or undermined or revised *on empirical grounds*.

Not surprisingly, we have come back to the issue of revisability. In one or another guise, independence of experience is taken to be intimately connected to unrevisability. Notably, for Field and others, the connection between apriority and unrevisability is *not* grounded on a further trait that apriority is supposed to have (necessity or analyticity) and in virtue of which a priori principles are immune to revision on empirical grounds. Still, it is unrevisability that counts for apriority. Here, or course, is territory in which Quine's argument against unrevisability looms large.

Have we then reached an impasse? Suppose that independence of experience has to do with unrevisability (on empirical grounds). If there is no further explanation of unrevisability in terms of a further characteristic that these principles are supposed to share and in virtue of which they are immune to revision on empirical grounds, the requirement of unrevisability simply boils down to an *intention* on our part to hold onto some principles, come what may. This is something that a Quinean would accept, while adding that this intention is a) a psychological state and b) revocable.

It seems to us that to tackle the issue of the connection between independence of experience and unrevisability, we need to draw a distinction between refutation and revision (and their modal analogues: refutability and revisability). A refutation of a theory (or a principle) on empirical grounds would require a direct contradiction between the theory (or the principle) and some sort of experiential situation (an experiment, an observation, some empirical law, a prediction and the like). But a theory may well be revised (on empirical grounds) without being refuted on empirical grounds. So revision is broader than refutation. The lesson drawn from the Duhem-Quine problem might be enough to drive this point home. (More will be said on all this is the next section.) For the time being, let us just consider briefly the case of Euclidean geometry. Euclidean geometry can be consistently negated – hence the non-Euclidean geometries. Can it be refuted on empirical grounds? *Qua* mathematical system, it cannot be refuted on empirical grounds. This is simply because it makes no empirical claims whatsoever. Hence it cannot be subjected to any present or future empirical test. Has Euclidean geometry been revised on empirical grounds? Well, *something* has been revised, but this is the claim that Euclidean geometry correctly describes the structure of *physical* space. *Qua* physical geometry, Euclidean geometry has been defeated, or undermined, on empirical grounds. The very possibility of this revision requires that Euclidean geometry has been applied to physical space. Physical geometry is revisable on empirical grounds because it can be subjected to empirical tests (*even though* it can be saved from refutation by a corrective move in some concomitant physical theory and the like).

What does all this mean? A principle (better, set of principles) can be independent of experience in that it cannot be refuted on empirical grounds, and yet this same principle (or set of principles) can be revised on empirical grounds in that experience can undermine it (that is, it can undermine its application to the physical world). Problems concerning the applicability of a certain set of propositions can explain their coming to conflict with experience and being undermined – but not *refuted* – by it.

Can there be a way to construe independence of experience that leaves room for the possibility that a principle is independent of experience and yet revisable on empirical grounds? Can there be, in other words, a middle ground between the absolute conception and the absolute rejection?

The *locus classicus* of this middle ground is found in Hans Reichenbach (1921). As is well known, he drew a distinction between two elements in the Kantian conception: a priori principles are meant to be necessarily true; and they are meant to be constitutive of the object of knowledge. Reichenbach accepted the second dimension but denied that a priori principles were necessarily true and unrevisable – rather, being framework-dependent, they are abandoned when the framework they are constitutive of is abandoned. This conception of the *constitutive a priori* brings to light the idea that there cannot be any systematic attempt to know the world unless the admissible empirical theories of it are restricted in such a way that they satisfy a set of principles that circumscribe the basic structure that the world must have in order for it to be knowable.

How can a principle be constitutive of a framework? In particular, how can it constitute the object of knowledge? A clear answer to these questions can be found in Poincaré's theory of conventions.[2] What we are about to claim is that constitutive principles – what Poincaré called conventions – satisfy some key requirement that makes them good candidates for a priori principles: they are independent of experience in that they constitute the object of knowledge, which is not, in the first instance at least, anything encountered in experience. This idea of constitutivity makes possible the thought that a priori principles might, after all, share something in common in virtue of which they are a priori, but this something is neither necessity nor analyticity, nor some deep explanatory trait, but rather the fact that some principles are elevated to conventions and are held to define the object(s) of knowledge.

4. Conventions

Conventions are *not* synthetic a priori principles, nor, of course, analytic a priori principles. For Poincaré, only the principles of arithmetic are synthetic a priori, mainly because the principle of mathematical induction requires synthetic a priori intuition. Synthetic a priori principles are »imposed on us with such a force that we could not conceive of the contrary proposition, nor could we built upon it a theoretical edifice« (1905, 74). The last requirement is very interesting, since it suggests that synthetic a priori principles are constitutive of our form of *understanding*. Conventions, on the other hand, are such that they can be meaningfully

[2] Reichenbach did speak about axioms of co-ordinations and distinguished them from the axioms of connection. But he left the constitutive status of the former unexplained.

negated and alternative theoretical structures can be built on their negations (see for instance, the non-Euclidean geometries and physical theories built on their basis).

4.1 Principles of Geometry

Poincaré argued that the principles of geometry are neither synthetic a priori nor empirical generalisations. The space studied by geometry – the geometrical space – is not the sensible space; it has properties radically different from the sensible space (e.g., it is homogeneous and isotropic). Geometry »is nothing but the study of formal properties of a certain continuous group; so that we may say space is a group«. This is a key idea. The very concept of a group (*qua* a freely created mathematical concept) is something that pre-exists in the mind. Better put, geometrical space is constituted as such by applying the concept of group to certain ideal (mathematical/abstract) entities – the displacements (without deformation). These are not empirical entities, but they can be seen, to some extent, as abstractions of empirical entities, namely rigid bodies. Geometrical space comprises the laws of displacements – that is, the group they form. That the set of displacements forms a group (e.g., that if two changes A and B are displacements, then the change A + B is a displacement) is not a priori true. Yet, if it were not taken to be the case, there would be no geometry. In other words, this principle is a presupposition for doing geometry – this is exactly what Poincaré calls ›convention‹. Geometrical space, in its turn, is a presupposition for doing science: »we *reason* about [external] objects as if they were situated in geometrical space« (1902, 82), though of course they are not.

In a loose sense, the group characterises the laws that displacements obey. If, as Hume observed, all events were distinct and separate, any displacement could be accompanied by any other. To put some *order* to the succession of displacements, some fixed laws of succession are necessary – and these are imposed by the group. By the same token, we could conceive of different laws of succession – and hence of different groups – that constitute the geometrical space. That is exactly what non-Euclidean geometries amount to. What's important, then, is the very concept of the group; this drives us away from empiricism and at the same time away from absolute apriorism, which would have it that only one geometry were possible. Still, experience plays some important role since it is experience that helps selecting among the many possible groups the ›standard‹ one, i.e. the one »to which we shall refer natural phenomena«.

Poincaré's geometrical conventionalism should be seen as a summary of this group-theoretic approach to geometry. Geometrical space is a framework within which all empirical objects are cast and in virtue of which we treat them as if they were geometrical objects. It's just fortunate that ordinary solid objects approximate the behaviour of ideal rigid bodies and, hence, their movements can be treated as if they were displacements. The conventionalism of this position lies in the fact that a) some geometrical framework should be chosen; and b) the choice is based on a free decision – hence, it is not *dictated* by empirical facts.

Why did Poincaré insist that geometry has nothing to fear from new experiments? It is noteworthy that Poincaré did not restrict this claim to Euclidean geometry. *Any* kind of geometry (that is, theory of geometrical space) has nothing to fear from experiments. His point was that the conventional character of a set of geometrical principles could not (and should not) be mistaken for anything else; in particular geometrical principles should not be mistaken for empirical principles – they are neither verifiable by experience (since geometrical objects are not empirical objects) nor refutable by experience.

Geometry is, in a sense, testable only when it is applied, with the aid of several auxiliary hypotheses, to the physical world and thus becomes part of the physical theory of the world. But then what is tested is the combination of geometry and physical theories and the experiments that do the testing are physical experiments (experiments that do *not* engage geometrical objects *per se* but solid bodies and light rays). What is therefore being tested is not the geometrical theory *per se* but rather a physical application of it.

4.2 Principles of Mechanics

Thinking about the principles of mechanics, Poincaré argued that although these principles are taken to be »rigorously true« (1902, 151), their truth can be neither a matter of demonstration nor established on a posteriori grounds. Take, for instance, Newton's law of inertia: if a body is not acted upon by external forces, its velocity remains unchanged. This is not a truth of reason; nor is it synthetic a priori. One can conceive of worlds in which, if a body is not acted upon by any external forces, either its position or its acceleration – and *not* its velocity – remain unchanged (cf. 1902, 113–15). Are, then, the principles of mechanics empirical generalisations established (and accepted) on the basis of experience? The answer is negative since the systems to which the laws of mechanics apply, such as perfectly isolated systems, are not to be found in nature. Their objects (the objects of knowledge of mechanics, at least in the first instance) are not worldly objects. No-one can verify these principles by recourse to experience. No experiential situation can afford us with perfectly isolated systems and the like (cf. 1902, 116). Besides, no experience can ever refute a principle of mechanics, for two reasons. *First*, since the principles of mechanics apply to systems not encountered in experience, they can never be submitted to a rigorous and decisive test. *Second*, even if a mechanical principle could be submitted to a rigorous test, it could always be saved from refutation by some sort of corrective move.

What, then, is the status of the principles of mechanics? Poincaré called them *conventions*, or »disguised definitions«. Though Poincaré insisted that conventions (the principles of mechanics, in particular) have their origin in experience, he also claimed that they »have been, so to speak, erected into principles to which our mind attributes an absolute value« (1902, 153). The crucial question then is: how can a universal principle, like a principle of mechanics, be neither synthetic a priori nor an empirical generalisation? The right answer to that question lies in the elevation of this principle to an unconditionally acceptable, strictly universal and rigorously true principle (convention). Kant himself distinguished between *strict* universality and empirically arrived universality – whatever is arrived at by induction cannot have the generality (exceptionless character) of an a priori principle. He took it that strict universality marks a sense of independence of experience characteristic of the a priori. Strict universality has to do with not being based on judgements of probability and evidence. For example, the reasonableness of accepting a universal principle of the form ›All As are B‹ is not always a matter of probability and evidence. A universal principle of the form ›All As are B‹ might mark a connection between concepts or a basic postulate for developing a theory. We may call this kind of universality ›unconditional‹ and claim that unconditionally universal principles are unconditionally acceptable, whereas empirical generalisations are conditionally acceptable (that is, conditionally on the available evidence). Principles that are constitutive of a theoretical framework are unconditionally universal. They are not (in the first instance at least) about empirical objects in the world; they constitute the objects of knowledge of the framework, and this marks a sense in which they are independent of experience. The objects of knowledge thus constituted have been the products of idealisations and abstractions. They are not empirical objects, though they can stand in (typically inexact) representational relations to empirical objects – in virtue of which the framework can acquire empirical content. It wouldn't be an exaggeration to say (though we shall not argue for this now) that for Poincaré the objects of knowledge thus constituted are mathematical models and idealities. However, not all statements of scientific theories can and should achieve the status of constitutive principles, since »if all the laws had been transformed into principles nothing would be left of science« (1902, 166).

The applicability of the constitutive principles of a physical theory to the empirical world requires that the objects met in experience are *analogous* (or similar) to the objects to which the constitutive principles apply. Still, it does not follow that these principles are empirical principles. To show this, let us take the statement:

(1) ›The stars obey Newton's law of gravitation.‹

This sounds like a testable hypothesis. But (1) can be decomposed into two other statements:

(2) ›Gravitation obeys Newton's law‹; and

(3) ›The only force exerted on the stars is gravitation.‹

Suppose, Poincaré (1902, 165) says, that astronomers discovered that the stars do not obey exactly Newton's law of gravity. What could they do? Two options are available. First, they might say that the gravitational force does not vary exactly as the inverse square of the distance between the two stars. If one thought of Newton's law as an *empirical* claim (an inductively established generalisation, say), such an observation would falsify it. The second option is for them to say that the force of gravity does vary exactly as the inverse square of the distance, but in this particular case – and others of similar deviations – gravitation is not the only force acting on the stars. If one thought of Newton's law as a definition of gravitation (as constituting the object of knowledge), this attitude would be more appropriate.

One is free to treat (2) as a *definition* that »escapes the verification of experience«. Accordingly, one can accept (2) independently of experience. Yet (3) is a substantive claim that can be tested. It is because of the testability of (3) that (1) is testable. One could, it seems, treat (2) as an empirical claim. But this would distort the situation. Due to its exactness etc., (2) does not apply, strictly speaking, to any worldly situation.

Treating the most general principles as *definitions* (of the objects of knowledge) implies that they end up being independent of experience, even though there is a sense in which they are based on experience: it is experience that »suggests«, or »serves the basis for«, or »gives birth to« the principles of mechanics (cf. 1902, 124). Our entitlement to them – and their elevation to conventions – is *independent of* experience: experience alone can neither force these principles upon us nor refute them. In adopting these principles and, subsequently, in interpreting experience in their light, there is always an element of *choice*.

Still, conventions (and the principles of mechanics, in particular) are revisable. In a rather marvellous passage, Poincaré drew a fine distinction between contradiction and condemnation – which underwrites our earlier distinction between refutation and revision. He was quite firm in that no experiments can ever contradict a principle of mechanics. For no experiment can conclusively refute such a principle. Yet, experiments can condemn a principle of mechanics, or even a whole mechanical framework, in that persistent failure to account for new facts renders a particular principle or a whole framework no longer convenient (cf. 1902, 146).

5. Implicit Definitions

Conventions, we have argued, can be seen as delineating a new sense of the a priori – the *conventional* a priori principles (which are constitutive of the object of knowledge, but their negation is conceivable and they are revisable). The natural way to understand this idea of constitutivity (suggested by Poincaré himself) is to think of conventions in terms of *implicit definitions* (or »*definitions by postulates*« as he, following Couturat, put it – 1908, 150). Clearly, Poincaréan definitions are not meant to be explicit definitions. As a rule, an implicit definition specifies the meaning of some concepts simultaneously and in a collective way. Unlike an explicit definition, it does not serve the purpose of eliminating the *definienda* in favour of the *definiens*. It was Hilbert's axiomatisation of Euclidean geometry and his point that the basic concepts that feature in a set of axioms get their meaning from their mutual logical relationships that set implicit definitions in motion.

The way an implicit definition is supposed to work is this: the meaning accrued to concept (or concepts) F (G, H) is such that a certain postulate(s) in which it (they) occur(s) is (are) true. Using standard notation, we may write ›#_‹ for the set of postulates with a blank wherever the *definiendum* F occurs; the definition then assumes the form ›$F =_{def} \#F$ is true‹. The objects of knowledge – those to which the implicitly defined concepts apply – are whatever entities satisfy the postulates (provided, of course, that the postulates are consistent). Actually, *any* system of entities that satisfy the postulates is such that the implicitly defined concepts apply to them. So the object of knowledge is not any independently given set of entities in particular but, in the first instance at least, a certain relational structure.

The chief attraction of implicit definitions is precisely that they fix meaning not by analysing already known and understood concepts but by legislating in a stipulative and conventional manner the truth of certain propositions of which the defined concepts are constituents. Hence, they *create* or *constitute* meanings: for something to be an F (that is, for the concept of F to be applicable to it), such and such conditions must be satisfied. The postulates (the conditions) by means of which implicit definitions are effected are conventional in that their form (let alone their content) is not dictated by experience. They are, hence, independent of experience in the sense canvassed by Poincaré. They are not empirical generalisations; they are rigorous; they involve a kind of selectivity in their constructions – they are, in the end, free choices. Implicit definitions can thus be seen as the locus of the constitutive a priori: they impose a priori restrictions on what the world is like – the only admissible empirical models should be such that they satisfy the stipulations.

Interestingly, definitions by postulates are not analytic. This has been pointed out by Arthur Pap (1946) and has been stressed by Wilfrid Sellars (1953). According to the latter, an explicit definition gives rise only to a priori *analytic* truths. So, if we search for principles that are true *ex vi terminorum* (*by means of the meanings of their terms*) and, at the same time, synthetic, then »If anything that has been called definition can serve this purpose, it is what … we shall call implicit definition …« (1953, 124–5). For Sellars, an implicit definition specifies that a set of statements in which a number of predicates occur is unconditionally assertable.

These statements are neither deduced from other statements nor derived from empirical situations, being thus independent of experience. Yet, on the basis of such definitions, conceptual frameworks are constituted, within which we learn to use certain symbols according to syntactic and semantic rules.

Presented as above, implicit definitions face a drawback. If the implicit definition of a set of concepts amounts to a definition of them by postulates, and if a scientific theory is presented as a deductive-postulational structure with empirical consequences, the role of implicit definitions becomes dubious. How can they function as definitions if they seem to generate new empirical content? This is a well-known problem that became entangled with the issue of separating the analytic (or meaning-fixing) and the synthetic (or factual) content of the theory. But if we do not tie implicit definitions to analytic truths (meaning: some statements might implicitly define some concepts without being *analytically* true, as in the case of the axioms of Euclidean geometry), the problem becomes how to separate a theory into two parts, one being a priori, the other a posteriori.

In the current discussion of implicit definitions, Hale and Wright (2001, 129–30) have pointed out that implicit definitions should be non-arrogant. Arrogant is a claim whose truth can be established by means of empirical investigation. Being a priori, implicit definitions should be non-arrogant. If an alleged implicit definition #F were arrogant, its employment would have to await empirical confirmation – hence, it would cease to be a definition of F. The drawback of implicit definitions by means of postulates is that it is not clear that they avoid arrogance. As we are about to see, this is a key thought that Carnap had and which led him to a breakthrough: the view that implicit definitions should be *conditional* in form.

6. Avoiding Arrogance

An interesting case of implicit definition is given by Carnap's reduction sentences in his (1936). A theoretical term Q is introduced by means of the following *reduction pair*:

$$\forall x\,(S_1x \rightarrow (O_1x \rightarrow Qx)) \qquad \text{(G1)}$$
$$\forall x\,(S_2x \rightarrow (O_2x \rightarrow \neg Qx)) \qquad \text{(G2)} \qquad \text{(RP)}$$

where S_1, S_2 describe experimental (test-)conditions and O_1, O_2 describe characteristic responses (possible experimental results). In the case that $S_1 = S_2$ (=S) and $O_1 = \neg O_2$ (=O), the reduction pair (RP) becomes a *bilateral reduction sentence*

$$\forall x\,(Sx \rightarrow (Qx \leftrightarrow Ox)). \qquad \text{(RS)}$$

Carnap rightly thought that the introduction of theoretical terms by means of reduction sentences was an improvement over the aborted attempts to define them explicitly by reference to an antecedently understood observational language. The reduction sentence (RS) does not fully define the predicate Q. Thus, the meaning of Q is *not* completely specified by virtue of observable predicates. At best, the reduction sentence gives, as Carnap (1936, 443) put it, »a conditional definition« of Q. And that's exactly the virtue of reduction sentences: being conditional in form, they can be seen as rules of introduction and elimination of theoretical concepts.

Can reduction sentences be treated as (implicit) definitions of theoretical concepts? Reduction sentences seem to play a factual role too, viz., they specify the empirical content of theoretical concepts. In Carnap's thinking, the problem was that reduction sentences had to be factually empty if they were meaning-fixers. But were they? There is an interesting difference between reduction pairs and bilateral reduction sentences. Take a reduction pair of the form of G_1 and G_2 above. It cannot be factually empty, since, as Carnap noted, it has synthetic consequences: the conjunction of G_1 and G_2 entails the sentence C: $\forall x \neg(S_1 \& O_1 \& S_2 \& O_2)$. This is a synthetic statement about the world. It asserts that the four predicates cannot be satisfied together. Carnap took this sentence to be the factual content of a reduction pair – what he called its »representative sentence«. Bilateral reduction sentences, however, do not face this problem. Their representative sentence is C‹: $\forall x \neg(S_1 \& O_1 \& S_1 \& \neg O_1)$, which is a tautology. Hence, they have no factual content and they can be taken to be implicit definitions.

What Carnap, in effect, suggested was that an implicit definition of a concept must, *qua* definition, be non-arrogant: its truth must be affirmed in a stipulative way and must not require any empirical investigation. By the same token, an implicit definition of a new concept in virtue of already understood vocabulary must be *conservative*: it must not entail fresh consequences stateable in terms of the already understood vocabulary. If it is not conservative (as RP is not), its truth cannot be just a matter of stipulation but depends on the truth-value of its representative sentence, which might well be a synthetic sentence.

The problem faced is to build an account of implicit definition which retains the advantages of the conditional form of reduction sentences while respecting the demand of conservativeness: the stipulative truth must *not* have non-tautological representative sentences expressed in the antecedently understood vocabulary.

Carnap's solution to it came in his (1952). Let G ($=G_1\&G_2$) be the conjunctions of the two reduction sentences. G entails the representative sentence C: $\forall x \neg(S_1 \& R_1 \& S_2 \& R_2)$. One can easily verify that since G implies C, G is equivalent to C & (C → G). So, G is analysable into two components, C and (C → G), the former expressing its factual content, while the latter – «which has no synthetic consequences« – captures its meaning-fixing component (cf. 1952, 71).

The strategy is generalisable. Assume that #F entails R. Then, #F will be equivalent to R & (R $\rightarrow$ #F). Being conditional in form, (R$\rightarrow$#F) is non-arrogant and can function as the a priori ›part‹ of #F and, in particular, the ›part‹ that implicitly defines F. The problem is to find the suitable R (that is, the suitable representative sentence). In the case of scientific theories, R should be such that it captures the empirical content of the theory, thereby bearing the empirical burden of the theory.

Carnap's ingenious idea was that the Ramsey-sentence ^{R}T of a theory T is the representative sentence of the theory and that given that the theory implies its Ramsey-sentence, a theory T is logically equivalent to the following conjunction: ^{R}T & ($^{R}T \rightarrow T$), where the conditional $^{R}T \rightarrow T$ is the so-called *Carnap-sentence*. The Ramsey sentence ^{R}T says that there are classes of entities which are correlated with the observable events in the way the postulates of the theory describe; but it does not say what exactly those entities are – it does not pick out any such class in particular.

7. Carnap-sentences as Implicit Definitions

Carnap's thought was that by manipulating the *form* that implicit definitions should have, their very possibility as a priori stipulations is secured. Using the notation above, Carnap's idea is this: Given that #F entails ∃x(#x), #F will be equivalent to ∃x(#x) & (∃x(#x) $\rightarrow$ #F). Hence, the implicit definition of theoretical concepts will have the conditional form (∃x(#x) $\rightarrow$ #F).

We have already encountered this idea of splitting a factual assertion into two components, one definitional and another empirical, in Poincaré (see section 4). The distinctive advantage of Carnap's way was that being conditional in form, the Carnap-sentence-style implicit definition is non-arrogant. This bears some elaboration.

Recall that Poincaré took the statement (2) ›Gravitation obeys Newton's law‹ to be a definition of gravity. Then, he took the empirical content of the statement (1) ›The stars obey Newton's law of gravitation‹, to be captured by the statement (3) ›The only force exerted on the stars is gravitation.‹ But (2) may be arrogant. Carnap's thought, in effect, was that (2) needs refining to be taken as an implicit definition of gravity. It should itself be analysed along the following lines:

(2a) ›There is something that obeys Newton's (inverse-square) law‹; and

(2b) ›if there is something that obeys Newton's (inverse-square) law, then *gravity* obeys Newton's (inverse-square) law.‹

In effect, (2) (which can be written as #F – where ›F‹ stands for gravity) should be analysed as: (∃x(#x) & (∃x(#x) $\rightarrow$ #F). The (implicit) definition then is not

(2), but (2b). This does not tell us how the world is – it makes no factual claims whatever. It merely defines or constitutes the object of knowledge, viz. gravity. Factual claims are made by (2a) and (3), which secure that (1) is still testable. (2b) is clearly non-arrogant.

More generally, the conditional ›*if* the Ramsey-sentence of the theory is true, *then* the theory is true‹ should be read thus: if there are entities that satisfy the Ramsey-sentence (and, of course, the Ramsey-sentence is empirically adequate), then these entities are those that render the theory true. Accordingly, the theoretical concepts of the theory are implicitly defined by the Carnap-sentence in such a way that they refer to whatever entities satisfy the Ramsey-sentence. This is clearly an implicit *definition*: it stipulates meanings without defining them explicitly. Carnap himself insisted that the Carnap-sentence is a meaning postulate, and in particular an analytic statement. But this is not quite right. The conditional $^{R}T \rightarrow T$ is not a logical truth; nor is it analytic in the Fregean sense (derived from logical truths by means of definitions and logical laws); nor is it the case that it cannot be denied without contradiction. The Carnap-sentence can be seen as an unconditionally acceptable principle constitutive of the conceptual framework of a scientific theory: it defines (implicitly) its theoretical concepts and *ipso facto* the object of knowledge of the theory, viz., whatever satisfies its Ramsey-sentence. It is neither a demonstrable truth nor an empirical generalisation. It is not based on judgements of probability and evidence. It is a convention in Poincaré's sense. The Carnap-sentence is a priori precisely because it is independent of experience. It does not assert anything about the world. As Carnap put it: »It does not tell us whether the theory is true. It does not say that this is the way the world is. It says only that *if* the world is this way, then the theoretical terms must be understood as satisfying the theory« (1974, 271). The Carnap-sentence poses a certain a priori restriction on the class of models that satisfy the theory: it excludes from it all models in which the Carnap-sentence fails.

Carnap's thought was that a) the Ramsey-sentence of the theory captures that part of its content that is a posteriori and b) the conditional $^{R}T \rightarrow T$ captures that part of the theory that is a priori. Hence c) the issue of whether a theory is true is divided into two separate issues: one dependent on experience (is the Ramsey-sentence of the theory true?) and another independent of experience (is the Carnap-sentence of the theory true?). As is now well-known, there is a drawback in this analysis. The Ramsey-sentence of a theory can be false since it might be empirically inadequate. But *if* it is empirically adequate, it cannot be false, provided that the universe of discourse has the right cardinality. So, if the Ramsey-sentence is empirically adequate, the only way in which the world might fail to satisfy it is by not having *enough* entities to make the Ramsey-sentence true. This does not impugn the apriority of the Carnap-sentence. But it does show its anaemic character. The only constrain it poses concerns the cardinality of the domain of discourse: there should be *enough* entities to satisfy an empirically adequate Ramsey sentence. If there are, then the implicitly defined theoretical concepts refer to them – the genuine object of knowledge of scientific theories becomes the cardinality of the domain of discourse.

So we have reached an impasse. If we follow the Poincaréan conception of conventions and develop it in terms of implicit definitions, we have a substantive conception of the constitutive a priori in science – where substantive principles constitute the objects of knowledge. But it is far from clear that we have succeeded in separating the a priori from the empirical. If we follow the Carnapian conception of implicit definitions, we secure a place for the constitutive a priori in science, but it is a rather anaemic one.

References

Boghossian, Paul & Peacocke, Christopher (2000) ›Introduction‹, in New Essays on the A Priori. Oxford: Clarendon Press.
Carnap, Rudolf (1936) ›Testability and Meaning‹. Philosophy of Science 3: 419–471.
– (1952) ›Meaning Postulates‹. Philosophical Studies 3: 65–73.
– (1974) An Introduction to the Philosophy of Science. New York: Basic Books.
Casullo, Albert (2003) A Priori Justification. Oxford: Oxford University Press.
Field, Hartry (2005) ›Recent Debates about the A Priori‹, in Tamar Szabo Gendler & John Hawthorne (eds), Oxford Studies in Epistemology. Volume 1. Oxford: Clarendon Press.
Hale, Bob & Wright, Crispin (2001) The Reason's Proper Study. Oxford: Clarendon Press.
Kripke, Saul (1972) Naming and Necessity. Oxford: Blackwell.
Pap, Arthur (1946) The A Priori in Physical Theory. New York: Russell & Russell.
Poincaré, Henri (1902) La Science et L'Hypothése. (1968 reprint). Paris: Flammarion.
– (1908) Science et Méthode. (English transl. 1952). Paris: Flamarrion
Psillos, Stathis (2006) ›Ramsey's Ramsey-Sentences‹, in M C Galavotti (ed.), Cambridge and Vienna: Frank P. Ramsey and the Vienna Circle. Springer, 7–90.
Reichenbach, Hans (1921) The Theory of Relativity and A Priori Knowledge. English translation by M. Reichenbach, 1965. Berkeley: University of California Press.
Sellars, Wilfrid (1953) ›Is There are Synthetic A Priori?‹. Philosophy of Science 20: 121–138.