A Priori truth in the natural world : a non-referentialist response to Benacerraf's dilemma

Novák, Z.

CONCLUSIONS

In this work, I have argued for a naturalistic construal of *a priori* truth and knowledge. My reasoning has consisted of three consecutive steps. Having elucidated the central notions and the major methodological assumptions of this inquiry, first, I argued that in the semantics of our paradigm *a priori* discourses (i.e. in the semantics of pure logic and mathematics) truth cannot be adequately interpreted along the standard referentialist line, in other words, assuming that the truth conditions of the relevant beliefs are to be specified in terms of those entities that these beliefs purport to be about ("non-referentialism about the paradigms of *a priori* truth"). Second, I argued that the truth conditions (as opposed to the intended referents) of our paradigm *a priori* beliefs are natural conditions in our heads, or, more exactly, that they are analytic relations among representations that we develop in our heads in the course of our cognitive interaction with our direct natural environment ("representationism about the paradigms of *a priori* truth"). Finally, I argued that we have good reasons to take representationality as an essential, defining trait of apriority, and thus as a necessary characteristic of *a priori* truths in general ("representationism about *a priori* truth, knowledge, justification and evidence").

My case against the standard referentialist construal of the paradigms of *a priori* truth was based on the observation that no conceivable version of the referentialist conception can support an acceptable account of all striking characteristics of these paradigms, and thus none of them can offer a viable solution to Benacerraf's updated and generalised dilemma in the philosophy of the relevant *a priori* disciplines. From among the *prima facie* plausible versions of the referentialist doctrine, the deflationist and anti-realist forms were shown to be incompatible with a suitable explanation of the objectivity of the examined truths, while the realist (i.e. platonist) form proved to support no

reasonable account of how we can acquire knowledge or reliable beliefs of the intended abstract domains.

My case for the suggested representationist construal of the same paradigms was based also on explanatory considerations. In particular, I argued that an advocate of this construal can explain all striking characteristics of the subject, and thus demonstrate that the conception amounts to a suitable response to Benacerraf's dilemma and at least a minimally adequate characterisation of the paradigms of *a priori* truth.

Finally, my case for the corresponding representationist construal of apriority in general was based on pragmatic considerations. My primary motive at this point was that by adopting this construal we obtain a relatively clear, informative, and scientifically confirmable idea of the nature and extension of apriority, which largely observes the received application of the term, which together with our best empirical theories of the world and human cognition explains the most important characteristics of our purportedly *a priori* claims, and which, last but not least, preserves an important connotation of the traditional epistemological notion, namely that *a priori* knowledge is, in a sense, prior to any further knowledge of the actual world.

The establishment of the explanatory adequacy of the naturalistic construal just summarised required a number of conceptual and theoretical assumptions concerning apriority, knowledge, truth, realism, objectivity, reference, abstractness, nature, representation, analyticity and semantic content, which have been made with an eye on our actual cognitive and linguistic practice and the most obvious, observable characteristics of the intended phenomena. For the sake of clarity, many of these assumptions were stated explicitly in chapter 1. Here and now, I do not wish to repeat all these statements. Instead, in what follows, I shall briefly review the most important generalisations that occurred among or can be drawn from the conceptual and theoretical claims made in this work and spell out what the collected findings imply concerning some neighbouring issues in the current literature and the relation of armchair philosophy to the deliverances of empirical inquiry.

The first general claim that I wish to emphasise was defended in chapter 3, and concerns the scope of the dilemma that initiated my case against the standard referentialist understanding of the paradigms of *a priori* truth. As has been observed, in its most generic form, the dilemma provides a challenge to the standard referentialist and realist construal of truth in any discourse in which we are supposed to acquire knowledge of conditions whose obtaining or absence can have no causal impact upon our knowing minds. Beyond the intended abstract referents of our beliefs in pure logic and mathematics, genuine (irreducible) normative properties and values, the entities of realistically construed non-actual worlds, and the future constituents of the actual spatiotemporal world are arguably also incapable of exerting causal influence on our present and actual cognitive faculties. If my arguments in chapter 6 are sound, then nothing in our currently available evidential pool can be reasonably taken as a reliable sign of the obtaining or absence of truth conditions involving such (realistically construed) normative, non-actual or future entities. So, if one believes that we can have knowledge of normative, non-actual or future states of affairs, then one must abandon the idea that the truth conditions of these knowledge claims are to be understood in terms of the respective intended referents.

My own view is that in the case of synthetic claims about realistically interpreted future or non-actual states of affairs referentialism about truth can be maintained, because we can acquire no synthetic knowledge of what will obtain in the future or what obtains in a non-actual world. Our synthetic claims about the future may be more or less well supported by our current evidence *if* future tokens of various entity types will occur along the same patterns as they did in the past, but the truth value of these claims will be fixed, and thus become knowable, only by the actual future obtaining or absence of the relevant (referential) truth conditions. Our synthetic claims about non-actual realms are similar to our synthetic claims about platonic entities: we can never reasonably take anything as evidence for the truth or the falsity of these representations. In chapter 7, I argued that the

truth conditions of analytic claims are non-referential and representational independently of what these representations are about. In line with this view, I do not query the possibility of analytic knowledge of the future or of alternative non-actual realms, but I suppose that this knowledge is knowledge of the (current and actual) obtaining or absence of some representational conditions in the subject's head, rather than of those future or non-actual conditions that these analytic claims purport to be about.

Our normative claims or value judgements are typically synthetic, so a representationist response to the dilemma in their case does not seem to be adequate. Note, however, that representationism is not the only conceivable form of non-referentialism. Simon Blackburn's quasi-realist in meta-ethics, for instance, as I observed in chapter 1, is best understood as an anti-realist about intended referents and a realist about truth conditions. This implies that she must be also a non-referentialist about ethical truths. Nonetheless, her construal is presumably not representationist in character. As Blackburn's or Gibbard's projectivist meta-ethics may exemplify, certain truth conditions can be regarded as non-referential, with realistically construed constituents (e.g. affective and conative attitudes) in the evaluating subject that are outside the domain of the subject's acquired representations. Be it as it may, despite the availability of this alternative non-referentialist construal, I believe that the semantic content of our normative or evaluative utterances is much more idiosyncratic (i.e. intersubjectively much less harmonised) and much less stable (i.e. changing considerably through contexts and time) than that of our utterances about abstract, not-actual or future states of affairs. If this belief is correct, then an adequate construal of truth in the case of these normative or evaluative beliefs may prove to be also much more heterogeneous than one in the semantics of other truth-apt representations.

In chapter 5, I argued that the explanatory failure of standard referentialism laid bare by Benacerraf's dilemma is merely epistemological, not semantical as sometimes believed.

What a referentialist cannot suitably explain is, in my view, our capacity to acquire knowledge or reliable beliefs about some problematic domains. Our ability to develop ideas or make claims about particular entities within those domains provides no explanatory puzzle for a referentialist, since a sufficiently sophisticated causal account of reference determination does not require that each mental representation be causally related to its intended referent. As we were able to learn from Putnam's model-theoretic argument, the semantic relation of model-theoretic satisfaction (specified by reference to the usual theoretical and operational constraints) is necessarily indeterminate. This, however, could undermine the possibility of a suitable explanation of determinate reference only if the two relations (i.e. reference and model-theoretic satisfaction) were identical. If we suppose that, beyond the existing (though merely Humean) causal links between our minds and the perceived aspects of the external world, our referential intentions and various concept-generating mental operations (as causally effective natural factors) also contribute to the determination of what our mental and physical symbols refer to, then the best-known arguments against the standard realist idea that we can determinately think of various (epistemically accessible or inaccessible) extra-mental entities prove to be equally inadequate.

A further important result about reference, anticipated in chapter 1 and then defended in chapter 4, is that the symbol-world relation that is (in my view rightly) invoked in the standard referentialist construal of some truths must be more fine-grained than the semantic link that Frege called reference. If the analysis that I provided of the slingshot arguments is correct, then we can lay down that the relation that stands at the heart of a suitable (referentialist) construal of most synthetic truths must be at least as fine-grained as the semantic aspect that Frege denoted by the term 'sense'.

It may be worth recalling also that the denial of standard referentialism in the semantics of various discourses is a doctrine about truth conditions, not a doctrine about referents. In other terms, by abandoning the idea that the truth conditions of our

truth-apt representations are to be specified always in terms of those conditions that the representations in question purport to be about one does not endorse any thesis about the nature or metaphysical status of the relevant intended referents. In particular, a non-referentialist need not deny the reality or the actual existence of these entities. In fact, she may embrace realism or anti-realism about them. In line with the intentionalist account of reference determination hinted at above, I believe that the adequacy of the latter doctrines (at least partly) hinges upon the subjects' referential intentions that accompany the application of the relevant symbols. Although these intentions are predominantly implicit in the subjects' cognitive and linguistic practice, they can be explicitly formulated in thought and expressed in public language as well. In pure mathematics, for instance, we do not entertain ideas or make claims about the nature or metaphysical status of mathematical objects. This, however, does not mean that our referential intentions in this discipline impose no constraint upon an appropriate "philosophical" characterisation of mathematical entities. Presumably, if questioned, most people would reject the idea that mathematical objects (i.e. the subject matters of pure mathematics) can appear in space and time, that they can be causally effective, or (I presume) that they are merely positions in a structure characterised categorically by a mathematical theory. These judgements are not mathematical, but they inform us about the speakers' referential intentions while doing mathematics, and they provide evidence against the adequacy of various revisionist theories (such as naturalism or structuralism) of the nature of mathematical entities. As regards the adequacy of realism about these referents, the decisive question is whether people conceive them as entities that can exist in the real world or as entities that can "exist" merely in a fictive domain invented by human minds. My guess is that most people would not exclude the possibility of the real existence of mathematical entities, as they would, in contrast, that of Little Red Riding Hood. If this is true, then realism may be an adequate doctrine about the referents of pure mathematics.

The observation that an adequate general construal of truth implies nothing about the metaphysical status of the subject matter of truth-apt representations may sound as an indirect embrace of deflationism in the theory of truth. Note, however, that the non-committal character of a general (not referent-invoking) construal of truth concerning the metaphysical status of subject matters need not prevent this construal from stating something substantive about the metaphysical status of truth itself. Beyond this conceptual point, in chapter 4, I argued that a proper explanation of objectivity calls for a realist construal of truth (or correct declarative applicability in general). Consequently, we can say that the general use-theoretic conception advocated in this work is not merely not entailing, but also diametrically opposed to deflationism in the theory of truth. As was emphasised in chapter 4, it is in fact a substantive realist account, according to which truth is a sort of correspondence, the property of having suitable semantic relation to some conditions that obtain in the actual world. The only difference between this and the existing referentialist versions of the correspondence theory is that this account does not assume that the worldly terms of the relevant correspondence relations are necessarily those facts that the true representations in question purport to be about.

Beyond the establishment of a viable naturalist construal of *a priori* truth and knowledge, the possibly most significant result of this work is the challenge it articulates to the standard, broadly Tarskian, referentialist conception of truth. Again, if the argumentation presented in the previous chapters is correct, then the idea that truth in general can be defined in terms of reference or satisfaction must be false. Now, one may think that the alternative use-theoretic construal advocated here is anti-Tarskian not merely in that it is not referentialist in character, but even in the stronger (and intuitively more problematic) sense that it fails to observe Tarski's T-schema or material adequacy condition for a theory of truth. As is well-known, the condition holds that a viable theory of truth must entail, for each sentence p of the language under scrutiny, the corresponding instance of the

schema "p is true if and only if *p*", where *p* stands for the metalinguistic translation of p. The conflict between the proposed use-theoretic construal and this requirement occurs if the truth or falsity of some sentences of the object language is not referential, while it is supposed that the metalanguage used for formulating the instances of Tarski's T-schema is still so (i.e. that the conditions "specified" by the metalinguistic expressions on the right-hand sides of the instances are always those that these expressions purport to be about). For instance, on a referentialist reading, Tarski's material adequacy condition holds that a proper theory of truth must entail, among other things, that the arithmetical sentence 'two plus three equals five' is true if and only if the abstract arithmetical conditions denoted by the metalinguistic translation of this sentence actually obtain. Under such a reading, the adequacy condition is certainly not observed by a general use-theoretic account that allows for a non-referentialist construal of this arithmetical truth. After all, if the truth conditions of arithmetical sentences are indeed non-referential, then this instance of Tarski's T-schema is not merely underivable from the use-theoretic conception of truth advocated in this work, but also false in the actual world.

Note, however, that Tarski's material adequacy condition need not be read along the standard referentialist line. What the condition intuitively requires is merely that the metalinguistic translation of each sentence of the object language be equivalent with the claim that the sentence in question is true, or (more exactly, as was observed in chapter 4) that the translations on the right-hand sides of the T-schema's instances in some way "specify" the truth conditions of the respective sentences mentioned on the left-hand sides. Now, if the truth of a sentence mentioned on the left-hand side of a schema instance is referential, then it is quite natural to suppose that the truth of its metalinguistic translation will be referential too, or (again, what is the same) that the translation in question does the required specification by referring to (i.e. being about) the relevant truth conditions. However, if the truth of the object sentence under scrutiny is not referential, then it is hard to see why it should be

supposed that the truth of its metalinguistic translation is still so. The intuitive assumption in this case is rather that the latter property is not referential either, which means that the translation "specifies" the relevant truth conditions merely by possessing them as such (i.e. possessing them as truth conditions, rather than as intended referents). Someone who understands the metalanguage, and therewith the translations provided on the right-hand side of the equivalence instances, will presumably understand what the truth or falsity of the corresponding sentences of the object language consists in even if the conditions "specified" are not referential. If Tarski's material adequacy condition is read along this line, then the suggested use-theoretic conception of truth turns out to observe this requirement just as much as its standard referentialist alternative. So, from the perspective of the current work, the problem with Tarski's condition is not that it imposes too strong a constraint upon a construal of truth. It is rather that its satisfaction does not guarantee the adequacy of a characterisation of this subject.

Having collected the most important general implications of this work concerning truth, reference and the relation of these semantic entities, in the remaining part of this conclusion, let me briefly review those general claims that an advocate of the proposed representationist construal must embrace concerning those characteristics that she has attributed to the paradigms of *a priori* truth in particular. In one sentence, the construal in question assumed that our logical and mathematical claims are analytic, their truth conditions non-referential, natural and representational, and their truth value necessary and *a priori* knowable. In view of the conceptual stipulations and theoretical assumption adopted in the previous chapters, it may be worth realising that some of these characteristics are more intimately related than the others.

In particular, I wish to draw attention to seven universal generalisations concerning the actual occurrences of these characteristics that seem to follow from the conceptual and doctrinal commitments made in this work. The first pair is a direct consequence of the suggested representationist construal of

a priori truth, and it observes that all *a priori* knowable truths are representational and all representational truths that are knowable from a first-personal perspective are *a priori* knowable. Since representationality is, at least in principle, an empirically discernible feature of truth conditions, whose occurrence can be established by a careful examination of the subject's cognitive or linguistic behaviour, the identification of apriority with this feature may promote the emergence of a systematic empirical theory of the actual nature and scope of *a priori* knowledge, justification and evidence as well as the sources of *a priori* epistemic mistakes.

The second pair of universal generalisations that I wish to bring into relief states that all knowable (explicitly or implicitly) modal truths are analytic and all analytic truths are (explicitly or implicitly) modal in character. This result can be derived from the semantical explanation of analytic modal truths presented in chapter 7, the idea that knowledge requires causal contact with obtaining truth conditions (so we cannot acquire knowledge of what obtains in realistically construed non-actual worlds), and Hume's (here merely granted) observation that we have no experience of modal (e.g. necessary, possible, probable and chancy) facts in the actual world. Note that this generalisation does not deny the existence of synthetic modal truths (or else claims made true by such modal facts). What it denies is merely that such truths can be discovered by a cognitive subject (unless it is one who, *pace* Hume, is capable of experiencing modal aspects in the actual world). The implication is nonetheless significant, since it runs counter to the popular view that we can acquire synthetic knowledge of the occurrences of various modal aspects (e.g. the necessity of identities, causal relations and essential property possessions, and the probability or chancy character of the obtaining of particular conditions) in the actual world.

Beyond subscribing to the premises mentioned in the previous paragraph, I advanced a few other considerations as well to weaken the appeal of this popular view. Discussing Putnam's case against metaphysical realism in chapter 5, for instance, I delineated a broadly Humean construal of causation, which does not assume the necessary character of this natural relation. If the construal is correct, then our synthetic knowledge of causal relations will not qualify as knowledge of the occurrence of modal aspects in the actual world. Further, in chapter 1, I argued that the apparent *de re* necessity of essential property possession can be explained by reference to our (content constitutive) referential intentions or explicit stipulations, without invoking modal occurrences in the actual world or non-modal occurrences in realistically construed non-actual worlds. If this explanatory strategy is correct, then our knowledge of necessarily possessed essential properties proves to be analytic in character. In chapter 7, I applied this strategy to the case of the scientific reduction of water to H_2O, and argued that if water possesses this chemical structure necessarily, then it is merely because we take it to be something that must do so (i.e. must possess this structure), rather than because of the obtaining of some modal property in the actual world. A complete case against the idea that we can acquire knowledge of synthetic modal truths would require a careful examination of our claims about identities and probabilities as well, something that could not be properly done in this work. Nevertheless, I believe that the result of such an investigation would be also in line with the above generalisation. While our identity claims are arguably also analytic in character, our synthetic claims about chances and probabilities are either unjustified or made true by non-modal truth conditions (e.g. relative frequencies) in the actual world, whose obtaining or absence can be established by human minds.

The fifth universal generalisation that I wish to mention here holds that all analytic truths are representational. The claim is a consequence of the representationist construal of analytic relations elaborated in chapter 7. The opposite generalisation (that all representational truths are analytic) does not follow from the commitments of this work. This is because the analytic relations of representations were not supposed to exhaust the representational conditions that may or may not obtain in the natural world. For instance, the truth conditions of someone's belief that a subject has representations in her head are clearly

representational. Still the belief is synthetic, since its truth value is not determined by the relations of the concepts that occur in its formulation.

The above characterisation of the relation of analyticity and representationality has at least two significant implications in philosophy. First, it provides a clear view of the factual ground of analytic truths, something that has been called for by many critics of post-Fregean moderate empiricist analytic philosophy. In particular, it suggests that analytic claims are true or false in virtue of the obtaining of some content-constitutive natural relations among representations in human heads (and in other representation-manipulating devices). Second, by acknowledging the existence of synthetic representational truths (which might be established from a first-personal perspective as well), the characterisation also concedes the possibility of synthetic *a priori* knowledge of a limited segment of the natural world.

The construal of analyticity adopted in this work gives rise to a sixth universal generalisation as well, according to which all analytic truths are non-referential. This time the fundamental observation to rely on is that the representational conditions whose obtaining or absence was supposed to determine the truth value of analytic beliefs are always different from those conditions that the beliefs in question purport to be about. Having recognised this fact, nevertheless, we must note that in the case of a metalinguistic claim about the obtaining of the truth conditions of an analytic claim (i.e. the obtaining of a content-constitutive natural relation between some constituents of the relevant analytic claim) the obtaining of the denoted conditions may be constitutive of the content of the metalinguistic claim as well. This situation occurs *if* the constituents of the analytic claim *mentioned* in the metalinguistic claim are supposed to be identified with respect to their semantic contents (rather than merely to their syntactic shape). What this means is that the truth value of such a metalinguistic claim may depend at least as much on the analytic relations of its own constituents as on the obtaining or absence of the conditions that it purports to be about.

Finally, the seventh universal generalisation that I wish to stress declares that all knowledge is knowledge of the actual (past or present) obtaining of some natural (i.e. spatiotemporal) conditions. The claim can be derived from the epistemological principle adopted here that knowledge is always knowledge of the obtaining of some truth conditions that must stand in a suitable causal relation with the knowing mind. Again, the opposite generalisation (that the obtaining or absence of any natural condition is knowable) is not meant to be implied in this work. Although there are reasons for assuming the existence of unknowable facts in the natural world, the argumentation developed in the previous chapters did not require taking a stance on this particular issue.

With this note, I have finished the promised survey of the general implications of the conceptual and doctrinal commitments of this work. The main purpose of this overview was to provide a sense of the broader significance of the proposals and claims that have been adopted and defended in the previous chapters. Although the central concern of the inquiry advanced in those chapters was the elaboration of a viable construal of *a priori* truth and knowledge, the realisation of this task required the development of a relatively clear notion of a number of other phenomena that are essentially related to these subjects. Thus we have witnessed, among other things, the emergence of a general realist use-theoretic construal of truth, a fine-grained, intentionalist, deflated construal of reference, a communicable metaphysical construal of realism that does not imply actual existence, a strictly naturalist causal contact theory of knowledge acquisition, a related representationist construal of analyticity that gives rise to a naturalist semantical explanation of knowledge of (analytic) modal truths, and a broadly causal account of how we learn to think of or speak about epistemically inaccessible entities or domains. Inflated with these conceptual and theoretical elements, the core tenets of this work acquire a wider significance than one might attribute to them after a brief reflection upon the primary subject of this investigation. In this conclusion, I wished to make sure that this collateral impact is

understood as clearly as the central claims and suggestions made concerning the nature of *a priori* truth, knowledge, justification and evidence.

Of course, wider significance means wider ground for potential criticism. In chapter 2, I presented a list of those phenomena whose proper explanation, with the assistance of other theories, I regarded as a condition of minimal adequacy for any construal of *a priori* truth. The stress on minimality was meant to keep awareness of the fact that a fully adequate notion of apriority must face the tribunal of a much larger evidential pool than the one (with the most striking characteristics of the subject) advanced in chapter 2. The implications collected here highlight a few latent segments of this larger evidential pool.

Having said that, I wish to conclude this work with a brief remark on the relation of the outcome of this philosophical inquiry to our current empirical (scientific and ordinary) theories of the spatiotemporal world. As often in philosophy, the primary purpose of this investigation was the development of a relatively sharp and theoretically useful notion of a phenomenon (viz. the apriority of truths and pieces of knowledge) that acquired a certain significance in the traditional concerns of the discipline. Although the work has been carried out from an "armchair", without the inclusion of new empirical data to test the adequacy of the adopted conceptual and theoretical elements, the commitments made here were nevertheless grounded on a considerable amount of empirical evidence of how things are in the natural world. In particular, conceptual proposals were made with an eye on the observable characteristics of our cognitive and linguistic practice (i.e. on how we used to apply the relevant mental and physical symbols in various symbolic and situative contexts) and the expected theoretical benefits of the suggested conventions, while theoretical claims were made on the basis of explanatory considerations and some epistemologically fundamental (though fallible) observations of the intended phenomena. Beyond attempting to satisfy these empirical adequacy conditions, one major methodological consideration behind the search for a viable naturalist (as opposed to, say,

platonist) construal of *a priori* truth and knowledge was that the schematic picture delineated by such an account is expected to be both confirmable and further specifiable by the results of future empirical inquiries into the relevant segments of the spatiotemporal world. In other terms, the suggested representationist construal of *a priori* truth and knowledge differs from its referentialist alternatives not merely in that it is supported by our current observations of the world, but also in that it can gain evidential support as well as further sophistication from the future findings of the sciences of human brain and mind. By calling for such support and sophistication, an advocate of this construal may inspire various inquiries in the above sciences. The continuity between her philosophy, on the one hand, and her ordinary or scientific theory formation, on the other, does not reduce to the regular adjustment of the former to the empirical adequacy conditions imposed by the latter. Beyond this adjustment, she can also inspire and shape new empirical research programmes in the purported disciplines. The fact that her philosophy is methodologically naturalised does not entail that it can have no impact on the development of our empirical theories of the world.