the case of sets: Wright's point that an anti-platonist who deploys this argument against numbers should deploy it against sets too is certainly correct. Wright's claim that the argument has equal force against rabbits is more controversial: certainly advocates of causal theories of reference would claim that causal considerations do much to constrain the reference of 'rabbit', and it is precisely the fact that such causal considerations seem inapplicable in the case of numbers and sets that makes those so much more problematic.

But suppose that Wright is correct. Suppose, as he says, that we don't need to assume that 'where standard uses and explanations are insufficient to determine uniquely the putative reference of an apparently referential expression, that expression is not genuinely referential'. Presumably if we are to draw this conclusion it is because our notion of reference is disquotational. On a disquotational view of reference, we are entitled to keep the disquotation schemata

If b exists then 'b' refers to b

and

'F' applies to Fs and to nothing else

independent of any 'theory of reference'. These schemata allow us to assert that 'set' applies to sets, that 'rabbit' applies to rabbits, that 'number' applies to numbers, that '2' refers to 2. But the schemata do not allow us to assert that 'number' does, or that it does not, apply to sets; nor do they allow us to assert, or to deny, that if '2' refers to a set then it refers to $\{\emptyset, \{\emptyset\}\}$. To assert or deny these things, we would need not only disquotational reference but also a claim about the identity or non-identity of mathematical entities in different theories. And it was of course the issue of identities and non-identities, not the issue of reference, with which Benacerraf's argument was concerned; Wright's points are irrelevant to the argument that Benacerraf actually gave.

I do not deny that there is a counter to Benacerraf with some similarity to Wright's that needs addressing: the counter is that (not the problem of reference but) the pervasiveness of arbitrariness about identifications is a feature of the non-mathematical realm as well as of the mathematical. (That claim is, I believe, the most interesting feature of Putnam's critique of 'metaphysical realism', in his 1981 and elsewhere.) A typical example: it is sometimes said that it is arbitrary whether we take a point of space to be simply a region of minimal size, with zero (or infinitesimal) volume; or instead say that a point is a convergent set of smaller and smaller regions, each region in the set having non-zero (and non-infinitesimal) volume. This example, of course, will not impress the anti-platonist: if one rejects sets, then *literally* speaking there are no convergent sets of regions; moreover, the task of making do with only

regions of non-zero (and non-infinitesimal) volume in developing a non-platonistic physics is highly non-trivial. Whether other examples are immune to this sort of reply (and to various other replies) is not an issue I can pursue here, but my own view is that we do not get the same kind of pervasive arbitrariness at the (non-physical) level that we do at the mathematical.

B. *Knowledge of Mathematical Entities*

Perhaps the most widely discussed challenge to the platonist position is epistemological. Here the *locus classicus* is again a paper by Benacerraf (1973). Benacerraf's formulation of the challenge relied on a causal theory of knowledge which almost no one believes anymore; but I think that he was on to a much deeper difficulty for platonism.

Very roughly, Benacerraf's challenge can be put like this: if there are mathematical entities of the sort that the platonist believes in (mind- and language-independent, having no spatio-temporal location, unable to enter into physical interactions with us or anything we can observe) then there seems to be a difficulty in seeing how we could ever know that they exist, or know anything about them; the platonist needs to explain how such knowledge is possible, and no answer is evident except one that posits mysterious powers of access to the platonic realm. (Note that it is not *just* the acausal character of the mathematical entities that gives rise to the apparent problem; rather it is a combination of characteristics that collectively make access to the entities seem mysterious.) It may seem that if the previous section is correct then we have an answer to this epistemological question: we know about mathematical entities because theories that postulate them and attribute specific properties to them are indispensable in our various theories – for instance, in our physical theories. Of course, this assumes that mathematical entities *are* indispensable (to physical theory or to some other important body of extra-mathematical belief), and this is an assumption that a fictionalist (of my sort) would question. But I think that there is a more fundamental problem with this answer to Benacerraf: I think that we can formulate his challenge more carefully, so as to make indispensability considerations of questionable relevance in answering it.

The way to understand Benacerraf's challenge, I think, is not as a challenge to our ability to *justify* our mathematical beliefs, but as a challenge to our ability to *explain the reliability* of these beliefs. We start out by assuming the existence of mathematical entities that obey the standard mathematical theories; we grant also that there may be positive reasons for believing in those entities. These positive reasons might involve only initial plausibility, for those who are unconvinced

of my treatment of initial plausibility in section 2. Alternatively, the positive reasons might be that the postulation of these entities appears to be indispensable for some important purposes. But Benacerraf's challenge – or at least, the challenge which his paper suggests to me – is to provide an account of the mechanisms that explain how our beliefs about these remote entities can so well reflect the facts about them. The idea is that *if it appears in principle impossible to explain this*, then that tends to *undermine* the belief in mathematical entities, *despite* whatever reason we might have for believing in them. Of course, the reasons for believing in mathematical entities (in particular, the indispensability arguments) still need to be addressed, but the role of the Benacerrafian challenge (as I see it) is to raise the cost of thinking that the postulation of mathematical entities is a proper solution, and to thereby increase the motivation for showing that mathematics is not really indispensable after all.

I am aware, of course, that this sketch of the form I take Benacerraf's challenge to have is highly schematic. To fill it out, one would have to do four things. First, one would have to formulate more clearly the claim that our mathematical beliefs are 'reliable' or 'reflect the mathematical facts'. In essay 7 below I argue that in doing this we need not rely on any notion of fact, or even on any notion of truth beyond a thoroughly disquotational one: the claim is simply that the following schema

If mathematicians accept 'p' then p

(and a partial but hard to state converse of it) holds in nearly all instances, when 'p' is replaced by a mathematical sentence. The second thing one would need to do is argue that a platonist needs to accept this 'reliability' claim; I think, though, that the platonist's need to do this is beyond serious question. The third thing one must do is argue that a platonist must not only *accept* the reliability, but must commit himself or herself to the possibility of *explaining* it. The idea is that the correlation between mathematicians' belief states and the mathematical facts postulated in the above schema (and its partial converse) is so striking as to demand explanation; it is not the sort of fact that is comfortably taken as brute. (The platonist can legitimately postulate brute facts about mathematical entities themselves, for instance, basic laws of set theory; and even certain kinds of brute facts about the relations between mathematical entities and physical entities, for instance that every physical entity is a member of some set. But special 'reliability relations' between the mathematical realm and the belief states of mathematicians seem altogether too much to swallow. It is rather as if someone claimed that his or her belief states about the daily happenings in a remote village in Nepal were nearly all disquotationally true, despite the absence of any mechanism to explain the correlation between those

belief states and the happenings in the village. Surely we should accept this only as a very last resort.) Fourth and finally, to make it believable that the Benacerrafian challenge is insurmountable, one would have to argue that it is impossible to explain the reliability claim in question: one would have to argue that various facts about how the platonist conceives of mathematical objects collectively rule out all possibility of finding any such explanation. (The relevant facts about how the platonist conceives of mathematical objects include their mind-independence and language-independence; the fact that they bear no spatio-temporal relations to us; the fact that they do not undergo any physical interactions (exchanges of energy-momentum and the like) with us or anything we can observe; etc.) Like Benacerraf, I refrain from making any sweeping assertion about the impossibility of the required explanation. However, I am not at all optimistic about the prospects of providing it.

Several points about this are worth making here. First, it seems to me that something like the problem here under discussion has been a main motivation for various versions of what I've called 'mathematical idealism': that is, for various views according to which mathematical entities are some kind of 'mental constructions' (or 'constructions out of our linguistic practices'). Advocates of such views assume, I think, that it would not be hard to explain why our beliefs are reliably correlated with facts that we ourselves have constructed. Whether or not they are right about this is hard to say: talk of mathematical entities as 'constructed by' the mind (or by our linguistic practices) strikes me as so obscure that until it is explained, no answer is possible. As I remarked earlier on, it may be best to interpret such talk of 'constructions' as simply a picturesque way of saying that mathematical talk should be interpreted along fictionalist lines.[16]

[16] If one does not so construe them as restatements of fictionalism, it seems to me that there are two dangers to which they *may* be liable. (Whether they really are so liable depends on how talk of 'constructions' is to be understood.) The first danger is that one may not be able to make sense of all of classical mathematics if one tries to impose an idealist construal of it. (It was an idealist view of mathematics that led Brouwer and Heyting to intuitionism – see for instance Heyting 1956.) The second danger is that on a limited idealist view, one that views mathematical entities as some sort of human construction but makes no such claim about the physical world, the application of mathematics to the physical world may turn out to be a mystery. The danger, in other words, is that in order to explain the applicability of mind-dependent mathematical entities to the physical world, the idealist about mathematics may have to become a full-blown idealist, and hold that even things like electrons and dinosaurs are somehow 'human constructions'. If this danger were indeed realized, I would regard that as a *reductio ad absurdum* of the idea that mathematical objects were human constructions. I do not want to assert that it is impossible to develop a mathematical idealism that avoids both dangers and is genuinely distinct from fictionalism and succeeds in solving the epistemological problem (and various other problems) that the idealist finds with the platonist position; but I must confess to having little idea how it might be done.

A second point about the Benacerraf problem as I have reconstructed it: it is sometimes said that there is a need to explain the reliability of our beliefs about entities of a certain type *when the facts those beliefs report are contingent*; but that in the case of mathematical entities the facts in question hold necessarily, and this makes the task of explaining the reliability of our beliefs trivial or unnecessary. I respond to such views at some length in essay 7 below, and will say no more about them here.

Third, the fact that some mathematical claims may seem initially plausible is no help in responding to the version of Benacerraf's problem that I have sketched. Claims of initial plausibility are of some help to the platonist in answering questions about justification; as I argued in section 2, they are helpful in answering questions about the justification of particular mathematical beliefs, at least relative to a certain practice of making plausibility judgements. (I also argued there that this relative justification did not give them any special authority in contexts where that practice is itself questioned; but my present point is independent of this.) But to give them a justificatory role does nothing to explain the reliability of this class of judgements. Someone *could* try to explain the reliability of these initially plausible mathematical judgements by saying that we have a special faculty of mathematical intuition that allows us direct access to the mathematical realm. I take it though that this is a desperate move, rather akin to the move of postulating a special faculty of intuition that allows the character three paragraphs back direct access to the events in Nepal.

The fourth point I want to make is more concessive to the platonist, or at least, to the platonist who bases his or her platonism on some sort of indispensability argument – especially one who stresses the indispensability of mathematics in application to the physical world. For one can try to invoke indispensability considerations not simply in the context of justification, but in the context of explaining reliability. One could argue, for instance, that if mathematics is indispensable to the laws of empirical science, then *if the mathematical facts were different, different empirical consequences could be derived from the same laws of (mathematized) physics.*[17] So, it could be said, mathematical facts make an empirical difference, and maybe this would enable the application-based platonist to argue that our observations of the empirical

[17] This does not conflict with the conservativeness of mathematics: that has nothing to say about what happens when you apply mathematics to platonistic physical laws. (Also, it's only the actual mathematical facts that a platonist must admit are conservative. It isn't obvious that the platonist should have to agree that mathematics would still be conservative if the mathematical facts were different.)

consequences of physical law are enough to explain the reliability of our mathematical beliefs. An advocate of this indispensability line might even argue that initial plausibility judgements play an important role in explaining the reliability of our mathematical beliefs: the idea would be that evolutionary pressures (biological and/or cultural ones) have led us to find initially plausible those mathematical claims which are empirically indispensable, and that this gives all the explanation of the correlation between our judgements and the mathematical facts that we should want.

I'm suspicious about this line of response (with or without the extension that encompasses initial plausibility judgements) to the Benacerrafian challenge; but my most general worries about it involve some large issues, and I think it better not to attempt to raise them here. (I suspect that it is impossible to deal adequately with these general worries separately from some of the issues briefly touched on at the end of section 3, about the differences between the explanatory roles of mathematical entities and of physical entities.) But I will raise two more specific doubts about the prospects for dismissing the Benacerrafian challenge in this way. The first specific worry is that the amount of mathematics that gets applied in empirical science (or indeed, in metalogic and in other areas where mathematics gets applied) is relatively small. This means that only the reliability of a small part of our mathematical beliefs could be directly explained by the proposal of the previous paragraph. To be sure, one could try to use the reliability of our beliefs in this relatively small part of mathematics to 'bootstrap up' to the reliability of larger parts, by hypothetico-deductive inference within mathematics: see the discussion of the quotations from Gödel in essay 2.[18] But I think that there is substantial room to doubt that such inferences are all that powerful: too many different answers to questions about, say, large cardinals or the continuum hypothesis or even the axiom of choice work well enough at giving us the lower level mathematics needed in science and elsewhere. (One could of course just admit that we are and always will be ignorant of the mathematical facts about the continuum hypothesis and the axiom of choice and even the small large cardinals, but I don't think that this is an attitude many mathematicians would find attractive. In section 8C of essay 7, I describe an alternative and I think more appealing viewpoint toward these axioms, one which allows there to be reasons for preferring some axioms to others while denying that the choice is a matter of truth value about which we might be mistaken.)

[18] For a more thorough elaboration of a position like Gödel's, see Maddy (1988).

My second specific reason for doubting the adequacy of the reply of two paragraphs back to the reliability worry is really an extension of the first. The first worry began with the fact that the amount of mathematics employed in empirical science (and elsewhere, e.g. in metalogic) is relatively small. This is so quite independently of any partial successes of the programme of nominalizing science (and metalogic, etc.). But if, which I take as true, the partial successes of the nominalization programme have been substantial, this very much weakens the case for the reliability of the mathematical beliefs that we apparently need in those cases where the nominalization programme has not been carried out. Suppose for instance that we could nominalize everything but quantum theory. If this were so (and if my earlier critique of autonomous platonism is correct) then the entire weight of our belief in mathematical entities would rest on quantum theory. Is it really believable that an adequate account of the reliability of our mathematical beliefs could be made on this basis?

Of course, in actual fact quantum mechanics is not the only thing that has so far resisted nominalization, but the general point is clear: the more the partial successes of the nominalization programme, the more the difficulties for the attempt to respond to the Benacerrafian problem on indispensabilist lines, and therefore the more the motivation to try to complete the nominalization programme so that we can maintain a fictionalist view on which the Benacerraf problem does not arise.

PART TWO

5 Logical Implication

What should a fictionalist say about such metalogical notions as logical implication and logical consistency? The standard definitions of logical implication and logical consistency (due to Tarski 1956) are in terms of models. Suppose that Γ is a set of sentences, and B is a sentence; then

(i) Γ logically implies B if and only if B is true in every model in which all members of Γ are true;

(ii) Γ is logically consistent if and only if there is at least one model in which all members of Γ are true.

Models here are mathematical entities – they are sets of a certain kind – so a fictionalist cannot literally believe talk of logical implication or

logical consistency if this is what it means. (The fictionalist can also n literally believe the talk about a set of sentences Γ, but this is easier eliminate.)

But I think that there are reasons why even a platonist should questic that Tarski's definitions give anything like an adequate account of tl meaning of 'logically implies' or 'logically consistent'.

One difficulty is that the proposed definition of consistency lool too strong, and the proposed definition of consequence too weak. Th comes out clearly when one takes the sentences in Γ to be about set Suppose for instance that Γ is the set of all truths of set theory. Sin all members of Γ are true, Γ should surely be consistent. But is obvious that there should be a model in which all members of Γ con out true? Well, if there were a model whose domain was the set of sets, and in which 'ϵ' stood for the membership relation, then tl answer would surely be 'yes', since all members of Γ are true, th would be true in this model. But everyone knows that there is no s of all sets, so there can be no model of the sort just contemplated. if the set of all truths of set theory is Tarski-consistent, it is so by virt of some model that does not have the full set-theoretic reality in domain (and in which 'ϵ' may not even stand for the membersh relation). Why on earth should anyone believe that there is such model?[19]

Of course, if the language in which the members of Γ are formulat is a first order language, there is a complicated argument for the existen of such a model. First, we argue that since all members of Γ are tru there can be no derivation of a contradiction from Γ; this seems *prin facie* plausible, and I will not raise any questions about it here. B second, we must go from this conclusion to the existence of a mod that makes all members of Γ true. And at this stage, the arguments (l Gödel, Henkin, etc.) are quite complicated (they are variations on proc of the Skolem–Lowenheim theorem), and the models of set theory th produce are quite unnatural (for instance, in being countable, and there being no guarantee that what gets assigned to 'ϵ' looks very mu like membership). The fact is that it is only by virtue of an 'accide of first order logic' that the Tarskian account of consequence gives t

[19] It is no good objecting that one can allow models to be proper classes instead insisting that they be sets. Yes, one can do that in a set theory that recognizes prop classes, like Gödel-Bernays; but then there will be no class of all classes, in which c it is unobvious why there should be a proper class model for the set of all truths abc *classes*.