THE JOURNAL OF PHILOSOPHY

VOLUME CVII, NUMBER 4
APRIL 2010

Published by **The Journal of Philosophy, Inc.**

THE JOURNAL OF PHILOSOPHY
VOLUME CVII, NO. 4, APRIL 2010

VALUE MATTERS IN EPISTEMOLOGY*

In what way is knowledge better than merely true belief? That is a problem posed in Plato's *Meno*. A belief that falls short of knowledge seems thereby inferior. It is better to know than to get it wrong, of course, and also better than to get it right by luck rather than competence. But how can that be so, if a true belief will provide the same benefits? In order to get to Larissa you do not need to know the way. A true belief will get you there just as well.

Is it *really* always better to know the answer to a question than to get it right by luck? In part I we ponder: Is knowledge always better at least in *epistemic* respects? The affirmative answer is subject to doubts deriving from a conception of belief as sufficient confidence, but is defensible against such doubts. In our search for the special value of knowledge, we then explore in part II the relation between knowledge and proper action. Part III goes on to consider how the value-of-knowledge intuition acquires further interest through its equivalence with the view of knowledge as a norm of assertion. Finally, part IV steps back to examine what we might mean in saying that to know *is* always, necessarily better than to get it right by luck while really in ignorance. In order to defend our value-of-knowledge intuition we need first to understand it more clearly. Part IV offers an explanation.

I. CONCEPTIONS OF BELIEF AND THEIR BEARING ON THE VALUE PROBLEM

I.1. The Threshold Conception of Belief.

I.1.a. Your degree of confidence on any given question ranges between absolute certainty in the affirmative and absolute certainty

* My thanks for helpful comments to Jason Baehr, Matthew Benton, Alex Jackson, Jennifer Lackey, Matt McGrath, Christian Piller, Baron Reed, David Sosa, and John Turri, and to a referee for this JOURNAL.

0022-362X/10/0704/167–190

in the negative. You can believe without being certain, if you are confident enough, above some threshold; you disbelieve when your confidence lies below a certain threshold of confidence. The segment between these two thresholds corresponds to confidence that amounts neither to belief nor to disbelief. Here the thinker consciously suspends. (We restrict ourselves to cases in which the thinker consciously considers the question.)[1]

I.1.b. Compare an actual degree of confidence on a certain question with the ideal degree, given the subject's epistemic position, including his total relevant evidence. Actual degree of confidence should align as closely as possible with the ideal. The best degree of confidence to have on a question is of course the ideal degree. The status of one's attitude declines, moreover, in direct proportion to the distance between the actual degree and the ideal. Call this the *proportionality intuition*.[2]

How important epistemically is the distance from the actual to the ideal? In fact, it can be relatively *in*significant, as is suggested by the following case. Suppose that, given the evidence at his disposal, Diffident should be *extremely* confident, while yet his great intellectual caution makes him much less confident. His belief may then still be highly justified *epistemically*, with the sort of full justification relevant to whether one knows. Diffident's belief could then be justified, surely, even if he properly could be much more confident than he is, with justification to spare. Compare Normal, who has much slighter evidence than Diffident on the question at issue. Sufficiently weightier evidence could make Diffident better justified, and might even trump the fact that Normal's actual degree of confidence is perfectly aligned with his ideal degree.[3]

[1] In place of such thresholds, we might of course have twilight zones separating belief and disbelief, respectively, from suspension. The following line of reasoning applies in the first instance to the threshold conception but should extend to cover also the twilight conception.

[2] We abstract here from truth. A true belief would of course be better epistemically in respect of truth than a false belief even when each subject accords the ideal degree of confidence to his belief. The proportionality intuition leaves aside the epistemic value that attaches to a belief in respect of its truth or falsity.

[3] Objection: "The sufficiency intuition leads into a problem if we also accept something like a probabilistic coherence constraint. As my actual degree of belief in p falls short of the ideal degree of belief in p, my actual degree of belief in not-p will lie above my ideal degree of belief. As going over commits one via coherence to going under in the opposite direction, we should not accept any departure from the ideal degree (one might argue)." Reply: If your evidence warrants a departure of a certain length from .5, then it warrants any shorter departure about as well. If your confidence level in <p> is .6 while the ideal is .8, then your confidence level in <not-p> required by coherence would be .4, while the ideal confidence level here is .2. By our principle,

Such intuitions seemingly oppose the proportionality intuition. There is here a *sufficiency intuition*: that, once highly confident belief is justified, lower degrees of confidence are also about as well justified, so long as these lie above the belief threshold.

Consider a yes/no question that one has no basis for answering either way. Here suspension is ideal. If one has excellent reason to believe, by contrast, then high confidence is in order. Compare those two cases: If each actual attitude corresponds to its ideal attitude, the two actual attitudes—suspension in one case, high confidence in the other—are the same epistemically *in respect of proportionality*. Yet the confident belief is far better epistemically, especially if it amounts to a clear case of knowledge.

Accordingly, proportionality can provide at best a prima facie or partial reason for assessment of a given degree of confidence. Other factors can play a role, and might easily prevail. If a highly confident belief is close enough to being ideally justified, then it is better epistemically than ideal suspension on some other question. Even if suspension on that other question is better aligned with the ideal, the highly confident belief is still epistemically better on the whole.[4]

I.1.c. Suppose next that your ideal degree of confidence is higher, perhaps even much higher, than your actual degree. This might detract little, if at all, from your overall epistemic justification. So long as you do believe, with some positive degree of confidence, your belief is justified, even if you should be more confident.

---

then, you are justified in a higher degree of confidence than is ideal. True, but still it seems correct that you are playing it safer through this higher degree of confidence, since your departure from .5 is shorter than it might properly be. And this seems the right intuition.

[4] This bears on a kind of internalism/externalism clash. The proportionality intuition is in line with evaluation of a subject, Intern, on an internal dimension wherein what matters is just how well the subject acts epistemically with the materials in his possession (where in addition we suppose him to be guilty of no negligence in possessing just those materials). One clearly falls short, internally, in direct proportion to the distance between one's actual conduct and one's ideal conduct. Suppose now that Extern has far better evidence than Intern on the question at issue. We have found that a dominant notion of *epistemic justification* allows more of this quantity to Extern than to Intern, *even when Extern falls short of Intern on the internal dimension*. Extern performs less well than Intern with the materials at his disposal, where we assume that neither subject is relevantly negligent to the slightest degree. An internalist intuition wants to evaluate the subject just on the basis of his doing as well as possible given his situation. That intuition, we can now see, does not give us the entire content of an intuitively plausible conception of epistemic justification. The evidence at the subject's disposal matters equally, and this is a factor beyond the subject's relevant control, once we assume him to be free of negligence. It is, in a relevant sense, an *external* factor.

Suppose *S*'s actual degree of confidence to be just slightly above his ideal degree. Normally, this has little effect on the epistemic standing of that degree of confidence. But there is an important exception: *when the belief threshold falls between the subject's actual degree and his ideal degree*. Something remarkable happens in this case. Now the subject believes what he should not believe: his belief is *un*justified! Take a subject with some actual degree of confidence barely above the belief threshold. If the ideal degree of confidence is also barely above the actual degree, then the small distance between the two is inconsequential. However, if the ideal degree lies below the actual degree, and *also* below the belief threshold, then the distance between them does matter, in direct proportion to its size. What accounts for this remarkable power of that particular threshold point in the scale of confidence?

These tensions are hard to resolve if we insist on regarding that particular point as a threshold, *and nothing more*. It is not easy to find a proper rationale that accommodates our intuitions while restricted to this "mere threshold" conception of belief. Moreover, this concern immediately spills over to our value-of-knowledge problem, to which we turn next.

I.1.d. Consider the following (where the sort of justification involved is, throughout, the *epistemic* sort, as is the relevant sort of evaluation).

(KA) *The Knowledgeable Answer Platitude*

> If one takes up a question, it is epistemically better to know the answer than not to know it. More specifically: One's conscious answer to the question is epistemically better than one's conscious suspension of judgment, provided one's answer constitutes knowledge.

(JA) *The Justified (Competent) Answer Platitude*

> If one takes up a question, it is epistemically better to have an answer than not to have an answer, provided the answer is justified (competent). More specifically: One's conscious answer to the question is epistemically better than one's conscious suspension of judgment, provided one's answer is justified (competent).

Compatibly with KA and JA, there are *pragmatic* dimensions of evaluation in which one might be better off by lacking knowledge and justification than by having them.

Despite their initial plausibility, KA and JA are problematic under the threshold conception of belief. Suppose Diffident is confident of a certain proposition to a degree just barely *below* the belief threshold, while Assertive is confident to a degree just barely *above* it. And suppose Assertive is justified in his belief, which even constitutes knowledge.

Is Assertive thereby epistemically better off than Diffident? More specifically, is Assertive's belief epistemically better than Diffident's conscious suspension on that same question? Is it not better epistemically to have a knowledgeable answer for a question than not to be able to answer the question?

Not only does Assertive barely believe as he does, however; suppose further that he is also just barely justified in so believing, with just enough evidence. By contrast, Diffident has a wealth of evidence. It is only his intellectual diffidence that keeps his degree of confidence just below the threshold of belief. Diffident then seems justified, indeed better justified than is Assertive. Remember, the difference in their degrees of confidence is *vanishingly small*, even though Assertive's is just above the belief threshold, while Diffident's lies just below.

A belief might after all be only *marginally* more confident than a conscious suspending on that question, while yet the suspending subject manifests better epistemic competence through his strong inclination to believe than does the believing subject through his weak belief. Both subjects have nearly the same degree of confidence, one just above and one just below the threshold. Diffident's degree of positive confidence is supported by a wealth of evidence, while Assertive has a paltry basis. Diffident then seems better off epistemically despite the fact that his confidence level falls just short of belief, whereas Assertive's confidence level lies just above the threshold and does barely constitute belief.

I.1.e. Also now in doubt is the following:

(AB) *The Apt-Belief Platitude*

> If one takes up a question, it is epistemically better to answer that question aptly than not to answer it at all. More specifically: One's apt answer to the question is epistemically better than any attitude that falls short of that, amounting only to suspending judgment and not venturing an answer.

Here we find a problem similar to those encountered earlier. A subject with a positive belief might be only marginally more confident than one who suspends on that question, while yet the suspending subject manifests more epistemic competence in getting it right through his strong *inclination* to believe than does the believing subject through his weak outright belief. After all, an inclination to believe can also be apt, if it is strongly positive, and also veridical, and it can even manifest epistemic competence by being veridical. (An inclination to believe is tantamount to a positive confidence level, above .5 but below the threshold of belief.) And, again, Diffident falls only slightly below Assertive in actual degree of confidence, while his ideal degree is much higher, since his body of evidence is vastly weightier on balance.

I.1.f. Platitudes KA, JA, and AB are thus in tension with the threshold conception of belief (and with the threshold conception of the epistemic attitudes generally: belief, disbelief, and suspension).

I.2. Affirmative versus Threshold Conceptions of Belief. We turn now to ideas that may help in our search for an alternative, ones with some independent interest of their own.

I.2.a. Consider a concept of *affirming that* p, defined as: concerning the proposition that *p*, either (a) *asserting it publicly*, or (b) *assenting to it privately*. For the present inquiry, let us take these notions as given.

I.2.b. So we have two ways to conceive of belief: *threshold-belief*, belief as sufficient confidence (above a threshold); and *affirmative-belief*, belief as disposition to affirm (as defined above). Two people might coincide in threshold-belief, since they share the same degree of confidence, while diverging in affirmative-belief, since one is naturally more assertive, the other more diffident. Correlatively, two people might coincide in affirmative-belief while diverging in threshold-belief.

I.2.c. Some advantages of the affirmative conception: Consider the bearing of this alternative conception on the platitudes that proved problematic for the threshold conception: KA, JA, AB. Take now any slight difference in degrees of confidence placed in one and the same proposition at any point across the belief spectrum. No such difference would seem any more significant than any other. If so, no special significance should attach to a slight enough difference that encloses a threshold. If the threshold is a mere threshold, that is how it seems.

Compare with that the difference between being disposed to affirm and not being so disposed. The importance of *this* difference might derive from the value of one's being a source of assertions and now a source of information for others. This seems a distinctive epistemic value of the state of belief defined as disposition to affirm. What constitutes this value would need to be clarified, since it is possible to talk too much, so as to pollute the dialectical space. The point concerns rather a necessary condition for a great good, the sharing of good, reliable information on matters of interest or importance. Without the disposition to affirm, there is no such sharing.

In addition, if one lacks the disposition to affirm, then one will be unable to use one's belief in conscious reasoning towards actionable or knowledgeable conclusions, regardless of how confident one may be. After all, such reasoning requires the affirmation of premises.

Whether or not that explanation pans out in the end, affirmative-belief escapes the discrimination problem of threshold-belief, in which the relevant degree seems indistinguishable from infinitely many others. The affirmative conception avoids this problem, since it does not make belief depend just on a particular point in the spectrum, one that seems epistemically insignificant in itself.

The affirmative conception can now be related to the threshold conception in either of two ways. The threshold can be allowed to vary from subject to subject and to be set by when the subject acquires the relevant disposition to affirm. Under this approach threshold-belief cannot possibly diverge from affirmative-belief. On another alternative, the threshold is set the same for every subject. What it takes to believe on this conception is to have a degree of confidence above the threshold. It is this alternative conception that runs into the difficulties laid out above.

In any case, on the affirmative conception we can still wonder what endows a disposition to affirm with its epistemic interest. We need this explained regardless of how the affirmative conception may be related to the confidence spectrum and to thresholds in that spectrum. Our suggestions about deliberative and social epistemic values aim to help fill this need.

II. HOW IS KNOWLEDGE CONNECTED WITH ACTION?

Consider the normativity that is fully constitutive of knowledge, the normative status and level that a belief must attain in order to constitute knowledge. Such normativity is a special case of performance normativity. Take any performance with an aim. If it is successful (by attaining its aim), then it is, let us say, "accurate." Moreover, if that performance is competent (if it manifests competence), then it is "adroit." And, finally, if its *success* manifests the competence manifest in the performance, then it is "apt." So we have an AAA structure under which performances (with an aim) generally can fall. Beliefs are a special case of such performances. They are cognitive performances that can be aimed at truth, and can then be apt by attaining that aim while thereby manifesting the believer's cognitive competence. In those cases they amount to knowledge on a first order: animal knowledge.

Just as beliefs are subject to the Gettier phenomenon, so are performances generally subject to a generalized Gettier phenomenon. The case of the archery shot that attains success through accidentally compensating gusts is a case in point. It is a shot that is accurate and adroit without being apt. A Gettiered belief is a special case of that general Gettier phenomenon. It is a belief that is accurate and adroit without being apt.

How is knowledge normatively related to action? Consider means-end action, of the form: *X*'ing in the endeavor to *Y*, as a means to

Y'ing, with the aim of *Y*'ing. Let us begin with such action, and perhaps generalize eventually. But let our initial conception of the relevant "means" be very broad, to include not only causally instrumental means, as when one flips a switch as a means to turning on a light, but also other sorts of means, as when one raises one's hand as a means to voting. Also for now, let us restrict ourselves to "definitely safe" means and exclude those that are merely "probabilifying."

A means-end intended action is constituted by a means-end belief. And if the intended action is successfully carried out, then the carried-out means-end action essentially involves that means-end belief.

Turn now to the evaluation of an intended means-end action. Say the agent flips a switch as a means to turning on a light. Such a performance with an inherent aim falls, of course, under our AAA structure. Among the things that constitute the relevant competence is the means-end belief involved. The competence whose manifestation might make the performance adroit or competent includes that belief on the part of the agent. Accordingly, that belief would need to be competent in order for the performance to be competent. (Even if this does not follow deductively, it seems plausible enough.)

Suppose the means-end belief is epistemically competent but not apt. Suppose it is Gettiered. It is competent and even true, but its correctness is due to luck and manifests no relevant competence. In that case, I submit, the means-end action itself fails to be apt. It falls short in this performance-normative way. It may attain its aim, and may even manifest competence: that is, the performance may manifest an overall competence that would include, in part, the epistemic competence manifest in the formation of that means-end belief. However, if the means-end belief essentially involved is not apt, if it hits the mark of truth by luck, then the performance itself also fails to be apt. It itself attains success by luck, in a way that is relevantly deplorable. Hence, it falls short in this performance-normative respect. The performance falls short simply because its success is *in that way* attributable to luck rather than fully enough to competence.

So we have a normative connection between knowledge and nonbasic, means-end action. Fortunately, it is now easy to generalize to basic action, as well, if basic action counts as a limiting case. Thus, a basic action of *X*'ing will be an action in which one *X*'s in the endeavor to *X*. Knowledge of this means-end proposition is easy to attain; indeed, it is hard to avoid. It is obvious that one can *X* by *X*'ing, if *X*'ing counts as a limiting case of a means to *X*'ing.

Moreover, we can obtain the further result that one's action falls short if it is based on ostensible reasons that one does not know to be true. What accounts for this result? Is it just that a proposition can constitute "your reason for *X*'ing" only if it is something you know to be true? No, this seems a fairly superficial feature of English. A better explanation derives from a deeper, closely related truth. This can be put in terms of one's rationale, of one's ostensible reasons, or of propositions adduced as reasons, or of stative reasons: that is, beliefs on which one bases some further belief, or some choice or decision. The deeper normative truth of interest is that if one acts based on an ostensible or adduced reason, then one's action falls short if the adduced reason is not something one knows to be true.

When someone flips a switch as a means to turning on a light, for example, he has an ostensible reason on which (in a broad sense) he bases his action, namely, that flipping the switch is a means to turning on the light. Now, any action taken as a means to a further objective will, of course, fall short if it does not bring about that further objective. Moreover, it will still fall short if the objective is attained by a certain kind of luck: that is, in a way that does not manifest the agent's competence. Suppose the relevant means-end belief to be true: I mean flipping the switch is a means to turning on the light. But suppose that belief to be competently acquired but Gettiered, so that it is correct only by epistemic luck. In that case, I say, flipping that switch still falls short, not because it does not bring about the light's going on, but rather because it brings it about in a way that does not fully enough manifest the competence of the agent, and thus is an inapt performance.

Inapt performances fall short not only in that they might have been *better* on relevant dimensions. They fall short in the fuller sense that they fail to meet minimum standards for performances. Because they are inapt, they are therefore *flawed*: not just improvable, but defective.

III. A KNOWLEDGE NORM OF ASSERTION?

Note next that assertion is itself an action. And suppose sincerity to be an epistemic norm of assertion. Suppose, that is to say, that an assertion falls short epistemically if it is insincere. As members of an epistemic community we are acting improperly if, in asserting, we lie rather than give voice to what we believe. Jennifer Lackey has argued that a creationist teacher might assert with full epistemic propriety when in her classroom she asserts propositions of evolutionary science that she does not believe. That is a very interesting case, which I propose to accommodate by means of a distinction between assertion in one's own person, as a human being who communicates with other human beings, and assertion as occupier of a role. As a newscaster or as a teacher one may be called upon to say things, and thereby to assert them, as in the classroom or in a newscast, even when one does not believe what one says. One may still proceed with epistemic propriety if one is playing one's epistemic role properly. To play one's epistemic role in such contexts may just require reading (assertively) from a script or from a teleprompter, or reporting from memory, where one serves as a mouthpiece for a deeper institutional source of the information conveyed, the deeper source that is the school or the news organization.

So I will assume that sincerity is a norm of assertion in one's own person, where one is not playing a role in some epistemic institution (for the delivery of information or the like). This means that in order to avoid falling short epistemically, assertion must be in the endeavor to assert with truth. To assert in disregard of what one takes to be true is to assert insincerely. Obviously, one is insincere if one asserts what one disbelieves. But one is not fully sincere even if one asserts what one fails to believe. So, to assert sincerely is to assert in the endeavor to thereby assert with truth, in line with what one takes to be the truth of the matter. And now our results concerning the propriety of means-end action apply to assertion as a special case. If one asserts that *p* as a means to thereby assert that *p* with truth, this essentially involves the relevant means-end belief. I mean the belief that asserting that *p is a means* to thereby assert with truth. And this belief is equivalent to the belief that *p*. Accordingly, if that means-end belief needs to amount to knowledge in order for the means-end action to be apt, then in order for a sincere assertion that *p* to be apt, the agent must know that *p*. In this way, knowledge is a norm of assertion. If an assertion (in one's own person) that *p* is not to fall short epistemically, it must be sincere, and a sincere assertion that *p* will be apt only if the subject knows that *p*. This is, moreover, not just a norm in the sense that the subject does better in his assertion that *p* provided he knows that *p*. Rather, if his assertion is not apt, it then violates minimum standards of performance normativity. Any performance (with an aim) that is inapt is thereby *flawed*.

That, then, is a way in which knowledge can figure as a norm of assertion more importantly than certainty. A performance may, perhaps, be even better if it involves certainty on the part of the performer that his means are means to his end in so performing. Compatibly with that, however, the performer meets minimum standards if his performance is apt, if its success manifests knowledge on his part. It need not be *flawed* even if the knowledge that it manifests does not amount to certainty.

Knowledge is said to be necessary for proper assertion. The propriety here must, of course, be *epistemic*. One *can* appropriately lie to a murderer looking for his weapon. So, the claim is that in order to assert with full *epistemic* propriety or worth you must know the truth of what you assert. And this now seems just one side of a coin whose other side is our value-of-knowledge intuition. That these are two sides of a coin gains plausibility through our conception of belief as disposition to affirm. If knowledge is the norm of assertion, it also plausibly is the norm of affirmation, whether the affirming be private or public. Affirmation that *p*, moreover, seems epistemically proper and worthy if, and only if, the disposition to so affirm is itself epistemically proper and worthy.[5]

We can now argue as follows:

(i) Knowledge is the norm of affirmation: that is, to affirm that *p* with full epistemic propriety or worth requires knowing that *p*.
(ii) Knowledge is the norm of belief: that is, to believe that *p*—to be *disposed* to affirm that *p* with full epistemic propriety or worth—requires knowing that *p*.
(iii) It is epistemically better to believe with full epistemic propriety or worth than to believe without such propriety or worth.
(iv) Therefore, knowledge is epistemically better than merely true belief, which is true belief that falls short.

And we can reverse direction as follows:

(v) Knowledge is epistemically better than merely true belief, which is true belief that falls short.
(vi) To believe that *p*—to be disposed to affirm that *p*—without falling short requires not merely believing correctly that *p*; it requires believing aptly that *p*, that is, knowing that *p*.
(vii) Knowledge is the norm of belief, of disposition to affirm that *p*: that is, to believe that *p* with full epistemic propriety or worth requires knowing that *p*.
(viii) Knowledge is the norm of affirmation: that is, to affirm that *p* with full epistemic propriety or worth requires knowing that *p*.

If each of (ii)–(iv), and each of (vi)–(viii), is made plausible by its predecessor, this argues for the equivalence of the knowledge norm of assertion and the value-of-knowledge thesis (that knowledge is better than merely true belief). The second half of the reasoning—from (v) to (viii)—gains plausibility, of course, if we replace 'the norm' with 'a norm', while the first half loses no plausibility through that replacement. Recovering the stronger claim would then require explicating

[5] Where the propriety of the former might even derive from the propriety of the latter, in the way skillful performance derives from the relevant ability or disposition of the agent to issue such performances; a performance might of course be skillful even when it happens to fail, perhaps due to unforeseeably unfavorable circumstances.

the respect in which this norm, the knowledge norm, is relevantly *distinctive*. And that might well rely on its claim to being the most fundamental relevant norm, from which others, such as a justification norm or a truth norm, might then derive. These issues are beyond our present scope. We are only pointing to a plausible equivalence that deserves to be explored.

Obviously, it may turn out that only the weaker equivalence is ultimately defensible: the equivalence between the value-of-knowledge thesis and the thesis that knowledge is *a* norm of assertion (not necessarily *the* norm). Compatibly, it may also turn out that only *certainty* or, alternatively, *knowing that one knows*, qualifies as *the* distinctive norm, the one that will explain all the others.[6] But why not KKK*p* rather than KK*p*? Indeed, good question! Yet the regress need not be vicious, if there is some top limit imposed by human limitations, even one that varies from subject to subject. That then would be the relevant norm for any given subject. Why think of this as a "norm"? It seems in the first instance a *standard*, not a *guiding principle*, even if it might on occasion serve in the latter capacity as well. Why KK*p* is a proper, higher standard emerges when we consider the effect on one's object-level belief of one's own take on whether one thereby knows. *Disbelieving* that one thereby knows reflects poorly on one's belief, as does *suspending judgment* on whether one knows. Clearly, it is epistemically better to affirmatively defend one's object-level belief as a case of knowledge (or even to be *able* to defend it, in the sense of having a defense at the ready). Plausibly, moreover, this epistemically enhances the object-level belief itself. And the same would seem to hold for any level to which the subject is able to ascend, above the object level.

Still, there is the following concern. *Evidence than which none greater is available* is not plausibly viewed as a norm of assertion or of belief. But if a belief falls short of such evidence, then it does fall short epistemically. Thus there is a better level to which one might have ascended in one's belief, with greater effort. So long as one was not negligent, however, that fact does not make one's belief epistemically reproachable or even flawed. It is not a flawed or faulty belief just because it might have been even better founded on a richer fund of evidence. The worry now is that the higher levels are like the greater available evidence. Yes, it would have been to the credit of the believer and would have added to the worth of the belief had it been guided by such higher-level knowledge on the part of the believer. The belief

[6] Compare David Sosa, "Dubious Assertions," *Philosophical Studies*, CXLVI (2009): 169–72.

would have been a better belief, in one clear, epistemically relevant respect, had it been guided by the believer's knowledge of his competence and situation. What is more, this is nothing peculiar to cognitive performance. Any performance with an aim is a better performance, in one clear performance-evaluation respect, if it is not only apt but also fully apt, that is, one whose aptness manifests the agent's meta-aptness, the agent's informed take on his relevant competence and situation. Is the thoughtlessness of an agent who acts "on automatic pilot" reproachable, and does it detract from his performance? Not always, surely. Much of what we do is done on automatic pilot, without being reproachable just for that reason. This suggests that while the K norm is true in full generality, the KK norm is true at most more restrictedly, when the issues are important enough to demand special care.[7]

In light of our most recent reasoning, it should be clear why our equivalence argument is better stated with certain qualifications, as follows:

(ix) Knowledge is the epistemic norm of affirmation: that is, to affirm that *p* without epistemic defect requires knowing that *p*.
(x) Knowledge is the norm of belief: that is, to believe that *p*—to be *disposed* to affirm that *p*—without epistemic flaw requires knowing that *p*.
(xi) Merely true belief is defective by comparison with the corresponding knowledge.

And, reversing direction:

(xii) Merely true belief is defective by comparison with the corresponding knowledge.
(xiii) To believe that *p*—to be disposed to affirm that *p*—without epistemic defect requires knowing that *p*.
(xiv) Knowledge is the epistemic norm of affirmation: to affirm that *p* without epistemic defect requires knowing that *p*.

Performance norms come in three sorts: (a) Assessment norms specify dimensions for the evaluation of a performance. (b) Minimal standards are criteria for the determination of satisfactory performance. A performance can fall short in such a way that it is a flawed performance, however, without the performer being at fault. (c) Criticism norms are norms whose violation not only makes the performance flawed but does redound to the discredit of the performer, who is thereby at fault.

[7] "Suggests," I say, advisedly, since alternatively one might say that no human ever believes, or even performs more generally, in complete disregard of his relevant competence and situation. The depth of reflection in the *Meditations* is not constantly required as a self-check, of course, but that does not mean that *nothing* is required, not even below the surface of consciousness.

So, there are three ways of falling short in one's performance, corresponding to those three norms (which we take in reverse order).

(1) The performance can be *discreditable*: that is, it can fall short and redound to the discredit of the performer, who is thereby at fault.
(2) The performance can be *flawed though not discreditable*: that is, it can fall short of even minimal standards—can fall short of a threshold for satisfactory performance—so that it is somehow defective, but the defect may not be the performer's fault.
(3) The performance can fall short of a higher level of performance that it might have attained, but it can fall short thus *without being either discreditable or even flawed.*

Take a batter who strikes out against a superb pitcher. As he swings and misses, his swing may or may not be flawed. Clearly it falls short, since it does not even connect with the ball. Is it flawed? Well, does it fall short of some minimal standard? It does perhaps if the batter was distracted avoidably on that particular pitch, so that he took his eye off the ball. But is it also his *fault* that he misses? This is not so clear. He may be a great batter, and this swing may be a fairly normal swing on his part. And batters are not required to be fully attentive on absolutely every pitch. They are cut some slack. So this is just one of those swings where our batter misses, even though he has the best batting average in the history of baseball and he is in his prime. He might of course have been at fault if he had downed a double martini just before game time. Then the performance might well have been not just flawed but also discreditable.

For another example, consider a dish prepared by following a recipe. The chef follows the recipe, which is obtained from an excellent source that he has every reason to trust and none to distrust. The stove is also one that he has every reason to rely on and none to distrust. Nevertheless, the recipe is defective, and the stove is also defective. Suppose each defect boosts the bad effect of the other, so that a dish that would have been overcooked is actually burned to a crisp. In that case, the performance is flawed but it may be no fault of the performing chef. It is still a defective performance though not reproachable. Moreover, if the two defects cancel each other out, then, I submit, the performance is still flawed, even if the dish comes out fine. It is a successful performance, but not an apt one. It is Gettiered, and succeeds by luck, and not in a way that manifests the chef's complete competence, the kind of competence that requires not only inner, constitutive competence, but also particular external aids.

It may be argued that a means-end performance can be apt even if the agent does not know that the means will definitely lead to the end.

But the foregoing commits us to nothing so strong. It all depends on whether something can be a means without being a fail-safe means. So far we have restricted ourselves to definitely safe means. But it seems reasonable also to recognize probabilifying means. We seem often to act on probabilifying means that do not guarantee success. When a batter swings he might have no better than a 15-percent probability of success. Indeed, that would be a highly competent batter. So, if his swing connects and results in a base hit, that success can surely be apt, even if the belief that swinging as he did would be a means to attaining a base hit could not plausibly have been the belief that it would be a definitely safe means. Rather, it may have been just the belief that it would *sufficiently probabilify* the attainment of the objective. Similarly for Diana as she hunts with her bow and arrows. Similarly for athletes generally. Aptness cannot require infallible competence. What is more, it cannot even require probability above 50 percent!

Compatibly with the foregoing, however, we can still plausibly require in an apt means-end performance that the agent know that the means are *likely enough* to secure the objective. The agent cannot just be taking a wild guess, a shot in the dark, and thereby manifest competence in his success. Arbitrary wild shots that score do not thereby manifest competence, if there is no glimmer of competence in the beliefs constitutive of that competence. If the constitutive belief is a probabilistic belief, however—say, that the chosen means are sufficiently likely to produce success—then we can ask about the status of that belief, and the AAA structure will be relevant. The means-end performance will then plausibly be apt only if the means-end belief constitutive of that competence is itself apt. The fact that we are now allowing probabilistic beliefs as constitutive means-end beliefs does not affect that requirement. The relevant probabilistic belief must itself be apt if the means-end action essentially involving that belief is to be apt.[8]

The foregoing has argued for the equivalence of certain knowledge-norm-of-assertion intuitions and corresponding value-of-knowledge intuitions. Earlier, we had also defended the view that knowledge is

[8] One might well wonder what is involved in the belief that the means are *sufficiently* likely to secure the end. And this is a place where pragmatic issues can enter very naturally and appropriately. However, consider pragmatic determination of whether the means "sufficiently" probabilify the end. Such pragmatic determination does not necessarily entail pragmatic encroachment on the determination of the level of competence or aptness required of the belief that the relevant means *are* thus sufficiently probabilifying of the end.

better than merely true belief from problems deriving from a particular conception of belief—the threshold conception. Doubts still remain, however, as will be seen presently, in part IV.

IV. THE VALUE PROBLEM: A STEP BACK

What is involved in epistemic evaluation? Can we clarify our meaning when we say that it is always "better" to know? Is it really clear that knowledge *is* always better than merely true belief, better at least epistemically?

In contemporary epistemology, this value problem has moved to center stage. Plato wondered how knowledge can be more valuable than its corresponding true belief, if a true belief will serve you equally well. True beliefs will guide you to your objectives no less efficiently than would the corresponding knowledge. In line with this, we ask: How can knowledge as such always improve on the corresponding merely true belief?

We assume that in order to constitute knowledge, a belief must satisfy some condition beyond being a belief and being true. If knowledge that *p* is always, necessarily better than merely true belief that *p*, then the additional condition must import some normatively positive content. And this further content should help explain how it is that knowledge is, as such, always better. When one ponders a question, for example, it would always be better to answer knowledgeably than to answer correctly but just by luck.

IV.1. The Value of Knowledge. The aim of belief is said to be truth. This is normally correct. When you pose a question to yourself, for example, you want a correct answer. When you reach an answer in adopting a certain belief, the aim of your belief is the truth of the matter. If the aim of a belief is thus truth, however, as it normally is, then *once true* that belief would seem to have what matters epistemically, irrespective of its etiology.

How then can a truth-reliably produced true belief be better than one that is no less true, regardless of how reliably it may have been produced? Conclusion: Knowledge is really no better as such than merely true belief.

"Any argument leading to that conclusion," comes the reply, "must have its premises examined. For one thing, perhaps the aim of belief required for knowledge is not just truth, but also knowledge. This would explain how and why it is that knowledge (with its required etiology) is after all better than merely true belief."[9]

[9] Here and in what follows, talk of "the aim of belief" is implicitly restricted to normal cases.

What follows will defend this reply by placing it in context, by explaining its content, and by drawing some implications.[10]

IV.2. How Indeed Is Truth Our Aim? How should we understand the value that we place on it? More explicitly, when we aim at truth, our aim presumably is to *have* the truth. So, it is the attained truth that has corresponding value. How then should we more fully describe our true objective? Is it just the accumulation of believed truths? Compare how we assess accurate shots, those that hit their targets. What is it that people value under this rubric? Is it the accumulation of accurate shots?

Someone casually draws a large circle on the beach right by his feet, aims his gun, and hits the target. Does he thereby attain, at least in some small part, a previously standing objective: namely, that of securing accurate shots? Is that an objective we all share, given how we all share the concept of a good shot? "Well, don't we all want good things (other things equal)? Aren't *good* shots good things?" This response, I trust we agree, is quite absurd.

The shot at the beach could be an accurate, good shot, nonetheless, as the marksman hits his target in the sand. Although, from one point of view, given the low or even negative value of the aim, this accurate shot has little value of its own, yet from another, performance-internal perspective it is graded as quite accurate, a good shot, maybe even an excellent shot if the marksman steps back far enough from the target. Even when the shot is difficult, however, its status does not derive from any standing preference of people for an accumulation of accurate, difficult shots. There is no normative pressure on us to bring about good shots, not even if we grasp perfectly well what it takes to be a good shot, and have this uppermost in our consciousness at the time. There is no *inherent* normative pressure to bring about even excellent shots, none whatever that I can discern. (What we are *not* normatively pressured to accumulate *for their own sake,* note well, is shots, not even excellent shots.)

Compare now our intellectual shots, our beliefs. A belief may answer a question correctly, but may have little value nonetheless, if the question is not worth asking. The value of its target, or of reaching it, will surely bear on the worth of any shot so aimed. Arbitrary selection of an area by your feet at the beach yields a silly target. Similarly, suppose you scoop up some sand and laboriously count the grains. You then take up the question of how many grains are contained in

[10] Moreover, the framework of performance normativity will be seen to accommodate also a broader reply that requires belief to aim at truth, but not necessarily at knowledge.

that quantity of sand. If you attain a correct answer, what is your performance worth? Do you thereby fulfill, at least in some small part, a previously standing objective, that of securing more and more true beliefs? This is no more plausible than is the corresponding view about the shot at the beach.

IV.3. In What Way, Again, Does the Truth of Our Beliefs Have Value? One thing that does plausibly have prima facie value is the satisfaction of our curiosity. Take again the silly question as to the number of grains of sand. If someone gets interested in that question anyhow, then the satisfaction of his curiosity will in an obvious way have value to him, which is to say just that he values it. And perhaps, to some small extent, it will even have some value for him by making his life better in that small respect. This is of course a way for the truth to have value to someone and for someone. After all, if one is curious as to (a) whether *p*, this is just to be curious as to (b) whether it is true that *p*.[11] So, what we want when we value the truth in that way is to have our questions answered, and of course answered correctly.

Sheer curiosity, whatever its basis, thus invests the right answer to a question with some value, though the value might be small and easy to outweigh, as with the question about the grains of sand. Having the answer to that particular question may add so little to the life of the believer, while cluttering his mind, that it is in fact a detriment, all things considered, if only through the opportunity cost of misdirected attention.

Similar considerations apply to the shot aimed from a foot away at the sandy beach. The sheer desire to hit that target, whatever its basis, gives value to the agent's hitting the mark. Still, hitting that mark might import little value for anyone. Spending his time that way may even be a detriment to the agent's life. Nor is it plausible that we humans have generally a standing desire for accurate shots, nor that we place antecedent value on securing such shots. Accuracy will give value to the shot at the sand only dependently on the gunman's whim to hit that particular target.

Even if that shot at the beach fulfills no human interest antecedent to the gunman's whim, it may still be a better shot, better as a shot, than

[11] Two distinctions need distinguishing: first, that between (a) and (b); second, that between (c) whether it is true that *p*, and (d) whether one's belief (whose content is that *p*) is a true belief. Many are the ways one could wonder whether one's belief is true without specifying its content. One could wonder whether that belief is true even when it is picked out by description, with no notice taken of its propositional content. (This latter distinction is important for understanding Descartes's epistemology, or so I argue in "Descartes and Virtue Epistemology," forthcoming.)

many with higher overall value. Take a shot at close quarters in self-defense that misses the targeted head of the attacker but hits him in the shoulder and stops the dangerous attack. A bad, inaccurate shot is this one, but more valuable than the accurate shot at the beach. (Had it been better as a shot, moreover, a more accurate shot, it might have constituted a terrible murder, since the attack did not justify shooting to kill.)

Are beliefs like shots in that respect? Is a belief a performance that can attain its internal aim while leaving it open whether it has any intrinsic value, and whether it serves or disserves any external aim? Let us explore this view of belief.

IV.4. Knowledge as a Special Case of Apt Performance: An Account of Its Special Value. A performance that attains its first-order aim without thereby manifesting any competence is a lesser performance. The wind-aided shot scores by luck, without thereby manifesting competence. It is hence a lesser shot by comparison with one that hits the mark and thereby manifests the archer's competence.[12] A blazing tennis ace is a lesser shot if it is a wild exception from the racket of a hacker, by comparison with one that manifests superb competence by a champion in control. And so on. Take any performance with a first-order aim, such as the archery shot and the tennis serve. That performance then involves also the aim of attaining its first-order aim. A performance *X* attains its aim <*p*>, finally, not just through the fact that *p*, but through the fact that it brings it about that *p*.[13]

The case of belief is just the special case where the performance is cognitive or doxastic. Knowledgeable belief aims at truth, and is accurate or correct if true. It has, accordingly, the induced aim of attaining that objective. Therefore, such belief aims not just at accuracy (truth), but also at aptness (knowledge). A belief that attains both aims, that of truth and that of knowledge, seems for that reason better than one that attains merely the first. That, then, is a way in which knowledge as such seems plausibly better than merely true belief.[14]

[12] A shot might manifest an archer's competence without its accuracy doing so. The shot with the two intervening gusts is a case in point. How does that shot manifest the archer's competence? By the arrow's having, when released, a direction and speed that would take it to the bull's-eye, in relevantly appropriate conditions.

[13] Just as its being true that *p* entails its being true that it is true that *p*, so one's bringing it about that *p* may entail that one brings it about that one brings it about that *p*, assuming such iteration always makes sense. It might be objected that one can bring it about that someone else brings it about that *p* without oneself doing so. But this is incoherent if we are flexible enough in allowing the use of others as means, and if we do not require exclusivity, so that one might bring it about that *p by* bringing it about that someone else does so more directly.

[14] Objection: "I do not think one gets a commitment to acting well from a commitment to acting. Often, I do not care how well I am doing what I am doing. It is just

Even if performances do not have the automatically induced aims just suggested, moreover, we still retain an account of why knowledge is better than merely true belief, since apt performances, in general, are as such better than those that attain success only by luck. Beliefs are a special case of that general truth. This account still requires our view of knowledge as apt belief, belief that manifests the relevant competence of the believer in reaching its aim of truth.

V. THE VALUE PROBLEM REDUX

However, it is not yet quite clear what sort of "value" we are attributing to knowledge when we consider it always "better" as such than would be the corresponding true belief. What is the respect of comparison, what is the dimension along which the value that knowledge has or would have is *always, necessarily* above the value of its corresponding true belief?

That question has been much discussed in recent epistemology. But what exactly *is* the question? In what way might knowledge be valuable? What does it mean to say that it is valuable? What are we saying when we claim that it must always be more valuable? An account of the meanings that the relevant phraseology has in English is of course welcome in its own way. But our puzzlement may admit a more direct cure. What we need is some take on our main question that will be clear enough and that will make it plausible enough that knowledge *is* "better" than would be the corresponding merely true belief. How should we take the question so that we comfortably can give it the answer that it seems so obviously to deserve? (Of course, this will leave it open that *other* ways of understanding the question may yield the same benefit. In our present straits, in any case, finding even one way would be welcome.)

One way at least in which knowledge is valuable is the way social interaction is valuable, or friendship, or nourishment. Here I mean to comment on the logic of such attributions of value. All of those things said to be valuable have some important role in a flourishing human life. Presumably, that is what makes them valuable. But this does not require that *every* instance of them be valuable, as an end or even as a means. Compare the sense in which good, apt shots are valuable for a hunting, warlike tribe, or why they have some important, valuable role

---

not important enough for me to invest myself in an activity in this way." Reply: "Competence," however, need not imply a high degree of competence; it can be minimal. And if a doing by an agent does not manifest even minimal competence, then it is unclear that it counts as an action attributable to that agent.

in the life of that tribe. That is what makes it true to say that they are valuable, and that good arrows, good bows, and good marksmanship are there valuable. Quite compatibly, however, many good, apt shots might have no value whatsoever, not even pro tanto or prima facie. None such can therefore contain more value than would be found in a corresponding shot that was not apt, nor even any good at all.

The value that knowledge in general has for the flourishing human life hence does not yet explain a way in which knowledge is *always* better than the corresponding merely true belief. Nor is knowledge necessarily better as a means to our relevant objectives. This is the point made in the *Meno*. Some true means-ends beliefs will help us attain our ends just as well as knowledge.[15]

Yet somehow, in some sense, knowledge would seem always to be preferable as such to the corresponding merely true belief. What is the relevant dimension?

If we think of knowledge as a kind of performance, in a broad sense, that may help us understand the apparent claim that knowledge enjoys such superiority.

Consider the following two theses of performance normativity:

> Success is better than failure.
>
> Success through competence is better than success by luck.

These are *im*plausible if interpreted as theses of absolute, objective value. And they gain little if interpreted as theses of instrumental value. It is implausible that the success of any endeavor is thereby always intrinsically valuable, independently of its specific content. Nor is it any more plausible that it must always be extrinsically valuable. Nor is it much more plausible that it always has at least pro tanto or prima facie intrinsic value. That success in *any* endeavor whatsoever would always, necessarily have some objective *intrinsic* value at least pro tanto or prima facie seems implausible. Consider the success of a wholly evil act of torture. Yes, there is logical space for the view that the evil aspects of the act only outweigh the objective prima facie value that nevertheless still resides in its success. But there is little to zero plausibility space, as far as I can tell.[16] I, at least, can discern no objective, intrinsic value in its success as such, not even prima facie.

[15] If it is replied that knowledge implies a confidence resistant to fruitless inquiry, then consider relevantly stubborn true belief. Won't such belief, if pertinently stubborn, be equally resistant without having to amount to knowledge?

[16] At a deeper level this does seem to turn, however, on a debatable issue in axiology: Is the satisfaction of *actual* preference a source of value, at least prima facie or pro tanto, regardless of how evil its content may be?

So we try another approach. Compare this: Anyone endeavoring to attain an objective would always prefer to *attain* his objective than not to do so; moreover, this would always be a proper preference, at least prima facie, though its propriety could of course be overridden. *Reaching* an objective must be distinguished, moreover, from *attaining* it, which requires that you reach it not just by luck. A rational, unakratic agent endeavoring to attain an objective already prefers attaining it, all things considered. Merely wishing for a certain outcome is weaker than endeavoring, or aiming for that outcome. Inherent in such aiming is endeavoring to bring about one's aim. Hence it is a requirement of basic coherence that if our agent compares the satisfaction of his preference with its frustration, he must rationally prefer the former.

Compare an agent who believes that *p* and considers whether his belief is true. Simple coherence would require that one consider one's beliefs true. Similarly, simple coherence requires that one prefer one's overall preferences satisfied.[17] This, I suggest, is why it seems so plausible that "success is better than failure." In making that judgment with such insouciant generality, one is adopting the point of view of the hypothetical agent. What then might the judgment mean in the mouth of the agent? As an agent, I am suggesting, one prefers the satisfaction of one's overall preferences, and this is a rationally proper preference to have, given how incoherent it would be to prefer the opposite or even to suspend preference.[18]

[17] Objection: "True, by believing I commit to regarding my belief as true. However, we would not say that this makes this latter belief correct. The distance between believing *p* and believing one's belief that *p* to be true is so small that our evaluation of the latter is always (exclusively) guided by our evaluation of the former. Similarly, in the case of preference we would answer the question whether it is good for the agent to succeed by evaluating his aim. The fact that consistency demands a preference for having one's preference fulfilled in virtue of having a preference in the first place is a similarly slim basis of evaluation. I want to count the grains of sand. Thus, I want my preference for counting the grains of sand to be fulfilled. This does not make my preference 'correct', or 'okay'."

Reply: Yes, I agree. But if we interpret the claim that knowledge is always, necessarily better along these lines, then it seems false. So this interpretation does not yield the truth that we feel can be contained in the claim that knowledge *is* always, necessarily better. The alternative suggestion is that in making the obviously true claim we adopt the position of the agent, and take note of the fact that *in respect of rational coherence*, he always, necessarily does well in endorsing on the second order the preference that he already has on the first order. This does not mean that he does well, *all things considered*, in proceeding that way. This latter is not always, necessarily the case; in fact, much too often it is false.

[18] The foregoing discussion illustrates one main problem with critiques of the use of intuition in philosophy. Apparent disagreement in intuitions too often reflects disagreement on the questions and not on the answers.

VI. CONCLUSION

We concluded in part I that our intuitions on epistemic propriety or worth can be accommodated only by a conception of belief as disposition to affirm and not by a confidence-threshold conception. Part II argued that knowledge is required for apt action, and of course has value in that way. And part III laid out how the disposition-to-affirm conception underwrites the equivalence of our value-of-knowledge intuition with the knowledge norm of assertion.

But we still faced the further value problem taken up in part IV: namely, that of understanding how knowledge can be said *always* to be better than would be the corresponding merely true belief. In this part, we considered ways to understand how such a saying might conceivably be plausible. What could we have in mind? In answer to this question, we settled on the following suggestion. In making such a general claim, we take the point of view of the believer and see that he would always *correctly* prefer his knowing, in at least one important respect, insofar as to know would be to attain aptness, which simple coherence requires one to prefer.

That then is the suggested explanation of how we speak with plausible truth in saying that knowledge is always, necessarily better than would be the corresponding merely true belief. We are saying that it would always, necessarily be proper for one to prefer one's knowing to one's merely believing correctly. This is just a special case of the fact that, for *any* endeavor that one might undertake, it is always, necessarily proper for one to prefer that one succeed in that endeavor, and indeed succeed aptly, not just by luck. That is always, necessarily proper in at least one important respect. And our relevant beliefs, endeavors after truth, are just a special case. One would always properly prefer to attain that which one endeavors to attain, and to attain it aptly, not just by luck. One would properly have that preference at least in the respect that it is the preference required for rational coherence.

Is there any more objective sense in which knowledge might plausibly be more valuable than merely true belief? Yes, surely knowledge is valuable because knowledge *of certain matters* adds so importantly to the flourishing of one's life individually, and of life in community. Mere true belief on those important matters falls short. This, however, is not to say that *every* instance of knowledge adds in those important ways, that such knowledge is *always* necessarily better than merely true belief. Nor does it even seem true that *every* instance of knowledge on such important matters adds to the flourishing of that knower or community. All that is required for it to be true that knowledge is a valuable commodity, more so than corresponding merely true belief, is that knowledge of certain important matters should normally make an important positive contribution as part of a life that flourishes individually, or as part of the flourishing of a community, above any contribution that would be made by corresponding merely true belief.[19]

ERNEST SOSA

Rutgers University

[19] Obviously, knowledge could even be valuable in that it is important to have *some amount of knowledge*, enough knowledge, of, say, one's loved ones, so as to be able to know *them* well, even if *no particular bit of knowledge* is essential for this outcome. Similarly, some amount of intimate interaction with one's intimates is a great good in any flourishing life, and such intimate interaction would plausibly need to be knowledgeable. But none of this would account for the respect in which *any* bit of knowledge would be better than would have been the corresponding merely true belief. The value of *some amount* does not necessarily extend to *each bit*.

THE NATURE AND PURPOSE OF NUMBERS*

Numbers are abstract entities introduced for the purpose of counting. The present paper is dedicated to the explication of this claim, and in particular it addresses the questions of what makes these entities "abstract," in what sense they are "introduced," and what we mean by "counting." Along the way, we investigate the logical status of arithmetic, the function of abstraction principles, and the respective merits of various strategies for reducing arithmetical notions to those of a theory that is viewed as more fundamental. The main emphasis is on the conceptual, foundational, and philosophical issues, with the technical details fully developed elsewhere.[1]

Two main conceptual threads are at the basis of the present approach: a *deflationary* conception of abstraction and a *nonreductionist* version of logicism. Each is implemented through a specific device, that is, respectively, an extra-logical operator representing numerical abstraction and a nonstandard (but still first-order) cardinality quantifier. The result is an account of arithmetic characterizing numbers as obtained by abstraction from the equinumerosity relation and emphasizing their *cardinal* properties (as used in answering "how many?" questions) over their structural ones.

Abstract entities are obtained through the application of an abstraction operator to what Frege would have called a *concept*, and which we will also refer to as a (possibly complex) *predicate*—as long as we take care to distinguish predicates from predicate *expressions*. However, not every mapping of the concepts into the objects represents an instance of abstraction. Abstraction operators are distinguished from other such assignments in that they are assumed to map concepts into objects while respecting a given equivalence relation. Frege, for instance, postulated an abstraction operator assigning objects of a particular kind—which he called *extensions*—to concepts in such a way that equi-extensional concepts (that is, concepts under which the same objects fall) are assigned the same object. This particular postulation, as embodied in an abstraction principle known as Frege's *Basic Law V*, turned out to be inconsistent. Nonetheless, many other abstraction principles are indeed consistent, and among them, most

* I am grateful to Elaine Landry, Albert Visser, Ed Zalta, and an anonymous referee for helpful comments, suggestions, and criticism.

[1] See G. A. Antonelli, "Numerical Abstraction via the Frege Quantifier," *Notre Dame Journal of Formal Logic*, LI, 2 (2010): 161–79.

0022-362X/10/0704/191–212

notably for our purposes, is a principle of *numerical abstraction*, also known as *Hume's Principle*. HP postulates an operator NUM assigning objects to concepts in such a way that concepts P and Q are assigned the same object precisely when P and Q are "equinumerous," that is, when just as many objects fall under P as fall under Q. The object NUM(P) assigned to P can then be regarded as "the number of P."

It is natural to think of the abstract objects delivered by such operators as having a somewhat mysterious nature, with such properties as nonspatiotemporal existence and causal inefficacy, for instance. But that would be a mistake, for abstraction principles do not force upon us any particular view of the entities they introduce. In fact, all that the adoption of a particular abstraction principle commits us to is the existence of *some* mapping of the concepts into the objects satisfying certain further conditions related to a given equivalence relation. The abstraction principles themselves do not tell us anything about the "ultimate nature" of these objects—and that is how it should be. The role of these abstraction principles is not to single out a special class of objects (the abstract entities living perhaps in a separate sphere of reality), but rather to make sure that the first-order domain of objects is large enough to accommodate representatives for the equivalence classes. Now, one might entertain specific worries about the ontologically inflationary nature of abstraction—the fact that, for instance, HP forces us to accept the existence of an infinite number of objects—but such worries are quite distinct from those concerning the causal inefficacy or the nonspatiotemporal existence of numbers.

According to the view of abstraction proposed here, there is nothing special about abstract entities, which can be drawn from whatever first-order domain we take our ordinary quantifiers 'there is' and 'for all' to range over. In this sense, abstract entities can be taken just to be ordinary objects recruited to serve as proxies for the equivalence classes of concepts generated by the given equivalence relation. Abstraction principles give a lower bound on the cardinality of the domain of objects, relative to the size of the class of all concepts, taken *modulo* a given equivalence relation. We characterize this view of abstraction as *deflationary* in that the main role it ascribes to abstraction is to provide such a lower bound, while denying the objects delivered by abstraction any special status.

Abstraction principles are represented linguistically by the explicit introduction of term-forming operators such as NUM(P) or, in the general case, $\Phi(P)$. For any possibly complex predicate expression P, the term $\Phi(P)$ will denote an object in the first-order domain. Such an object is introduced "by abstraction" as long as the assignment Φ of objects to predicates satisfies the constraints associated with the corresponding equivalence relation. This of course does not preclude the possibility that the same object might also be denoted by *other* terms—a possibility that, as we will see, points toward a possible dissolution of the so-called "Caesar problem."

The deflationary account of abstraction goes hand-in-hand with the view that abstraction principles are properly regarded as extra-logical principles that, as such, do not enjoy a logically or epistemologically privileged status. On the contrary, on many accounts inspired by some form or other of logicism, the logical character of arithmetical notions is made to depend on some kind of reduction of arithmetical truths to Hume's Principle, which in turn is claimed to be constitutive of the notion of number and therefore somewhat close to an analytical principle.

These logicist or neo-logicist views are based, however, on an equivocation. When logicism is properly understood in *nonreductionist* fashion, it is the notion of *cardinality*, rather than that of *number*, that appears logically privileged. Numbers, as we have seen, are objects, and matters regarding the existence of objects fall outside the purview of logic proper. On the other hand, from a logicist point of view, cardinality *can* be viewed as logically innocent. When taken at face value, this broad construal of logicism opens up the possibility of including cardinality as one of the basic building blocks of a language suitable for the representation of arithmetic. The notion of cardinality that turns out to be at issue here, as we will see, is a *comparative* notion, specifying a relation between concepts P and Q that holds if and only if there are no more objects falling under P than there are falling under Q. This notion of comparative cardinality is linguistically represented through the introduction of the Frege quantifier F, binding two formulas ϕ and ψ, and expressing the fact that there are at least as many objects satisfying ψ as there are satisfying ϕ.

Numerical abstraction and the Frege quantifier are the two main devices that will be employed to give a first-order representation of arithmetic emphasizing the cardinal properties of numbers.

It is worth noting here, before we get to some of the details of the project, that there seems no obvious way to extend the present treatment to ordinal notions, except trivially in finite domains, where ordinal and cardinal numbers coincide. Certainly, it would seem that no such treatment could be developed using the Frege quantifier and the abstraction operator, which are aimed squarely at cardinal notions. But one might think that a similar treatment of ordinal notions could be developed by introducing *ordinal abstraction*. As we will see, however, such a principle gives rise to paradox. In this respect, ordinal notions, while usually regarded as on a par with their cardinal counterparts (or perhaps even as primary, as in set theory, where cardinals are defined as initial ordinals), instead appear to be intrinsically more complex than the latter, and quite possibly beyond the reach of a first-order treatment.

I. NUMBERS AS *ABSTRACTA*

Arithmetic is the theory of the natural numbers 0, 1, 2, ... with which we are all acquainted. Because of the basic character of the natural numbers as the foundation upon which mathematics and science rest, ever since Frege and Dedekind philosophers have been concerned with the proper formalization of arithmetic.[2] This line of investigation has lead to a variety of approaches, including the Dedekind-Peano axioms that are nowadays standard, several set-theoretic reductions, and finally a renewed interest in the Fregean project as championed by the neo-logicist school of Hale and Wright.[3]

One possibility is to regard numbers as primitive objects that need no reduction to others. A proponent of this view thus would fully embrace the axioms of standard first-order arithmetic as extra-logical characterizations of the fundamental properties of numbers. These axioms, first formulated by Dedekind[4] and Peano,[5] identify the basic properties of the successor operation on the natural numbers (as well as, possibly, the properties of addition and multiplication), and postulate an induction schema expressing that any properties of natural numbers that hold of zero and are preserved by the successor operation, hold of all natural numbers. Although Peano Arithmetic (PA), as the theory has come to be known, is sometimes supplemented by a second-order induction principle, it is standardly expressed at the first-order level. The insistence on a first-order axiomatization is motivated by the desire to preserve certain properties of first-order logic, such as axiomatizability, compactness, Löwenheim-Skolem, and so on, which fail at the second-order level.[6] Let us refer to this view as the axiomatic approach.

[2] Gottlob Frege, *Begriffsschrift, eine der arithmetische nachgebildete Formel-sprache des reines Denkens* (Halle: Nebert, 1879); English translation in J. van Heijenoort, ed., *From Frege to Gödel: A Source Book in Mathematical Logic* (Cambridge: Harvard, 1967). See also Frege, *Die Grundlagen der Arithmetik, eine logisch-mathematische Untersuchung über den Begriff der Zahl* (Breslau: Koebner, 1884); J. L. Austin, trans., *The Foundations of Arithmetic: A Logico-Mathematical Enquiry into the Concept of Number* (New York: Oxford, 1950).

[3] Robert Hale and Crispin Wright, *The Reason's Proper Study: Essays toward a Neo-Fregean Philosophy of Mathematics* (New York: Oxford, 2001).

[4] Richard Dedekind, *Was sind und was sollen die Zahlen?* (Brunswick: Vieweg, 1888).

[5] Giuseppe Peano, *Arithmetices Principia, nova methodo exposita* (Turin: Bocca, 1889). English translation in van Heijenoort, ed., *From Frege to Gödel.*

[6] For an accessible, clear, and rigorous treatment of second-order logic see Herbert Enderton, "Second-order and Higher-order Logic," *Stanford Encyclopedia of Philosophy,* ed. Edward Zalta (2008). URL: http://plato.stanford.edu/entries/logic-higher-order/.

A second approach was championed both by Frege (in his *Grundlagen* and, later, the *Grundgesetze*[7]) and Whitehead and Russell,[8] each one of whom provided an account of the natural numbers as intimately related to classes of equinumerous concepts, that is, as equivalence classes under the "having the same cardinality" relation. In Whitehead's and Russell's (unramified) theory of types, numbers are "concepts of concepts" (that is, concepts of propositional functions) such that between any two of such functions there is a relation that is both one-one and onto. As a consequence of the rigid type-theoretic discipline of the theory, numbers are reduplicated at each type higher than 2. This result makes it impossible to compare cardinalities across types and hardly squares with our intuition that there is, in fact, just one class of natural numbers. Frege's framework avoids such undesirable consequences by characteristic recourse to a type-lowering device, that is, concept extensions.[9] Concept extensions derive from the application of a particular abstraction principle, the crucial function of which is to assign first-order objects to concepts in such a way as to respect the equi-extensionality relation among the latter. Frege's strategy ultimately failed because it was driven by the desire for unattainable generality—the idea that the assignment of extensions to concepts needs to be universal—whereas, as we know from Cantor's theorem, there are many more concepts than objects, even when concepts are taken *modulo* equi-extensionality.

Finally, ever since the acceptance, by some, of Zermelo-Fränkel set theory as the privileged framework for mathematics, set-theoretic reductions of arithmetic have become standard. Such reductions proceed by identifying particular representatives for Frege's and Russell's equinumerosity classes. These representatives are given in the form of a linearly ordered sequence of sets, having a first element, and with the additional property that any element has only finitely many predecessors in the sequence. Sometimes, a particular n-membered set is selected as representative for the number n (as when the sequence comprises the von Neumann finite ordinals). However, this need not always be the case, as with the Zermelo numbers (standardly, but somewhat inappropriately, referred to as "Zermelo numerals"), in which each element is the singleton containing its unique predecessor,

[7] Frege, *Grundgesetze der Arithmetik* (Jena: Hermann Pohle, 1903); Montgomery Furth, trans., *The Basic Laws of Arithmetic* (Berkeley: California UP, 1967).

[8] Alfred N. Whitehead and Bertrand Russell, *Principia Mathematica*, vol. 1, 2d ed. (Cambridge: University Press, 1925).

[9] A reconstruction of Frege's extensionalist—as opposed to logicist—program can be found in Antonelli and Robert May, "Frege's *Other* Program," *Notre Dame Journal of Formal Logic*, XLVI, 1 (2005): 1–17.

so that the representative chosen for the number *n* does not itself have *n* members. In fact, quite generally, representatives chosen for equivalence classes relative to some relation *R* need not themselves be in the field of *R* (although they often will): all that is required is that the assignment of representatives respect the equivalence relation.

Each of the above-mentioned approaches is wanting in some respect or other. Formalizing arithmetic as a first-order theory, PA ultimately leaves the nature of numbers unexplained. There is certainly something to be said for taking an approach that focuses on the intrinsic order-theoretical properties of the natural numbers, for instance, the fact that they form an ω-sequence, without being concerned with their ultimate nature. But this approach fails to take into account other, crucial and extrinsic, properties of the natural numbers: first and foremost their *cardinal* properties, as they are put to use in connection with the act of counting. Natural numbers find their primary domain of application in answering questions such as "How many?" It is therefore a desirable feature of any formal account of arithmetic that the cardinal properties of the natural numbers take center stage, at least on a par with the mathematical properties of ω-sequences. It is this emphasis on the applicability of arithmetic that is lacking in any account that privileges order-theoretical properties of ω-sequences over cardinal ones. In this respect, PA still needs to be supplemented by a separate account of how numbers are used in counting.

Set-theoretic reductions fare somewhat better than PA when it comes to their employment in the assignment of cardinal numbers, if anything because they are embedded into a richer theory allowing for many sorts of maps between sets of different kinds. The details here depend on the particular set-theoretic reduction being adopted, but set theory does provide the beginnings of an account of the cardinal properties of the natural numbers. But again, this account is not based on an explicit consideration of such properties. Moreover, set-theoretic reductions suffer from a problem first pointed out by Paul Benacerraf,[10] who argues that there is no privileged way to select one particular set-theoretic reduction over any other, and that in fact, in the end, "any ω-sequence will do." Given that all set-theoretic reductions (of which there are indeed infinitely many) provide for the same intrinsic order-theoretical properties of the natural numbers,

[10] Paul Benacerraf, "What Numbers Could not Be," *Philosophical Review*, LXXIV, 47 (January 1965): 47–73, reprinted in Benacerraf and Hilary Putnam, eds., *Philosophy of Mathematics: Selected Readings*, 2d ed. (New York: Cambridge, 1983), pp. 272–94; and "Recantation *or* Any old ω-sequence would do after all," *Philosophia Mathematica*, IV, 3 (1996): 184–89.

how is one to assess the relative merits of and decide between, say, Zermelo numerals and von Neumann cardinals? Benacerraf 's answer is that there is no principled reason to choose between them, and that since the natural numbers cannot be both Zermelo numerals and von Neumann ordinals, numbers cannot be sets at all.

This leaves us with a last option, the Frege-Russell account of the natural numbers. The great advantage of this approach is that the intrinsic mathematical properties of the natural numbers are derived from their cardinal properties, rather than the other way around. Whereas for both the axiomatic approach and set-theoretic reductions cardinal properties require a separate account, according to the Frege-Russell conception such properties are central to the account of the natural numbers. The essential lines of such an approach, therefore, appear to be intuitively well motivated and mathematically elegant. Unfortunately, the mathematical implementations are rife with problems: Frege's own attempt in *Grundgesetze* was notoriously inconsistent, and Whitehead's and Russell's imposition of a type discipline, while blocking the inconsistency afflicting Frege's theory, led to reduplications and restrictions that hardly do justice to actual mathematical practice.

Lately, Hale and Wright (*op. cit.*) have championed a somewhat different, "neo-logicist" approach, which addresses directly the idea that numbers are related to equivalence classes under the equinumerosity relation, dispensing with the whole apparatus of concept extensions.[11] They introduce a theory of numbers based on what we have been referring to as the Num operator and postulating that such an operator is to satisfy Hume's Principle. Since, as it was already known to Frege, the axioms of PA can be derived, within second-order logic, from Hume's Principle, the resulting neo-logicist system is thus adequate for the representation of second-order arithmetic (and in fact equiconsistent with it).

There are two main issues with such an approach. The first is both philosophical and conceptual: Hale and Wright rely on the logical character of Hume's Principle in order to characterize their project as continuous with Frege's original logicist views. However, the extent to which Hume's Principle enjoys a logically or even epistemologically privileged status is debatable, for instance in light of the fact that there are models of arithmetic where Hume's Principle fails.[12]

[11] Antonelli and May (*op. cit.*) go in the opposite direction, developing an explicitly extensionalist program dispensing with the logical character of arithmetical principles.

[12] A counterexample to the right-to-left direction of Hume's Principle can be found in George Boolos, "On the proof of Frege's Theorem," in Adam Morton and Stephen

Even discounting these worries, a main technical obstacle remains, namely the fact that the neo-logicist program is carried out wholly within the framework of second-order logic, characterized by the already-mentioned failure of those meta-theoretic properties that make first-order logic so attractive. Nonetheless, the Frege-Russell approach appears to be conceptually superior in its characterization of numbers as *abstracta* of the equinumerosity relation, in the way it derives basic mathematical properties from cardinal ones (and the concomitant emphasis on the applicability of arithmetic), and in its intuitive motivation.

II. DEFLATIONARY ABSTRACTION

In their most general form, abstraction principles such as HP govern the assignment of objects to concepts according to a given equivalence relation:[13]

$$f(P) = f(Q) \Leftrightarrow R_f(P, Q)$$

A principle of this form asserts that the object f assigns to the concept P is the same as the object it assigns to the concept Q if and only if P and Q are appropriately related to each other by the equivalence relation R_f.

Because abstraction in some sense "lives" in conceptual space—as evidenced by the connection of abstraction to concept formation, for example, in children or in science (whereby a concept is "abstracted" from a variety of instances)—abstraction principles often have been thought to enjoy a particularly privileged status. But clearly not all such principles are acceptable. For instance, one could not have a principle of the form $f(P) = f(Q) \Leftrightarrow P = Q$, for then we would have at least as many objects as there are concepts over a given domain, contradicting Cantor's theorem (independently of whether the identity $P = Q$ is taken extensionally or intensionally). There are two possible strategies, a combination of which can be used to obviate such a situation: the equivalence relation appearing on the right of the equivalence can be made *coarser*, allowing possibly a great many distinct concepts to be assigned the same object; or the function f on the left can be made to apply only to a subset of all the concepts (the definable ones, for instance).

Stitch, eds., *Benacerraf and his Critics* (Malden, MA: Blackwell, 1996), pp. 143–59. A counterexample to the left-to-right direction—usually considered the less questionable one—can be found in Antonelli and May, *op. cit.*

[13] See Gideon Rosen, "Abstract Objects," *Stanford Encyclopedia of Philosophy* (2006). URL: http://plato.stanford.edu/entries/abstract-objects/.

Hume's Principle implements the first of these strategies, by taking the equivalence of the right to be *equinumerosity*. As mentioned, neo-logicists such as Hale and Wright claim for HP a status not unlike that of a logical or analytical truth, and consider HP somehow "constitutive" of the notion of number. However, several objections have been raised against such a privileged status, beginning with the so-called "Bad Company" objection, that is, the fact that there are principles very much like HP that turn out to be inconsistent. One such example, of course, is Frege's own Basic Law V. Another example replaces equinumerosity with order-isomorphism: according to such a principle, given relational predicates $\varphi(x, y)$ and $\psi(x, y)$, one says that the *type* $\tau(\varphi)$ of φ equals the type $\tau(\psi)$ of ψ if and only if $\varphi(x, y)$ and $\psi(x, y)$ are order-isomorphic. Attractive as this treatment of ordinal notions might appear, it is inconsistent, for it gives us the Burali-Forti paradox.[14]

Against this, neo-logicists can (and did) rebut that consistency itself ought to be taken as the hallmark of acceptability for abstraction principles. Both R. Heck[15] and A. Weir[16] have criticized this line of argument, however, pointing out that there are individually consistent but pairwise incompatible abstraction principles: if both principles in such a pair are acceptable, hence at least to an extent analytic, then both need to be regarded as true, which is of course impossible (this is Weir's so-called "Embarrassment of Riches" objection). There is a vast literature attempting to find a general and well-motivated demarcation between acceptable abstraction principles and unacceptable ones.[17] But in the end there are reasons to believe that any such attempt to find a completely satisfactory demarcation might turn out to be futile.

The issue of the privileged status of abstraction principles also can be approached from a different angle. There is a long tradition, going back to the work of Alfred Tarski, according to which logical notions are those that are *invariant* under permutations of the domain of objects, ostensibly because logical notions are completely general and do not have any specific subject matter.[18] Tarski first introduced

[14] See John Burgess, *Fixing Frege* (Princeton: University Press, 2005), pp. 164–70, for an excellent treatment of these issues.

[15] Richard G. Heck, Jr., "On The Consistency of Second-Order Contextual Definitions," *Noûs*, XXVI, 4 (December 1992): 491–94.

[16] Alan Weir, "Neo-Fregeanism: An Embarrassment of Riches," *Notre Dame Journal of Formal Logic*, XLIV, 1 (2003): 13–48.

[17] See Øystein Linnebo, ed., *The Bad Company Problem*, special issue of *Synthese*, CLXX, 3 (2009).

[18] This proposal has since come to be known as the Tarski-Sher thesis after the proposal was endorsed and further developed by Gila Sher in *The Bounds of Logic* (Cambridge: MIT, 1991).

this idea in its full generality in a posthumously published 1966 lecture,[19] explicitly inspired by Klein's "Erlangen" program.

Let us call a predicate P *logically invariant* if and only if for any permutation π and object a, $\pi(a)$ falls under P if and only if a does. This idea can be generalized to notions of arbitrary type, including connectives, quantifiers, and other higher-order objects. While Tarski's proposal is nowadays considered too liberal (in that it countenances as logical notions that appear to be properly mathematical, rather than logical, in character), there is widespread consensus that permutation invariance provides at least a necessary (although likely not sufficient) condition for logicality of predicates and quantifiers.[20] Somewhat surprisingly, the invariance criterion was not applied to the question of the status of abstraction principles until relatively recently.

Kit Fine first considered criteria of logical invariance for abstraction principles in their full generality.[21] Among the several possible ways in which such criteria can be formulated, we follow the reconstruction given by John Burgess,[22] and call an abstraction principle of the form

$$f(X) = f(Y) \Leftrightarrow R_f(X, Y)$$

invariant if and only if for any permutation π, it holds that $R_f(X, \pi[X])$, where $\pi[X]$ is the point-wise image of X under π. Clearly, this implies the weaker criterion according to which $R_f(X, Y)$ holds precisely when $R_f(\pi[X], \pi[Y])$ also holds.

Since permutations preserve the cardinality of a set, it follows immediately that HP is logically invariant, in the sense that if R is the equinumerosity relation between concepts (viewed as sets), then for any such concept X, we have $R(X, \pi[X])$. But we should not make too much of this. First, it is not at all clear that there is a privileged way to express invariance for abstraction principles.[23] Even considering just this version of invariance, however, there is another principle, very close to HP, that is invariant in this sense, but inconsistent: this is the principle mentioned in the "bad company" objections to HP assigning abstracts to binary relations R and S in such a way that R and S are assigned the same object precisely when there is an order isomorphism between them. The principle is invariant because a

[19] Alfred Tarski, "What are Logical Notions?" ed. John Corcoran, *History and Philosophy of Logic*, VII (1986): 143–54.

[20] See Denis Bonnay, "Logicality and Invariance," *The Bulletin of Symbolic Logic*, XIV, 1 (2008): 29–68.

[21] Kit Fine, *The Limits of Abstraction* (New York: Oxford, 2002).

[22] Burgess, *op. cit.*

[23] See Antonelli, "Notions of Invariance for Abstraction Principles," *Philosophia Mathematica*, forthcoming.

relation R and its image S under a permutation π are order-isomorphic (with π itself providing the isomorphism), but as already mentioned it gives rise to the Burali-Forti paradox.

The approach of this paper aims to sidestep the issue of the privileged status of abstraction principles completely. Abstraction principles, when properly understood, just provide an assignment of representatives to the equivalence classes induced by equivalence relations. Nothing more is said about these representatives, other than that the assignment must respect the equivalence relation. Accordingly, these principles are best viewed as *extra-logical* devices, the main function of which is to provide "inflationary thrust" on the first-order domain D of objects by imposing a lower bound on the cardinality of D, relative to the size of the space of concepts over D.[24]

Some abstraction principles are good and some bad, and among the bad ones we should certainly count those that are inconsistent. Good abstraction principles can be put to use in a variety of philosophical contexts, and they are capable of accomplishing a variety of tasks. But by classifying them as "good" no principled good/bad demarcation is implied—only the statement that they have turned out to be useful in some context or other.

We characterize this view of abstraction as *deflationary*, in that it denies abstract objects any privileged status (the fact that the deflationary account emphasizes the *inflationary* role of abstraction principles should not—it is hoped—lead to confusion). Any worries about the special ontological status of abstract objects, or the special logical status of abstraction principles, to the extent that they have any cogency at all, are no longer foisted upon us by a consideration of abstraction. For, from a *logical* point of view, we need not assume anything logically (or epistemologically) privileged about such principles; and from an *ontological* point of view, we need not assume that abstract objects make up a separate, privileged ontological realm. *Anything at all*—even ordinary objects—can play the role of these *abstracta*, as long as the choice respects the equivalence relation.

One of the recurring worries connected with the introduction of natural numbers via Hume's Principle is the so-called "Caesar Problem," that is, the fact that HP gives us enough information to settle the truth-value of identities in which both terms are abstracts, but says nothing about identities involving an abstract and a term of a different kind. The worry, then, is that no account of the natural

[24] Fine, and others following him, use "inflationary" in connection with abstraction principles in a different sense.

numbers can be regarded as complete unless their identity conditions settle all such questions.

Such worries are misplaced. The answer to the question, "What prevents the number of the planets from being equal to Julius Caesar?" is: *nothing*. The number of the planets will indeed be equal to Julius Caesar in some models, and distinct from it in some other models. Nothing much is to be made of this, for the corresponding abstraction principle, HP, is silent about it. To worry about this is not to understand the proper nature of abstraction as imposing a lower bound on the size of the domain of objects, rather than opening up a metaphysically separate domain of "abstract" objects.[25] The existence of such a separate realm has nothing to do with abstraction, for abstraction is perfectly compatible with there being only one domain of discourse, populated by objects to which we should have no qualms appealing for whatever philosophical, logical, or mathematical purpose we might be pursuing. So, in this sense, the present account is deflationary as regards the Caesar problem as well.

It is worth noting that in his classic paper on the nature of numbers (and its sequel) Benacerraf also reached a generally deflationary point of view. His conclusion that "any ω-sequence will do after all" deflates metaphysical worries about the ultimate nature of numbers. But Benacerraf's account moves in a different direction from the present one, focusing on the order-theoretical rather than the cardinal properties of the natural numbers. Being an ω-sequence is an order-theoretical property *par excellence*, and such properties are best viewed as supervenient upon cardinal ones—or so we submit. Benacerraf's account, then, is *incomplete*, but not because it fails to provide a characterization of the ontological status of numbers; he himself showed that no such strategy could succeed. Rather, the account is incomplete because it does not address the role of cardinal properties in arithmetical applications such as counting.

Benacerraf's views have long been regarded as promoting a version of *structuralism* as regards arithmetic. His claim that any ω-sequence can play the role of the natural numbers has been interpreted as implying that there is nothing more to the natural numbers than their order-theoretical properties. Be that as it may, Benacerraf certainly

[25] It is then perfectly possible for *abstracta* to have "additional nature" beyond the properties they inherit in virtue of HP. This issue recently was taken up by Hale and Wright in "Abstraction and Additional Nature" (*Philosophia Mathematica*, XVI, 2 (2008): 182–208) in response to Michael Potter and Peter Sullivan's "What is Wrong with Abstraction?" (*Philosophia Mathematica*, XIII, 2 (2005): 187–93). The line taken here is that abstracta are allowed to have such "additional nature," a fact that presents no obstacle to their employment in mathematics and science.

did argue against the claim that there is a privileged way to select objects to be arranged in an ω-sequence. A similar argument can then be put forward after we shift the focus from the order-theoretical properties to the cardinal ones. The lesson to be learned from Benacerraf's argument is this: once we have an account of natural numbers in terms of cardinal properties, it does not make any difference which objects are chosen as representatives of the equinumerosity classes, as long as we have enough of them to satisfy the inflationary thrust of the corresponding abstraction principle.

III. NONREDUCTIONIST LOGICISM

Traditional logicism and neo-logicism rely on the special status of abstraction principles such as HP to establish the logical character of arithmetical notions. But once we adopt the deflationary point of view and regard these principles as extra-logical and not epistemically privileged, what is then left of logicism? What are the prospects of the ambitious program initiated by Frege and revived by the neo-logicist school?

Frege's original program aims to combine two largely incompatible views: *logicism*, construed as the view that arithmetic is interpretable into (higher-order) logic; and *extensionalism*, construed as a theory of concept extensions *qua* abstract objects. Dummett[26] referred to the latter as Frege's "platonism"—the view that there are logical objects in the form of concept extensions—and pointed out that the view is not only independent from, but in fact in direct tension with Frege's logicism, a fact that underpins the contradiction uncovered by Russell. Indeed, on the "natural view" of logic, there are *no* logical objects: all that is needed or required for the completion of the logicist program is an interpretation of all mathematical statements (or at least the arithmetical ones) into a logical language.

In recent versions of neo-logicism, Frege's extensionalism has been replaced by a theory of numbers *qua* logical objects delivered by HP. It is possible, however, perhaps for the sake of conceptual purity, to pursue one part of Frege's original program independently of the other. For instance, one approach[27] is to pursue extensionalism by providing a theory in which concept extensions are explicitly governed by extra-logical principles, and arithmetic is recovered as a second-order theory identifying numbers with particular concepts. But can logicism also be pursued for its own sake without relying on a theory of numbers as logical objects?

[26] Michael Dummett, *Frege: Philosophy of Mathematics* (Cambridge: Harvard, 1991), p. 301.
[27] Antonelli and May, *op. cit.*

Arithmetical logicism[28] generally is characterized as the view that arithmetic is, in a substantial sense, *logic*. This view is usually taken to comprise the two distinct claims that arithmetical notions are definable in terms of purely logical ones and that, under this interpretation, arithmetical theorems can be proved from purely logical principles. Ever since Frege, the view that arithmetic is logic is most often articulated in a *reductionist* fashion by identifying some principle, claimed to enjoy some logically or epistemologically privileged status, to which (translations of) arithmetical truths turn out to be proof-theoretically reducible. The need for such a principle is clear. Even when arithmetical notions have been appropriately translated in a purely logical language, one cannot expect translations of arithmetical *theorems* to be provable using only the most general axioms and rules characterizing reasoning in terms of connectives and quantifiers. For one thing, arithmetic implies the existence of a great many objects, and pure logic alone cannot establish existence claims. Hence the need is felt for some intermediate principle carrying enough inflationary thrust to allow the derivation of such arithmetical truths while remaining purely logical in character. Frege identified such a principle in Basic Law V, the completely general nature of which does indeed lend some plausibility to the claim that it is a proper part of logic. Unfortunately, Basic Law V is *too inflationary* and therefore unsatisfiable. The option pursued by the neo-logicists, instead, is Hume's Principle. While HP is inflationary, it is also consistent. It is arguable, however, whether HP enjoys the full generality that made claiming logical status for Frege's Basic Law V plausible. And, as we have seen, there are reasons to question the logical character of HP.

The reductionist implementation of logicism thus seems to fall short of the desired goal. But why go the reductionist route in the first place? Reductionism did enjoy a certain currency in the philosophy of science, a role later codified in the work of Nagel.[29] He championed inter-theoretic reduction via "bridge principles," the role of which was not too dissimilar from that of Basic Law V or Hume's Principle in logicist or neo-logicist theories. However, even if reductionism could be defended in the case of empirical science, in the case of arithmetic such a strategy is not the only, neither the most general, nor even the most natural interpretation of logicism.

[28] By "arithmetical" logicism we mean logicism as applied to arithmetic. There is another, more ambitious version of logicism claiming that mathematics as a whole is interpretable as logic in the same sense. Here we restrict our attention to the more limited version, to which we refer as "logicism" *simpliciter*.

[29] Ernest Nagel, *The Structure of Science* (New York: Harcourt, Brace & World, 1961).

On the broadest interpretation of logicism, cardinality is *already* a logical notion, and it does not need a definition in terms of a special kind of logical objects to make it so. This is a point already made by Dummet,[30] although for Dummett it applies to the notion of cardinal *number*, rather than directly to the more basic notion of cardinality. The two, as mentioned, are indeed distinct. Whereas cardinal numbers are objects introduced by abstraction, cardinality expresses a property of concepts or, more generally (in the case of equinumerosity), a *relation* between concepts. And while the case for the logical character of cardinal numbers is philosophically suspect, those same objections hardly apply to the general notion of cardinality.

Once we recognize that cardinality is also available to the logicist as a genuine logical notion *per se*, independently of the status of any abstraction principles involved, we come to a broader and more natural construal of the logicist project, in which cardinality is employed as one of the main building blocks (or perhaps *the* fundamental block) in designing a formal framework adequate for the representation of arithmetical facts. As mentioned, the basic notion of cardinality involved is that of a relation F between concepts X and Y that holds whenever there are no more X's than Y's or, in set-theoretic terms, whenever there is an injection of the X's into the Y's. That this notion is more basic than that of equinumerosity can be seen from the fact that the latter is definable from the former (but not vice versa): X is equinumerous to Y if and only if both $\mathsf{F}(X, Y)$ and $\mathsf{F}(Y, X)$ hold.

A relation between concepts, as we will see, is a *quantifier*.[31] We will accordingly develop a formal framework for the formalization of arithmetic having the F quantifier as one of its basic building blocks. We refer to such a quantifier as the Frege quantifier, and use it to develop, in a nonreductionist fashion, an account of arithmetic in which the logicist standard is not carried by the abstraction principle but rather by the quantifier itself. Thus, the very idea of cardinality is seated firmly at the center of the resulting logical framework as a primitive logical notion.[32]

[30] Dummett, *op. cit.*, p. 224.

[31] The most comprehensive survey on generalized quantifiers can be found in Stanley Peters and Dag Westerståhl, *Quantifiers in Language and Logic* (New York: Oxford, 2006).

[32] The Frege quantifier is not, of course, alone among cardinality quantifiers. Most notably, quantifiers in the same vein were first introduced by K. Härtig, "Über einen Quantifikator mit zwei Wirkungsbereichen," in L. Kalmár, ed., *Colloquium on the Foundations of Mathematics, Mathematical Machines and their Applications* (Budapest: Akadémiai Kiadó, 1965), pp. 31–36; and Nicholas Rescher, "Plurality-Quantification and Quasi-Categorical Propositions," *Journal of Symbolic Logic*, XXVII (1962): 373–74. The Härtig quantifier $\mathsf{I}(A, B)$ holds if and only if A and B have the same cardinality, and the Rescher quantifier $\mathsf{R}(A, B)$ holds if and only if there are strictly more A's than B's.

The defining feature of the Frege quantifier is that it deals with cardinality notions *directly*, without appealing to any separately given mathematical machinery.

Compare this to the situation in set theory, where in order to express certain relationships between the cardinality of two given sets, one has to appeal to the existence of certain *other* objects in the domain of quantification—such objects are, in turn, sets of a certain kind, containing ordered pairs as members and satisfying certain further conditions. The same holds if one instead decides to express such cardinality notions at the second order, by asserting the existence of relations satisfying certain further constraints.

Eschewing both of these options, we instead take cardinality notions as linguistic primitives and explore the expressive power of the resulting linguistic framework. We propose to adopt a language—inspired by a nonreductionist approach to logicism—in which the Frege quantifier is the only primitive logical machinery besides predication and sentential connectives.[33]

IV. THE MODERN VIEW OF QUANTIFIERS

According to the modern view, a first-order quantifier over a domain D is a collection of, or more generally a relation among, subsets of D. This idea can be traced back to the work of Frege, and specifically §21 of his *Grundgesetze der Arithmetik*, where he asks us to consider forms of the "conceptual notation" corresponding to the modern formulas $\exists a(a^2 = 4)$ and $\exists a(a > 0)$: these forms can be obtained from the general form $\exists a\varphi(a)$ by replacing the function-name placeholder $\varphi(\xi)$ by names for the first-level functions $\xi^2 = 4$ and $\xi > 0$ (a function is at the first level if it takes its arguments from the domain of objects). These two functions take numbers as arguments and return the value *true* if those numbers are square roots of 2 or (respectively) positive, and *false* otherwise. In other words, they are exactly what Frege refers to as "concepts." It follows, then, that the general form of a quantifier, $\exists a\varphi(a)$, is that of a second-level concept, that is, a function taking first-level concepts as arguments and returning truth values. The modern view, made precise in the theory of *generalized*

Both have been extensively studied from a mathematical point of view (a survey can be found in Heinrich Herre et al., "The Härtig Quantifier: A Survey," *Journal of Symbolic Logic*, LVI, 4 (1991): 1153–183).

[33] The present approach thus differs from the usual study of generalized quantifiers, in which first-order logic is taken for granted: whenever logicians and linguists are interested in the properties of some quantifier Q, they explore the expressiveness of the language L_Q obtained by adding Q to full-fledged first-order logic (for instance, see Peters and Westerståhl, *op. cit.*).

quantifiers,[34] identifies such second-level concepts with collections of subsets of the domain. So for instance:

- The ordinary existential quantifier ∃ can be identified with the collection of all nonempty subsets of *D*;
- Dually the universal quantifier ∃ can be identified with the collection of subsets of *D* that contains *D* itself as its only member: ∃ = {*D*};
- The quantifier "there exist exactly *k*," usually written $\exists!^{k}$, can be identified with the collection of all *k*-membered subsets of *D*.

These examples apply to a single open formula φ(*x*) at a time: they are, as we will say, *unary*. However, some quantifiers are not only best viewed as applying to more than one such formula; they are also such that no other interpretation is possible. One such example is the quantifier Most. The statement Most*x*(φ(*x*), ψ(*x*)) represents "most φ's are ψ's," and it is true when more φ's are ψ's than φ's that are *not* ψ's. It is well known that Most cannot be represented by a formula of ordinary first-order logic.[35] In contrast, the quantifier Only, applying to formulas φ and ψ just in case all ψ's are φ's, can be expressed using the ordinary universal quantifier and Boolean connectives.

A further distinction concerns the dimension of a quantifier's arguments, as distinct from its number. For instance, a quantifier can simultaneously bind two variables *x* and *y* (thus having dimension 2), as in the case of the quantifier Q*xy*φ(*x*, *y*) which returns value true if and only if φ expresses the universal binary relation over *D*. All the above-mentioned quantifiers are *first-order*, a notion that can be characterized precisely in semantic terms. A unary quantifier is first-order if and only if it represents a collection of subsets of *D* (and similarly, a binary quantifier is first-order if and only if it expresses a relation between subsets of *D*). According to this definition, some quantifiers are first-order even though, like Most, they are not definable by a first-order formula.

The same is true of the Frege quantifier. The Frege quantifier represents a relation between subsets of the domain—the relation that holds between *F* and *G* when there are no more *F*'s than *G*'s. Hence, the quantifier relates *F* and *G* precisely when there is an injective function mapping the *F*'s into the *G*'s. Thus, it might appear that the Frege quantifier inherently appeals to a *second-order* notion. After all, existence claims for relations, functions, and so on, are properly expressed at the

[34] The theory originated with Andrzej Mostowski, "On a Generalization of Quantifiers," *Fundamenta Mathematicæ*, xliv (1957): 12–36; and Richard Montague, "English as a Formal Language," in R. H. Thomason, ed., *Formal Philosophy* (New Haven: Yale, 1974, originally published 1969).

[35] See Peters and Westerståhl, *op. cit.*

second order. But appearances are deceiving: the Frege quantifier is no more at the second order than ONLY or MOST, and just like them it *expresses*, but does not *assert*, the existence of a relation between the concepts appearing as arguments. The distinction between *expressing* and *asserting* the existence of higher-order entities is a crucial one, one that properly demarcates the first- from the second-order realm.

The property of *permutation invariance* also plays a crucial role in the modern conception of quantifiers. Quantifiers such as $\exists$ and $\forall$ answer the question "How many?" with no concern for the specific nature of the objects in question and are therefore invariant under permutations that swap around objects of the domain. Whereas notions of invariance for abstraction principles can be formulated in at least a few nonequivalent ways, it is easy to make precise such a notion in the case of quantifiers:

> If π is a permutation of D, then a binary first-order quantifier Q is *permutation invariant* if and only if for all subsets A and B of D, $\mathsf{Q}(A, B)$ holds precisely when $\mathsf{Q}(\pi[A], \pi[B])$ holds as well.

While the standard quantifiers of first-order logic are permutation invariant in the above sense, many more quantifiers enjoy this property, most notably those dealing with *cardinality* constraints (including the Härtig and the Rescher quantifiers). Among the latter, of course, is the Frege quantifier F. Our first task is to explore the expressive properties of the logical framework resulting from taking the Frege quantifier as the basic building block.

V. THE LANGUAGE OF THE ARITHMETIC

Formally, we consider a first-order language L_{F} with formulas built up from (individual or predicate) constants by means of Boolean connectives and the quantifier F; specifically, F takes two arguments, so that if $\varphi(x)$ and $\psi(x)$ are formulas, so is $\mathsf{F}x(\varphi(x), \psi(x))$.[36] The language of the Frege quantifier can be given a standard interpretation by supplying a recursive truth definition *à la* Tarski. Models for L_{F} look just like first-order models, providing a nonempty domain D and interpretations for nonlogical constants. The recursive clauses for the connectives are as usual, and formulas of the form $\mathsf{F}x(\varphi(x), \psi(x))$ are satisfied in the model if and only if there are no more objects in D that satisfy $\varphi(x)$ than there are objects satisfying $\psi(x)$.

The language thus defined is quite expressive. First, observe that the standard first-order quantifiers are expressible in L_{F}: a universally

[36] In the most general presentation, we will allow the formulas $\varphi(x)$ and $\psi(x)$ to contain parameters, and the quantifier to bind one or more variables simultaneously.

quantified formula $\forall x\varphi(x)$ can be represented by saying that the complement of φ is empty, that is, that there are no more objects satisfying $\sim\varphi(x)$ than there are satisfying $x \neq x$. Dually, an existentially quantified formula $\exists x\varphi(x)$ can be represented by saying that there is no injection of φ into the empty set. But the language turns out to be much more expressive than ordinary first-order logic. For instance, while infinity cannot be characterized in first-order logic using only the standard existential and universal quantifier, there is an axiom of infinity in the pure-identity fragment of L_{F} (such an axiom states that the universe is Dedekind-infinite). The negation of such an axiom, then, is true in all and only the finite domains, a fact that shows that, as a consequence, compactness fails.[37] In a similar way, for any formula φ one can express the fact that the set of objects satisfying φ is Dedekind-finite.

There is, however, an alternative interpretation of the Frege quantifier, which we regard as equally attractive—we refer to it as the *general* interpretation of F—on which the Frege quantifier is much less expressive. Recall that *second-order* quantifiers can be given, beside a standard interpretation, also a so-called general interpretation (first introduced by Henkin[38]). On such a general interpretation, second-order quantifiers are taken to range not over the "true" power-set of *D*, but over some previously given "universe" comprising some, but not necessarily all, subsets of *D*. So while standard models for second-order logic are indistinguishable from first-order models, general models carry, beside a domain *D*, also a universe of *n*-place relations over *D* (for each *n*).[39] All this is well known.

Perhaps more surprisingly, *first-order* quantifiers can also be so interpreted (an apparently hitherto unnoticed fact). Consider for instance the ordinary existential quantifier: as we have seen, in a classical first-order language this quantifier ranges over the collection of *all* nonempty subsets of the domain. But an alternative, "general" interpretation is possible as well, on which $\exists$ ranges over *some* collection of nonempty subsets of the domain.[40] The set of sentences valid on

[37] Since Härtig's quantifier is interpretable in L_{F} (by the Schröder-Bernstein theorem, sets *A* and *B* have the same cardinality if and only if there are injections from *A* to *B* and vice versa), any results about the expressiveness of I as detailed, for example, in Herre et al. (*op. cit.*), carry over to the Frege quantifier.

[38] Leon Henkin, "Completeness in the Theory of Types," *Journal of Symbolic Logic*, xv (1950): 81–91.

[39] In practice, such a universe of relations will satisfy some closure conditions—it will be, for example, closed under definability, thereby satisfying the second-order comprehension axiom.

[40] So the nonstandard existential quantifier can range over any collection of subsets omitting the empty set; likewise, then, the nonstandard universal quantifier would range over any collection of subsets as long as that collection contains *D* itself.

such an interpretation of the quantifiers turns out to be well known, if unexpected: it is the set of validities of positive free logic.[41]

In a similar vein, we consider a less expressive interpretation of the first-order quantifier F, which, just as in the case of $\exists$, is specified by singling out a particular class of models. By a general model for L_{F} we understand a structure providing a nonempty domain D and interpretations for the nonlogical constants, as well as a collection **F** of 1-to-1 functions between subsets of D. On this account, a formula $\mathsf{F}x(\varphi(x), \psi(x))$ is satisfied in the model if and only if there is a function f in **F** taking the set of objects satisfying φ into the set of objects satisfying ψ. In practice, we want the collection **F** also to satisfy certain *closure conditions*, which ensure that the language is still powerful enough for an adequate formalization of arithmetic.[42] One such closure condition, for instance, ensures that if there are no more F's than H's then there are no more F's and G's than H's.

Thus, we have two equally attractive ways to specify a semantics for the language of the Frege quantifier. This language can be given either the standard interpretation, in which F ranges over all injections between subsets of the domain, or the general interpretation, in which less comprehensive collections of functions are allowed. We regard the two interpretations as equally attractive.

One ingredient is missing in order to specify completely the language, namely, the abstraction operator. We thus introduce an abstraction operator Num mapping formulas into terms: Num(φ) picks out an object to be construed as "the number of φ." (Strictly speaking, Num is a variable-binding operator, so it properly should be written as Num$x\varphi(x, y)$ where y is a placeholder for possible parameters; in practice, the bound variable is understood). It is clear what a model for such a language would look like: on the *general* interpretation, besides supplying a nonempty domain D and interpretations for the nonlogical constants, a model would also supply both a collection **F** of injections between subsets of D and a function η taking subsets of D into D: the former, of course, is used for the interpretation of the Frege quantifier, while the latter provides an interpretation for the abstraction operator Num. On the *standard* interpretation, on the other hand, there is no need to specify the collection **F** (or, equivalently, **F** can be taken to be the collection of *all* injections between subsets of D.)

[41] See Antonelli, "Free quantification and logical invariance," in A. Paternoster, M. Andronico, and A. Voltolini, eds., *Rivista di estetica: Il significato eluso. Saggi in onore di Diego Marconi* (Essays in honor of Diego Marconi), xxxiv, 1 (2007): 61–73.

[42] The reader is referred to Antonelli, "Numerical Abstraction via the Frege Quantifier," for the technical details. See note 1.

VI. FORMALIZING ARITHMETIC

We now have both main components of our approach to arithmetic: the Frege quantifier, embodying a nonreductionist take on logicism; and the NUM operator, construed according to a deflationary view of abstraction. Special extra-logical axiom schemas formulated in the language L_{F} augmented with NUM will be needed to govern the interaction between the cardinality quantifier and the abstraction operator. We will give just a sketch of these axioms here, since the details are fully developed elsewhere.[43]

These extra-logical axiom schemas naturally fall into three main categories. The first group of axioms contains *definitional* and *uniqueness* principles, beginning first and foremost with Hume's Principle. HP can be expressed easily by asserting that the identity $\text{NUM}(\varphi) = \text{NUM}(\psi)$ holds if and only if there is a bijection between the φ's and the ψ's, that is, if and only if both $\mathsf{F}x(\varphi(x), \psi(x))$ and $\mathsf{F}x(\psi(x), \varphi(x))$ hold. In a similar vein, one can define an ordering relation by means of a schema saying that $\text{NUM}(\varphi) \leq \text{NUM}(\psi)$ if and only if $\mathsf{F}x(\varphi(x), \psi(x))$ holds (where $\leq$ is taken as primitive). Next, one characterizes the natural numbers by introducing a primitive predicate $\mathsf{N}(x)$ along with the axiom stating that $\mathsf{N}(x)$ holds if and only if x is the number of the predicate "natural number less than x," that is, $x = \text{NUM}(\mathsf{N}(y) \,\&\, y < x)$, and moreover there are only finitely many y's satisfying the predicate $\mathsf{N}(y) \,\&\, y < x$.

The second group of axioms comprises *existence* and *closure* principles. Among these a prominent role is played by an "infinitary" principle, stating that if a formula $\theta(x, y)$ defines a function from φ to ψ—that is, if for all x in φ there is exactly one y in ψ such that $\theta(x, y)$)—then $\mathsf{F}x(\varphi(x), \psi(x))$ holds. This axiom expresses the closure under definability of the space of injections used to interpret the Frege quantifier. A further axiom expresses an induction principle of the form "every finite and nonempty set of numbers has a maximum." One can then show the equivalence between the latter axiom and the standard induction schema ("every class of numbers containing zero and closed under successor contains all the numbers").

So far, we have been concerned uniquely with the basic arithmetical theory of successor and ordering. The third group of axioms is used to extend the theory to an account of the other arithmetical operations, especially addition and multiplication. While again we will not give the details here, we note that, on the most natural treatment, multiplication requires the version of the Frege quantifier binding two variables simultaneously (so that we can count pairs). Thus, beginning with cardinal

[43] *Ibid.*

properties, we have been able to recapture the basic structural features of the natural numbers within the context of a first-order language containing both the Frege quantifier and an abstraction operator.

VII. CONCLUSION

We have thus developed an account of arithmetic inspired by the twin principles of deflationary abstraction and nonreductionist logicism. The account has three main conceptual features:

1. It emphasizes the cardinal properties of the natural numbers over the structural ones, deriving the latter from the former, rather than the other way around, as is the case, for instance, with set-theoretic reductions.
2. It follows the Frege-Russell intuition that natural numbers are delivered by abstraction as representatives of equinumerosity classes.
3. It proceeds entirely at the first order from a semantical point of view—regardless of whether the standard or the general interpretation is chosen for the Frege quantifier.

In accordance with the deflationary view of abstraction, the "ultimate nature" of numbers is left completely unspecified on the present account, since abstract entities are picked from within the same domain over which our ordinary quantifiers range. So we open the way for a possible deflation of general worries concerning abstract objects in general, and numbers in particular. Abstract objects need not form a separate realm, but can be recruited—via abstraction principles such as HP—from among the ordinary objects of our ontology, a fact that leaves them accessible for use for whatever philosophical, logical, or mathematical purpose we might be pursuing. In this sense, the present account agrees with Benacerraf's general structuralist stance (as it often has been characterized) in emphasizing that the intrinsic nature of numbers is irrelevant for their role as abstract entities. Where the account diverges from Benacerraf's is in de-emphasizing order-theoretical properties, such as forming an ω-sequence, which are best viewed as "supervenient" upon cardinal ones.

Indeed, in keeping with the broadest and most general construal of logicism, cardinality notions do take center stage in the present account. Insofar as they deal directly with properties and relations of concepts—rather than matters of existence of objects such as numbers—cardinality notions properly can be regarded as having a logical character. Accordingly, we take the logicist claim that cardinality is a logical notion at face value, and rather than *arguing for it* (perhaps by providing a reduction to some other principle), we set out to explore its consequences by introducing cardinality, in the form of the Frege quantifier, as the main building block in the language of arithmetic.

G. ALDO ANTONELLI

University of California, Davis

COMMENTS AND CRITICISM

REALIZATION, REDUCTIOS, AND CATEGORY INCLUSION*

Thomas Polger and Laurence Shapiro argue that Carl Gillett's much-publicized dimensioned theory of realization is incoherent, being subject to the following *reductio*:

(P1) Everything that is realized is a property instance, and at least one property instance is realized. (Gillett's account)
(P2) Some things that are realized are multiply realized. (assumption)
(P3) No property instance is multiply realized. (trivial)
(C1) Some property instances are multiply realized. (from P1 and P2)
(C2) Not (P1). (*reductio* from P3 and C1)[1]

The *reductio* turns on the fact that (P1) makes property *instances* the exclusive relata of the realization relation, while the conjunction of (P2) and (P3) implies its denial, namely, that *properties* are the relata of the realization relation on occasions of multiple realization. Polger and Shapiro are correct to see an apparent puzzle here. Gillett defines realization in terms of property instances, and he accepts the multiple realization of properties.[2] He is not alone. Sydney Shoemaker appears to accept (P1) when he says: "to speak of one property as realizing another is shorthand for saying that instances of the one are among the possible realizers for instances of the other."[3] And he also accepts the multiple realization of properties, thus giving credence to (P2) and presumably (P3). But I am not primarily concerned about whether Polger and Shapiro are correct in attributing the corresponding beliefs to Gillett, or whether they can be extended to others like Shoemaker.[4] Rather, I am interested in the more general issue their argument raises for theories of realization and their underlying metaphysics. In particular,

* I thank John Carroll, Carl Gillett, and Jaegwon Kim for some helpful comments and discussion.

[1] Polger and Shapiro, "Understanding the Dimensions of Realization," this JOURNAL, CV, 4 (April 2008): 213–22, at p. 214.

[2] See Gillett, "The Dimensions of Realization: A Critique of the Standard View," *Analysis*, LXII, 4 (October 2002): 316–23, at p. 322, where he defines realization in terms of property instances.

[3] Shoemaker, *Physical Realization* (New York: Oxford, 2007), p. 3.

[4] In personal communication Gillett says he rejects (P1). See also his "Multiply Realizing Scientific Properties and their Instances: A Response to Polger and Shapiro," (forthcoming in *Philosophical Psychology*). But his reasons are different than mine. See my discussion in section III.

0022-362X/10/0704/213–219

regarding (P1) they say: "On Gillett's account, realization is defined as a relation between property instances. It follows that only property instances can be realized on his view."[5] But a definition of realization cast in terms of property instances does not imply (P1) by itself, since one can add an auxiliary assumption that allows the concept of realization to range over multiple ontological categories. To that end I will supplement a dimensioned theory with the necessary category-inclusive proposition. Consequently, one may consistently reject (P1) on grounds that properties are also realized. Alternatively, one may consistently reject (P3) on grounds that property instances are also multiply realizable.[6] I will then offer a few reasons to justify the proposed category-inclusive view of realization.

I

Consider a parallel case. Jaegwon Kim describes a concept of *event* supervenience, which picks out a relation of determination between particulars.[7] Kim also allows for the possibility that certain properties have *multiple* subvenient bases, as required by certain strong interpretations of multiple realizability.[8] So event supervenience concerns particular instances, while multiple supervenient bases concern properties. That fact notwithstanding, one should not accept the following *reductio* of Kim's ideas about event supervenience:

(P1*) Everything that supervenes is a property instance or token event, and at least one property instance or token event supervenes.
(P2*) Some things that supervene have multiple subvenient bases.
(P3*) No property instance or token event has multiple subvenient bases.
(C1*) Some property instances or token events have multiple subvenient bases. (from P1* and P2*)
(C2*) Not (P1*). (*reductio* from P3* and C1*)

This *reductio* is unacceptable because Kim defines supervenience in terms of families of *properties*, and then he explains event supervenience

[5] Polger and Shapiro, *op. cit.*, p. 214.

[6] One may find the denial of (P3) especially objectionable on grounds that only properties can have multiple instances. But compare Richard Boyd's claim about the transworld compositional plasticity for token events in his "Materialism without Reductionism: What Physicalism Does Not Entail," in Ned Block, ed., *Readings in the Philosophy of Psychology*, vol. 1 (Cambridge: Harvard, 1980), p. 99. The denial of (P3) is also possible on metaphysical schemes that identify particulars with properties, say, a bundle theory minus individual nonduplicating heacceities.

[7] Kim, "Epiphenomenal and Supervenient Causation," *Midwest Studies in Philosophy*, IX (1984): 257–70.

[8] *Ibid.*, p. 261; see also his "Concepts of Supervenience," reprinted in *Supervenience and Mind* (New York: Cambridge, 1993), pp. 53–78, at p. 65.

in terms of property supervenience.[9] Specifically, for the latter project he offers the following category-inclusive coordinate definition:

(CD_s) An event, *x*'s having *F*, supervenes on the event, *x*'s having *G*, just in case *x* has *G* and *G* is a supervenience base of *F*.[10]

This permits Kim to say that both properties and events stand in a supervenience relation, contrary to the exclusive claim represented by (P1*). However, the unacceptable *reductio* is just Polger and Shapiro's *reductio*, only cast in terms of supervenience rather than realization. This suggests that one can avoid their argument on the model provided by Kim, specifically, by defining a dimensioned theory of property realization and then sketching a larger category-inclusive theory whereby property-instance realization is explained in terms of property realization.

II

Gillett says that a "flat" theory presents realization in terms of the same object instantiating the realized and realizing property, while a "dimensioned" theory presents realization in terms of an object and its proper parts, where the causal powers of the former are composed from the several causal powers of the latter in a way befitting that dimensioned mereology.[11] Hence, a dimensioned theory of *property realization* must incorporate these claims. I propose:

(DR_P) Property *F* is dimension-realized by properties P_1–P_n if and only if (i) there is an object *s* with proper parts p_1–p_n such that *F* is instantiated by *s* and P_1–P_n are instantiated by p_1–p_n, and (ii) the causal powers that *F* bestows upon *s* are composed from or otherwise determined by the distinct causal powers that P_1–P_n bestow upon p_1–p_n.

The notion of dimensioned realization for *property instances* can then be explained on the basis of this notion of dimensioned property

[9] Kim defines supervenience as a relation between families of properties in various places, for example, in "Concepts of Supervenience," p. 65, and "Supervenience as a Philosophical Concept," reprinted in *Supervenience and Mind*, pp. 131–160, at p. 140.

[10] "Epiphenomenal and Supervenient Causation," *Midwest Studies*, p. 262.

[11] Gillett, "The Dimensions of Realization," and "The Metaphysics of Realization, Multiple Realizability, and the Special Sciences," this JOURNAL, C, 11 (November 2003): 591–603. Gillett states his version of a dimensioned theory as follows: "Property/relation instance(s) P_1–P_n realize an instance of a property *F*, in an individual *s*, if and only if *s* has powers that are individuative of an instance of *F* in virtue of the powers contributed by P_1–P_n to *s* or *s*'s constituent(s), but not vice versa" ("The Dimensions of Realization," p. 322, with variables changed to match my own). Like Gillett's definition, my definition (DR_P) only mentions the realized *F* and the realizing proper part properties P_1–P_n, not some additional organizational feature, or structural property, or role-player possessed by *s*. So the definition only presents a core notion of dimensioned property realization.

realization by means of the following category-inclusive coordinate definition:

(CD_r) Property instance, *s* having *F*, is dimension-realized by the collection of property instances that constitute its proper parts, p_1–p_n having P_1–P_n, just in case p_1–p_n instantiate P_1–P_n and the collection of properties P_1–P_n is a realization base for the property *F*.

Like Kim's coordinate definition for supervenience (CD_s), I formulate (CD_r) so that the realization of properties is a necessary condition for the realization of property instances. As such, it does not provide a reductive analysis.[12] Even so, with (CD_r), both properties and property instances can be said to stand in a dimensioned realization relation, contrary to premise (P1) of the *reductio*. My proposal also accords well with a large number of statements about realization in the philosophical literature that involve different ontological categories. Thus, Ernest Lepore and Barry Loewer say that an event with a physical property realizes an event with a mental property.[13] But Colin McGinn says that a physical property realizes an intentional property.[14] Even the same philosopher crosses ontological categories. Putnam speaks about not only the realization of objects like Turing machines but also the realization of their state types.[15] A category-inclusive view of realization explains this linguistic variation. Of course, other explanations are possible. For example, some philosophers might maintain a category-exclusive view by treating the language of property realization uniformly as a convenient shorthand for longer statements about property-instance realization.[16] But if so, it is incumbent on those philosophers to provide a

[12] For example, the definition of property realization (DR_p) only mentions objects instantiating properties, so it does not imply property instances without an additional existence condition for property instances mentioned in the paragraph that immediately follows.

[13] Lepore and Loewer, "More on Making Mind Matter," *Philosophical Topics*, XVIII (1989): 175–91, at p. 179.

[14] McGinn, "Philosophical Materialism," *Synthese*, XLIV, 2 (June 1980): 173–206, at p. 196.

[15] See Hilary Putnam, "Minds and Machines," reprinted in *Mind, Language and Reality: Philosophical Papers*, vol. 2 (New York: Cambridge, 1975), pp. 362–85, at p. 371; and "The Nature of Mental States," reprinted in *Mind, Language and Reality*, pp. 429–40, at pp. 434, 438. Technically (CD_r) does not address object realization. But one can formulate the appropriate coordinate definition by linking object realization to property realization.

[16] As I stated in the introduction, Shoemaker appears to take a category-exclusive position when, in *Physical Realization*, he says: "to speak of one property as realizing another is shorthand for saying that instances of the one are among the possible realizers for instances of the other" (p. 3). But, after stating that what is realized is a property instance for both "same object property realization" and "microphysical realization," Shoemaker also says: "it is not excluded that other sorts of entities should be said to be realized" (p. 4). My category-inclusive proposal can reconcile such remarks.

clear meaning for the language of multiple realization involving properties when, on their view, realization is always property-instance realization. That may be a small chore, but the alternative category-inclusive view faces no such task, since it takes the language of property realization at face value.

It is also worth emphasizing that a category-inclusive result is not achieved by the definition of property realization (DR_p) alone. First, (DR_p) only refers to an object instantiating a property, which does not imply that there exist items from the further category of property instances unless it is taken in conjunction with the appropriate metaphysical assumption according to which a property instance *s* having *F* exists when the object *s* instantiates property *F* (there are more sparse ontologies which deny that assumption). Second, even if one accepts the required existence condition, (DR_p)'s domain of discourse is explicitly stated to be properties in the definiendum. Hence, the role of the coordinate definition (CD_r) is to extend the interpretation of its key predicate '*x* is realized by *y*' to include property instances. Similar remarks would apply if one were to begin with a definition of dimensioned realization stated in terms of particular instances. The appropriate category-inclusive coordinate definition would then serve to extend the interpretation of its realization predicate to include properties.

III

I think the category-inclusive view of realization represented by (CD_r) succeeds in avoiding the stated *reductio*, meaning that it permits a consistent set of beliefs regarding properties and their instances as the relata of the designated realization relation. But it does not establish that such a view is *plausible*. Is it justified to extend the interpretation of the realization predicate, or the concept expressed by that predicate, in a category-inclusive way?

One might think that a category-inclusive view of realization is justified with the aid of general metaphysical principles. For example, a doctrine of property immanence ensures that particulars are present whenever properties are realized, and an account of realization adorned with a causal theory of properties makes both properties and their instances relevant to the realization relation because such an account requires that the realized and realizing properties contribute powers to their instances.[17] But such metaphysical doctrines only ensure that items from the stated categories are present and relevant when realization occurs. A category-inclusive coordinate definition is still needed to ensure that

[17] Gillett argues along these lines in his "Multiply Realizing Scientific Properties and their Instances."

a theory of realization picks out the desired items as the relata of the realization relation. After all, a given property and its instance are distinct entities. The former can exist apart from the latter. Moreover, a concept or theory of realization can track the property but not the particular instance. Compare that a concept can track just one among multiple co-occuring events, or just one among multiple co-extensive properties (by some accounts, even just one among multiple lawfully co-extensive properties).

Worse, other metaphysical relations can elicit the opposite category-exclusive judgment in spite of the truth and application of the same general metaphysical principles. For example, some philosophers believe that token events are the relata of the causal relation, not objects *per se*, and not properties either.[18] Yet certain properties of those events are certainly present and relevant on occasions of causation. Something can be judged present and relevant but not the relata of the relation in question. So a category-inclusive view does not appear to stand by doctrines of property immanence and causal powers alone.

Nevertheless, I think a category-inclusive view of realization can be justified. As a start, given the parallel between the proposed category-inclusive view of realization and Kim's treatment of property and event supervenience, it seems warranted to conclude that the former is justified if the latter is justified. At least this conditional judgment seems warranted until one explains why realization has certain distinguishing features that prevent the parallel category-inclusive analysis. Yet this result is not entirely satisfying, since one might also challenge the category-inclusive treatment of supervenience, and since other metaphysical relations like causation can elicit the opposite category-exclusive judgment. So an important question remains—why is it justified to treat supervenience and realization in a category-inclusive way, and causation in a category-exclusive way?

I think there is a plausible answer. Brain McLaughlin points out that, while 'supervenience' and 'realization' are philosophical terms of art whose stipulative meanings can only be judged by their theoretical

[18] Most notably, Donald Davidson, "Actions, Reasons, and Causes," and "Causal Relations," both reprinted in his *Essays on Actions & Events* (New York: Oxford, 1980), pp. 3–19 and 149–62, respectively. This view of particular events as the relata of causation also leads some philosophers to use a different term for the role of properties in causation, such as 'causal relevance' or 'quausation'. See Terence Horgan, "Mental Quausation," *Philosophical Perspectives*, III (1989): 47–76. As Stephen Yablo puts it: "Although causes and effects are events, properties as well as events can be causally relevant or sufficient," from his "Mental Causation," *The Philosophical Review*, CI, 2 (April 1992): 245–80, at p. 247, fn. 5. Of course, not all philosophers exclude properties from the causal relation. See Frank Jackson, "Essentialism, Mental Properties and Causation," *Proceedings of the Aristotelian Society*, XCV (1995): 253–68, at p. 254.

utility, 'causation' is a term grounded in common usage that carries substantial pre-theoretic intuitions.[19] This suggests the hypothesis that, while a category-inclusive analysis of supervenience and realization are justified by their theoretical utility, a category-inclusive analysis of causation falters, if it does falter, when weighed against pre-theoretic intuitions and ordinary language sentences that favor causes as concrete particulars. Moreover, like a category-inclusive analysis of supervenience, a category-inclusive analysis of realization is, indeed, theoretically useful. Why? Because, in either case, the pertinent coordinate definition provides a measure of unification under a core notion of determination. For supervenience, there can be no difference in one set of entities (the supervenient ones) without a corresponding difference in another (the subvenient ones). That notion holds true for properties and property instances alike, and the coordinate definition (CD_s) fixes the extension accordingly. For realization, one entity (the realizer) lawfully necessitates another entity (the realized).[20] That notion holds true for properties and property instances alike, and the coordinate definition (CD_r) fixes the extension accordingly. A competing category-exclusive position would seem to multiply meanings beyond necessity, creating different senses for property and instance determination where there appears to be none.

RONALD P. ENDICOTT

North Carolina State University

[19] McLaughlin, "Mental Causation and Shoemaker-Realization," *Erkenntnis*, LXVII, 2 (September 2007): 149–72, at pp. 149–50. John Carroll also made the same point in discussion. I add that, because 'realization' is largely a technical term of art when used by philosophers, it can pass certain ordinary language tests for category inclusion in a trivial way. Compare how Lowe explains why the verb 'to cause' has a different sense when applied to object causation versus event causation, namely, that while it is not incongruous to say "Smith and Jones together caused the collapse of the bridge," it *is* incongruous to say "The explosion of the bomb and Jones together caused the collapse of the bridge." See E. J. Lowe, *A Survey of Metaphysics* (New York: Oxford, 2002), p. 196. But statements of mixed categories are perfectly fine for realization. So it is not incongruous to say "John's mind and its mental properties" (object and property); or "John's brain event and its physical properties together realized John's mind event and its mental properties" (event and property); or "John's brain and its physical events together realized John's mind and all its mental events" (object and event). In my view, these statements do not violate ordinary language conventions because, at present, ordinary language has *no* well-established conventions about the philosophical vocabulary of realization to violate.

[20] Philosophers typically analyze realization as a determinative relation that implies one-way conditional laws. See Lepore and Loewer, "More on Making Mind Matter," p. 179; Michael Tye, *Ten Problems of Consciousness* (Cambridge: MIT, 1995), p. 41; and Kim, *Philosophy of Mind* (Boulder, CO: Westview Press, 1996), p. 133.