Reference Fixing and the Paradoxes

(Draft, May 23, 2021. Revised version forthcoming in M. Petrolo and G. Venturi (eds.), Paradoxes between Truth and Proof, Springer, forthcoming.*)*

Mario Gómez-Torrente
Instituto de Investigaciones Filosóficas
Universidad Nacional Autónoma de México (UNAM)
mariogt@unam.mx

Abstract: I defend the hypothesis that the semantic paradoxes, the paradoxes about collections, and the sorites paradoxes, are all paradoxes of reference fixing: they show that certain conventionally adopted and otherwise functional reference-fixing principles cannot provide consistent assignments of reference to certain relevant expressions in paradoxical cases. I note that the hypothesis has interesting implications concerning the idea of a unified account of the semantic, collection and sorites paradoxes, as well as about the explanation of their "recalcitrance". I also note that it does not necessarily imply that one should not expect the sometimes hoped for "unique" solution to a paradox of these kinds.

In this paper, I will consider and defend the hypothesis that the semantic paradoxes, the paradoxes about collections, and the sorites paradoxes, are all, in a certain sense, paradoxes of reference fixing. On the hypothesis, they are all paradoxes of reference fixing in the sense that they arise from implicitly adopted, purported reference- or content-fixing principles or sets of principles for certain relevant expressions, principles that fail to work as intended in the cases revelatory of the paradoxes—and in other cases. We have reason to postulate implicitly adopted, purported reference-fixing principles for many expressions of our languages, and we know that, although such principles typically do their job reasonably well (otherwise they would have probably ceased to be conventionally adopted principles), if they exist as adopted principles then they do occasionally give rise to problems. A given set of reference-fixing principles may fail to determine a referent for an expression in its intended range of application in at least a couple of familiar ways:

the principles as applied in a particular case may assume or imply some empirical falsehood, or they may not contradict any truth but just be insufficiently strong to eliminate lurking indeterminacies and determine some one referent.[1] In the case of the semantic, collection and sorites paradoxes, we will point out that the relevant reference-fixing principles as applied in the relevant particular cases do not assume or imply empirical falsehoods, but instead *a priori* imply that certain expressions cannot have referents assigned in a consistent way in those cases; and we will also emphasize how those same principles also give rise to indeterminacies in other cases.

Our hypothesis here is, then, that the semantic, collection, and sorites paradoxes all involve reference-fixing principles, and we will defend it by arguing that all these paradoxes arise from comparatively localized failures of principles of this kind that work well in many other cases. My main aims are, first, to make the hypothesis that the semantic, collection and sorites paradoxes are paradoxes of reference fixing appear initially plausible through direct consideration of it as it applies to the different kinds of paradoxes (section 1); and second, to discuss some consequences and non-consequences of the hypothesis, in the hope that the discussion will make it appear still more plausible (section 2). In particular, I will compare the kind of unified explanation of the sources of the sorites and the semantic and collection paradoxes that the hypothesis provides with other ideas about what unifies the sorites with the other paradoxes, and I will relate the present idea to an explanation of the "recalcitrance" of these paradoxes. The investigation here will be "diagnostic", in the sense of Chihara (1979), i.e. one intended to throw light on the roots of the paradoxes, rather than one intended to devise "treatments" or solutions of some kinds. In fact, I think that the hypothesis by itself has no strong implications concerning the search for solutions to the paradoxes; I will conclude with a brief discussion of this latter claim, as I suspect that the hypothesis might be thought to imply more or less directly that there is no such thing as the sometimes hoped for "unique solution" to the

[1] There are many places where one can find expositions of problems of this sort for reference-fixing conventions. See my book Gómez-Torrente (2019) for discussion and references.

paradoxes, while, if I'm correct, it doesn't really have this implication.

1. Paradoxes and reference fixing.

1.1. Semantic paradoxes and reference fixing.

Let's begin by considering the semantic paradoxes, taking the Liar as our first example. In a standard version, one gives the name '(1)' to the sentence

(1) (1) is false,

and reasons as follows: (1) must be either true or false; suppose first that (1) is true; then, what it says must be the case; but what it says is that (1) is false, so if (1) is true, (1) is false; since this can't be, suppose, second, that (1) is false; but that (1) is false is what (1) says, and so since what (1) says is the case, (1) is true, and then we have established that if (1) is false, (1) is true, which can't be either.

One key tacit assumption in this reasoning is the assumption that a sentence is true just when what it says is the case. As many have noted, although this assumption may give rise to a number of metaphysical worries concerning the idea of something "being the case", an apparently metaphysically innocuous version that suffices to develop the Liar reasoning above is the so-called "Truth convention" involving the so-called "T-schema":

> *Truth convention:* All instances of the following schema are assertable/ usable in reasoning/ true:
>
> (T-schema) 'S' is true if and only if S,
>
> where in place of " 'S' " goes a name of a certain sentence (e.g. a quotation of it) and in place of 'S' goes that sentence.

The instance of the T-schema implicitly used (a couple of times) in the Liar reasoning above is, of course,

> (1) is true if and only (1) is false.

As a second example of semantic paradox, take Yablo's paradox. This arises from consideration of the following infinite list of sentences, where intuitively each sentence says that the sentences that follow it in the list are false:

(S_1) For any $n > 1$, S_n is false.
(S_2) For any $n > 2$, S_n is false.
(S_3) For any $n > 3$, S_n is false.
.
.
.
(S_m) For any $n > m$, S_n is false.
.
.
.

The paradoxical reasoning here is as follows: either all the sentences in the list are false, or some are true; suppose first that at least one, S_a, is true; if so, for any $n > a$, S_n is false; then S_{a+1} is false; but S_a being true also implies that for any $n > a+1$, S_n is false, and this in turn means that S_{a+1} is true, which can't be; so suppose, second, that all the sentences in the list are false; then S_1 is false, and so is S_n for any $n > 1$; but this is precisely what S_1 says, so S_1 must be true, which again can't be. Note that here again the T-schema is employed a couple of times, namely when

S_a is true if and only if for any $n > a$, S_n is false

is used, and when

S_1 is true if and only if for any $n > 1$, S_n is false

is used.

As a final example, consider Curry's paradox, in the following form. Let (2) be the sentence

(2) If (2) is true, then $0=1$ and $0\neq1$.

We reason thus: either (2) is true, or it is false; first suppose that (2) is false; then its antecedent is true (as (2) is a material conditional); but that antecedent is *(2) is true*, and so (by an application of an instance of the T-schema) (2) is true, against our supposition. So suppose, second, that (2) is true; then (again by an application of an instance of the T-schema) if (2) is true, then 0=1 and 0≠1; and so, by *modus ponens*, 0=1 and 0≠1, which can't be.

There are at least two reasons why it is plausible to see these (and other) semantic paradoxes as paradoxes of reference fixing. The first reason is that, as we have seen, they all make implicit use of (something like) the Truth convention involving the T-schema. This makes it plausible to think of them as paradoxes of reference fixing because it's natural to think of the Truth convention as an implicit conventional principle purporting to fix the reference of the predicate 'true'. The implicit idea is that it should do this since it gives the conditions under which each particular sentence is true, and these should in turn determine (along with the corresponding relevant facts) the set of sentences that constitute the extension of 'true'. The second reason, however, is that the Truth convention arguably fails to achieve its reference-fixing purpose, at least fully, and that the way it fails mirrors the way purported reference-fixing conventions fail to achieve their purpose in certain occasions. We will later turn to this second reason why it's plausible to see the semantic paradoxes as paradoxes of reference fixing, but first let's describe in appropriate detail the reasons why the Truth convention and the T-schema fail, which are intimately involved in the paradoxes.

First we must emphasize that even if the Truth convention does not fully achieve its purpose, it does achieve it partially, in the sense that many sentences do get acceptable truth conditions by means of the T-schema. For example, given the instance of the T-schema

'0=0' is true if and only if 0=0,

and given that it is the case that 0=0, we know that *0=0* is true, that it is in the extension of 'true'; and given the instance of the T-schema

> '0=1' is true if and only if 0=1,

and given that it is not the case that 0=1, we know that *0=1* is not true, and so not in the extension of 'true'. Also, given the instance

> 'In 2021, Betelgeuse has already gone supernova' is true if and only if in 2021, Betelgeuse has already gone supernova,

the sentence *In 2021, Betelgeuse has already gone supernova* will be in the extension of 'true' if in 2021, Betelgeuse has already gone supernova, and outside it if it is not the case that in 2021, Betelgeuse has already gone supernova, even if we don't know which of the two things is the case.

But in some cases, the Truth convention doesn't work well. Specifically, at the very least, it doesn't work well when it is applied to "ungrounded" sentences (in the sense of Herzberger (1970) and Kripke (1975)). Consider this sentence, "the truth-teller":

(3) (3) is true.

Is (3) in the extension of the truth predicate, or outside it? The corresponding instance of the T-schema,

> (3) is true if and only if (3) is true,

is evidently of no help in answering this question, and not because of any lack of knowledge on our part: the condition under which that instance of the T-schema tells us that (3) is true simply repeats that (3) is true, and this is precisely what we wanted conditions for. To be told that (3) is true if and only if (3) is true doesn't determine whether (3) is in the extension of the truth predicate in any independent, non-circular way, not even in the presence of relevant additional but possibly unknown facts. (3), however, is not paradoxical, in the sense that the assumption that it is either true or false does not lead us to any contradiction; so we don't have that sort of argument for the conclusion that (3) determinately lacks content and truth value.

The Liar sentence (1) gives rise to a similarly failed instance of the T-schema, namely *(1) is true if and only if (1) is false*. Here the condition under which this instance of the T-schema says that (1) is true is, at least in part, that (1) is not true (assuming, as is natural, that 'false' means the same as something like 'contentful and not true'). But again this is evidently circular: one cannot give reference-fixing conditions for the application of a predicate to a thing in terms of whether that same predicate applies (or not) to that same thing. Furthermore, in this case things are even worse; while no contradiction could arise if the reference-fixing condition were not defective, in this case its defectiveness allows the *a priori* derivation of a contradiction from the assumption that (1) is either true or false. The paradoxical reasoning thus shows in an *a priori* way that no consistent assignment of truth-value to (1) is possible given the instance of the T-schema at stake.

Turn now to Yablo's paradox, which is especially relevant as it does not appear to involve circularity, while circularity seems a clear part of what leads to the problems for reference fixing just outlined. Consider for example the appeal to the T-schema instance

S_1 is true if and only if for any $n > 1$, S_n is false,

which gives the condition for the truth of S_1. Now in a clear sense (and against Yablo's claims about his own paradox) there is a certain kind of self-reference in S_1 (and in all the S_n's), at least from a classical understanding of the notation 'S_n': this notation involves reference to a function or functional sequence one of whose components (as value of 1) is S_1 itself (and the same holds for all the S_n's), so reference to S (or S_n) involves reference to S_1 (and to all the S_n's). But I don't think this remark by itself shows anything about the failed nature (from the point of view of reference fixing) of the instances of the T-schema involved in Yablo's paradox. They are failed instances, however, in this case because the search for a truth condition determining how S_1 (or any S_n) is true or false would involve us in an infinite regress. Again, the condition for the truth of

S_1 is that for any n > 1, S_n is false; well, to see what this amounts to, we must turn to the S_n's for n > 1, beginning with S_2, whose truth condition according to the T-schema would be that for any n > 2, S_n is false; and to see what this condition amounts to, we must turn to the S_n's for n > 2, beginning with S_3, whose truth condition according to the T-schema... It's clear that no real truth condition for S_1 (or any S_n) is ever actually fixed by the infinite relevant instances of the T-schema. The contradiction additionally *a priori* shows that no truth-value assignment to them would be consistent.

For our purposes, the case of Curry's paradox is not substantively different from the case of the truth-teller or the Liar sentence. The condition under which (2) is true according to the relevant instance of the T-schema is that, if (2) is true, then 0=1 and 0≠1. Again this doesn't determine whether (2) is in the extension of the truth predicate in a non-circular way, not even in the presence of additional relevant facts: the condition for the truth of (2) involves an appeal to the idea of the truth of (2) itself. No determination of whether (2) is in the extension of 'true' or not can arise from such a condition, and in fact, given the relevant instance of the T-schema, we can show that (2) could be neither true nor false.

If paradoxical sentences are not determined to be either true or false by what, to all appearances, is the reference-fixing principle we have for truth ascriptions, we lose one of the key assumptions that set the paradoxical reasonings going, namely, that the paradoxical sentence or sentences at stake are either true or false. In fact, as we have seen, the paradoxical reasonings can be seen as *reductiones ad absurdum* of the implicit idea that the Truth convention determines that the relevant sentences have a truth-evaluable content and thus must be either true or false.

At this point something ought to be said about the so-called "revenge problem" for views which, like the present one, postulate or imply that paradoxical sentences are neither true nor false because they are contentless or semantically defective. The idea of revenge is that if we name '(4)' the evidently problematical sentence

(4) (4) is not true,

(the so-called "Strengthened Liar", where 'not true' means the same as 'either false or contentless'), we can reason much as in the Liar paradox to reach a contradiction from either of the three suppositions that (4) is true, false, or contentless. The reasoning in the case of the first two suppositions is clear. In the case of the supposition that (4) is contentless, the following is a representative formulation:

> Assume that [(4)] is [contentless]. If [(4)] is [contentless], then it is either false or [contentless]. It *says* that it is either false or [contentless]. So it is true. Hence it is both true and [contentless]. This violates the Law of Non-Conflict [which says that no sentence is more than one of true, false, or contentless]. Contradiction. (Cook (2013), 81, my emphasis)

It is not infrequent to find in the literature arguments, based essentially on this reasoning, to the effect that views on which paradoxical sentences (like (4)) are neither true nor false are inconsistent: these views—so go these arguments—say that (4) is neither true nor false, and hence not true, so they accept that (4) is true, and this is contradictory. (See e.g. Bacon (2015), sec. 1, Armour-Garb (2017), 10.) In my view, these arguments exploit a conflation of the informal diagnostic exposition of the philosophical reasons why sentences like (4) are contentless (such as the exposition we have given above) with the (presumably formal) way in which such a diagnosis would be formulated under the strictures of a solution or treatment of the paradoxes. Thus, in the informal exposition we say that (4) and the like are neither true nor false, and hence not true, which suggests that we are condemned to contradict ourselves accepting that (4) is true (isn't that (4) is not true what (4) *says*?); but if we were to set out to formulate the claim that (4) and the like are neither true nor false using our preferred solution (presumably one of the formal solutions inspired by the idea that paradoxical sentences are neither true nor false; see below), the way in which we would formulate it would be perfectly consistent—for example, we might say that (4) is "not true" using a Tarskian truth predicate of a higher level than that in (4), or say that (4) is "neither true nor false" using a Kripkean "(un)groundedness" predicate indefinable within the language fragment that we take (4) to belong to.[2] The proponents of these arguments can (and do) additionally claim either that it is a defect of the *informal diagnosis* that it cannot be consistently stated at the purely informal level, or that it is a crippling defect of the formal solutions inspired by the informal diagnosis that the semantic predicates they use are not all at the same language level, but I see no convincing reason to accept either of these claims.[3]

We can now develop more fully the second reason why it's plausible to see the semantic paradoxes as paradoxes of reference fixing. This is the idea that the wrong results of the Truth convention parallel the problematic results occasionally produced by

[2] Inevitably, the arguments in question appeal to some premise that will not be validated within a formal solution inspired by the idea that paradoxical sentences are neither true nor false (much as some of the proponents of the arguments may believe that the proponents of the idea *ought to accept* that premise). For example, the argument cited from Cook (2013) in the main text appeals tacitly to the supposition that (4) *says* that (4) is not true, and is thus contentful. But if we suppose that (4) is contentless, then it *doesn't* say anything; in particular, it doesn't say that it itself is either false or contentless, as the T-schema instance

(4) is true if and only if (4) is not true

(where 'not true' means the same as 'either false or contentless') is a failed instance that will have no correspondence in our formal theory. In the case of Armour-Garb's (2017) reasoning, there is a similar step going from *σ is either meaningless or not true* (σ is 'σ is either meaningless or not true') to *σ is meaningful*, which would again be unwarranted in any of the familiar formal reconstructions. In the case of Bacon's (2015) argument, the problematic premise is the "SRT schema" to the effect that all sentences of the form

If 'ϕ' is meaningful, then ('ϕ' is true iff ϕ),

are assertable/ usable in reasoning/ true, which is of course unacceptable in a formal reconstruction inspired by a diagnosis of our kind (at the informal level, we can see that when 'ϕ' is replaced by a paradoxical sentence in SRT, the result is just as meaningless as the paradoxical sentence itself).

[3] Thus, for example, Armour-Garb ((2017), 10), protests that "[σ] is perfectly good English" and, as such, ought to be dealt with within English itself. But part of the point made by our diagnosis is that, regardless of the appearances to the contrary, σ is *not* "perfectly good English".

other implicit reference-fixing conventions: such reference-fixing conventions also give both indeterminacy results (as in the case of the truth-teller) and determinate failure of reference results (as in the Liar and the other paradoxes). The reference-fixing task, so to speak, is known to generate problems of the same sort as the principles implicitly involved in the paradoxes, and this suggests again that the source of the problems lies in both cases in the adoption of problematic reference-fixing conventions.

To illustrate this, consider the widely accepted idea that underlying the common use of ordinary "natural kind terms" is a convention of roughly this form:

> *Ordinary natural kind term convention.* A term intended to be a natural kind term from ordinary language refers to the natural kind (substance, species, natural phenomenon, etc.) shared by at least most of the things that are ordinarily taken to be paradigms of the kind.

Although there are philosophical disputes involving strongly skeptical views that natural kinds don't really exist (and thus cannot be referred to) or that the Ordinary natural kind term convention can never fix referents for ordinary natural kind terms due to a number of sophisticated problems,[4] for present purposes it's reasonable to grant that in many, perhaps a vast majority of cases ('water', 'dog', 'heat', 'measles', etc.), the convention does deliver a certain, possibly vague natural kind as referent for a relevant term. However, it's also reasonable to think that the Ordinary natural kind term convention sometimes fails to deliver a natural kind as referent for a term intended to be a natural kind term, and sometimes it leaves things unclear and probably indeterminate.

As an example of the first kind, take the term 'miasma', presumably an ordinary natural kind term which by the lights of all sensible people just has not gotten a natural kind as referent via the Ordinary natural kind term convention, and certainly did not have a referent even when belief in miasma was popular. (It may have

[4] See Gómez-Torrente (2019), ch. 5, for discussion of some of these philosophical issues surrounding (refined versions of) the Ordinary natural kind term convention.

eventually gotten some more or less special descriptive and/or fictional kind as referent in a special use, as we will mention later.) There is no natural substance behind the foul odors of rotting materials once taken as paradigms of miasma (not even behind most of them), and thus the assumption that there is simply contradicts that empirical truth. As an example of the second kind, think of a term such as 'madness', again presumably a natural kind term from the ordinary point of view. Here, although it surely has turned out that the traditional paradigms are instances of diseases or other phenomena that are often quite different—epilepsy, dementia praecox, all kinds of so-called neuroses and psychoses, etc.—it is unclear that this places 'madness' in the same category as 'miasma', as an intended natural kind term that determinately lacks reference. Although this option can be defended, it also seems defensible to claim that, assuming there is such a (high-level) natural kind as *psychiatric disorder*, 'madness' refers to this kind, as presumably most of the traditional paradigms are instances of psychiatric disorders. Probably there is no fact of the matter whether 'madness' got a determinate reference via the Ordinary natural kind convention or not, and the question is an indeterminate one.

Thus, the problematic results of the Truth convention do indeed get mirrored by problematic results of other implicit reference-fixing conventions.[5] Much as our reference-fixing conventions may initially seem to provide with referents the expressions that we intend the conventions to apply to, the conventions will on occasion be insufficient to fix referents for these expressions in a determinate way, and in other cases they will even lead to contradictions, whether *a priori* or in an empirical fashion. This was our second reason for thinking that the semantic paradoxes are paradoxes of reference fixing.

I think that yet a third, more indirect and somewhat weaker reason for holding this view can be derived from the fact that the most widely accredited, standard formal solutions of these paradoxes arguably rely in some way on the idea that the reference-

[5] See again my book Gómez-Torrente (2019), chs. 2 and 3, for discussion and examples of analogous failure of reference and indeterminacy examples that arise for implicit reference-fixing principles in the case of demonstratives and proper names.

fixing mechanisms for ‘true’ and other semantic paradoxical expressions work well in general, even if they turn out to fail in certain relatively localized cases—an idea that is surely part of the description of reference fixing and its failures. The exact delimitation of the problematic cases is certainly a disputed matter among the different theories, but undoubtedly the theories follow this abstract pattern. Tarski’s theory of truth makes completely explicit the central role of instances of the T-schema (the T-biconditionals) in the specification of the truth conditions of truth ascriptions. Tarski then proposes a hierarchy of truth predicates in such a way that meaningful truth ascriptions are restricted to ascriptions to sentences not containing the ascribed truth predicate nor any truth predicate of a higher level in the hierarchy. Truth ascriptions of other kinds are considered contentless. Kripke’s theory of truth relaxes the Tarskian restriction on T-biconditionals, noting for example that in some natural non-paradoxical cases two speakers can make truth attributions about each other’s sets of claims. This cannot be straightforwardly accommodated within the Tarskian hierarchy of truth predicates, and Kripke proposes a theory with one unique truth predicate. Nevertheless, contentful truth ascriptions, and thus truth ascriptions susceptible of getting their truth conditions fixed by T-biconditionals, are delimited in another way: they are restricted to those where the sentence to which truth is ascribed is grounded: for them, truth conditions can be established, without circularity or infinite regress, and ultimately traced to statements not containing semantic vocabulary. In this way, both the Tarskian theory and the Kripkean theory fit a certain abstract pattern that one might expect a theory of the paradoxes to fit if it bases its solution, whether explicitly or implicitly, on a view about the fixing of the content of ‘true’.[6]

1.2. Collection paradoxes and reference fixing.

Let’s turn now to the paradoxes about the notion of collection, class or set, taking Russell’s paradox as our first example. It would seem that we can, if we please, name ‘R’ the collection formed exactly by

[6] Other accredited though arguably less standard theories that fit this pattern include those in Parsons (1974), McGee (1990) and Gupta and Belnap (1993).

the collections that do not belong to themselves, in such a way that for an arbitrary object x,

> x belongs to R if and only if x is a collection that does not belong to x.

But it turns out that there is no such thing as R so characterized, for an instance of this condition for membership in R is the sentence

> R belongs to R if and only if R is a collection that does not belong to R,

which contradicts the supposition that R is a collection. Note that, in fact, any substitution for 'P' and 'Q' in *~∃R(Q(R) & ∀x(P(x,R) ≡ Q(x) & ~P(x,x)))* gives a first-order logical truth, so that the postulation of the existence of R contradicts a truth of pure logic.

One key tacit assumption in this reasoning is the assumption that given any properties expressible by predicates P, Q, there exists, and we can name as we please, the collection of things having those properties, or to which the predicates P and Q apply. A version of this assumption that suffices to develop the Russellian reasoning above is the so-called "Comprehension convention" involving the so-called "Comprehension-schema" (or C-schema):

> *Comprehension convention:* All instances of the following schema are assertable/ usable in reasoning/ true:
>
> (C-schema) ∃!S(S is a collection & ∀x(x belongs to S ≡ Q(x) & P(x))),
>
> where in place of 'P' and 'Q' go predicates.

The instance of the C-schema implicitly used in the Russellian reasoning above is, of course,

> ∃!S(S is a collection & ∀x(x belongs to S ≡ x is a collection and x does not belong to x)).

It is the initial belief in the existence of a unique S with these characteristics that leads to a (in the presence of that belief) legitimate introduction of the name 'R' for such a unique S.

As a second example of collection paradox, take the Cantorian paradox of the universal set. It would seem that we can, if we please, name 'U' the collection formed exactly by all the collections, in such a way that for an arbitrary object x,

> x belongs to U if and only if x is a collection,

(here the predicate 'is a collection' replaces both 'P' and 'Q' in the schema, so that we can avoid the repetition), or so an appeal to the (repetition-free) instance of the C-schema

> ∃!S(S is a collection & ∀x(x belongs to S ≡ x is a collection))

leads us to think. But it turns out that there is no such thing as U so characterized, for Cantor's theorem would imply that the power set of U would have more elements than U, and so collections not in U, contradicting the logical implication of the above instance of the C-schema that all collections are elements of U. Here the non-existence of U is not a matter of pure logic, but it follows from the supposition of the mentioned instance of the C-schema plus some logic and set-theoretic mathematics.

The same holds for our final example of collection paradox, the Burali-Forti paradox. It would seem that we can, if we please, name 'Ω' the collection formed exactly by all the ordinals, in such a way that for an arbitrary object x,

> x belongs to Ω if and only if x is an ordinal,

and certainly the introduction of 'Ω' is made legitimate by an appeal to the instance of the C-schema

> ∃!S(S is a collection & ∀x(x belongs to S ≡ x is a collection and x is an ordinal))

(or just

> ∃!S(S is a collection & ∀x(x belongs to S ≡ x is an ordinal)),

given that all ordinals must be collections). But it turns out that there is no such thing as Ω so characterized, for theorems about the ordinals imply that Ω must be well-ordered if it exists and thus have an ordinal number, one which, however, must be outside Ω, contradicting the logical implication of the above instance of the C-schema that all ordinals are elements of Ω. Thus the non-existence of Ω is again not a matter of pure logic, but it follows from the supposition of the mentioned instance of the C-schema plus some logic and set-theoretic mathematics.

As with the semantic paradoxes, I see three reasons why it is plausible to see these (and other) collection paradoxes as paradoxes of reference fixing. The first reason, again in close parallel with the semantic paradoxes, is that, as we have seen, the collection paradoxes all make implicit use of (something like) the Comprehension convention involving the C-schema. This turns them into paradoxes of reference fixing because it's natural to think of the Comprehension convention as an implicit conventional principle purporting to give sufficient conditions for the constitution of collections referred to by names appropriately introduced by corresponding descriptions. The implicit idea is that it should do this since it gives the conditions under which an arbitrary object belongs to the set for which a name is being introduced, thus uniquely characterizing this set. But the second reason, again in parallel with the semantic paradoxes, is that the Comprehension convention arguably fails to achieve its reference-fixing purpose, at least fully, and that the way it fails mirrors the way purported reference-fixing conventions fail to achieve their purpose on certain occasions. In order to see that these two reasons are in fact such, let's describe in appropriate detail when and how the Comprehension convention and the C-schema succeed and fail.

First we must emphasize that the Comprehension convention does achieve its purpose at least in a vast number of cases. For example, given the instance of the C-schema

> ∃!S(S is a collection & ∀x(x belongs to S ≡ x is a person and x is a European head of state on January 1st, 2021)),

and given that we know that it is both in principle and as a matter of fact determinable whether the condition of being a person who is a European head of state on January 1st, 2021 is possessed or lacked by an object, the introduction of a name for the set of European heads of state on January 1st, 2021 is made legitimate; and given the instance of the C-schema

> ∃!S(S is a collection & ∀x(x belongs to S ≡ x is a star and x has gone supernova by January 1st, 2021)),

and given that we know that it is at least in principle determinable whether the condition of being a star that has gone supernova by January 1st, 2021 is possessed or lacked by an object, the introduction of a name for the set of stars that have gone supernova by January 1st, 2021 is made legitimate—even if the determination of whether the condition of being a star that has gone supernova by January 1st, 2021 is possessed or lacked by an object may not be feasible as a matter of fact.

But in some cases, the Comprehension convention doesn't work well. If in the case of 'true' the problems we know arose with ungrounded sentences, in the case of collection constitution the known problem is a related one of what we might also call "ungrounded" collections. For the paradoxes we have considered, the problem can be put by saying that, in the words of Gödel (1944, 454-5, 459), one seeks to introduce a collection containing members "involving" or "presupposing" this collection.[7] (But, in order to take

[7] Gödel distinguishes (essentially) two forms of Russell's "vicious-circle principle" (which Russell used to diagnose and treat the semantic and set-theoretic paradoxes): one like the one mentioned in the text, that Gödel sees as basically correct, and according to which no totality can contain members involving or presupposing this totality, in some appropriate sense of 'involving' and 'presupposing'; and one according to which no totality can contain members *definable* only in terms of this totality. This second, definability version of the vicious-circle principle is not so plausible and

account of "Yabloesque" paradoxes with the C-schema (see Goldstein (1994)), one ought to extend Gödel's idea of an "ungrounded" set so as to include sets that are given conditions of membership that exploit a vicious infinite regress.) Consider the condition of self-membership for collections:

> x is a collection that belongs to itself.

Applying the C-schema to this condition we get

> $\exists!$S(S is a collection & $\forall$x(x belongs to S $\equiv$ x is a collection and x belongs to itself)).

Suppose this collection exists and under that supposition call it A. Is it at least in principle determinable whether the condition that defines it is possessed or lacked by an object? In order for this to happen, the condition that defines it ought to be independent of the very question of whether the object at stake belongs to A. But when, assuming A exists, we consider A itself and ask ourselves whether it belongs to A, the instance of the C-schema that we have is evidently of no help in determining independently this matter, and not because of any lack of knowledge on our part: it merely implies that

> A belongs to A if and only if A is a collection and A belongs to A,

and this just doesn't determine whether A belongs to A in any independent, non-circular way, not even in the presence of some additional, relevant but possibly unknown facts. A, however, is not a paradoxical collection, in the sense that the supposition that it exists does not lead to contradiction.

The attempt to define and name the Russell set gives rise to a similarly failed instance of the C-schema, namely *$\exists!$S(S is a collection & $\forall$x(x belongs to S $\equiv$ x is a collection and x does not belong to x))*. Here the condition under which this instance of the C-schema says that

leads to a prohibition of all impredicative definitions that is problematic from the point of view of classical logic and mathematics.

an arbitrary object belongs to R is, at least in part, that x does not belong to x. But again this gives us no way of telling whether R belongs to R that is independent of the very question of (non-)membership in R: we are just given conditions for membership in R in terms of the idea of (non-)membership in that very set. Furthermore, the defectiveness of the condition makes possible the derivation of a logical contradiction from the supposition that R exists, made explicit above.

In the case of the Cantorian paradox of the universal set, the problem from the point of view of reference fixing is not much subtler. Here the failed instance of the C-schema is

> ∃!S(S is a collection & ∀x(x belongs to S ≡ x is a collection)).

To see why it fails, let's ask ourselves: Is it at least in principle determinable whether the condition of being a collection is possessed or lacked by an arbitrary object, in a way that is independent of the very question of whether the object at stake belongs to U? It is surely determinable for the vast majority of objects, but when we consider U itself and we ask ourselves whether it would be a collection, part of what we are implicitly asking is whether it would be determined, for an arbitrary object, whether it belongs to U or not; so whether U would belong to U or not is part of what would determine whether U is a collection or not: that questions of membership in U are determined goes hand in hand with the question of whether U is a collection. And then we have no way in which it could be determined whether U would belong to U that is independent of the condition of whether U belongs to U, not even in the presence of some additional, relevant but possibly unknown facts. Furthermore, the circularity again makes possible the *a priori* derivation of a contradiction from the supposition that U exists.

Much the same holds for the Burali-Forti paradox. The relevant instance of the C-schema, let's recall, is

> ∃!S(S is a collection & ∀x(x belongs to S ≡ x is an ordinal))

(given that all ordinals must be collections). Let's ask again: Is it at least in principle determinable whether the condition of being an ordinal is possessed or lacked by an arbitrary object, in a way that is independent of the very question of whether the object at stake belongs to Ω? It is surely determinable for the vast majority of objects, including typical ordinals: for a typical ordinal α, what determines that α is an ordinal is that the things in α form an initial segment of well-order types in the natural order, i.e that α is the first well-order type greater than all of an initial segment of well-order types, i.e the first to which all of an initial segment of well-order types belong without α being one of them—a condition that doesn't involve in a circular way the question whether α is in Ω. But when we consider Ω itself and we ask ourselves whether it would be an ordinal, part of what we are asking is thus whether Ω is an initial segment of order types without being one of them, i.e. without being one of the things that belong to Ω. So ultimately whether Ω would belong to Ω or not is part of what would determine whether Ω is an ordinal or not, and we have no way in which it could be determined whether Ω would belong to Ω that is independent of the condition of (not) belonging to Ω, not even in the presence of some additional, relevant but possibly unknown facts. Once more, the circularity makes possible the *a priori* derivation of a contradiction from the supposition that Ω exists.

Thus, paradoxical collections are not given appropriate constitution conditions by what, to all appearances, is the conventional principle providing the descriptions by means of which collection names are introduced. In this way, we lose one of the key assumptions that set the collection-paradoxical reasonings going, namely, that the described paradoxical collections exist. In fact, as we have seen, the paradoxical reasonings can be seen as *reductiones ad absurdum* of the implicit idea that the Comprehension convention determines that they exist.

The second reason why it's plausible to see the collection paradoxes as paradoxes of reference fixing was the idea that the wrong results of the Comprehension convention parallel the problematic results occasionally produced by other purported reference-fixing conventions. We discussed these parallels when we developed our second reason for the hypothesis as applied to the semantic paradoxes, and now we have seen how the situation is much the same with the Comprehension convention: it also gives both indeterminacy results (as in the case of the collection of self-membered collections A) and contradictory results (as in the paradox of the collections that do not belong to themselves and the other paradoxes). The argument for our second reason has thus already been laid out.

As in the case of the semantic paradoxes, I think that a third and weaker reason for the view that the collection paradoxes are paradoxes of reference fixing in our sense can be derived from the fact that the most standard theories of these paradoxes arguably rely in some way on the idea that the naïve reference-fixing mechanism for intended names of collections, i.e. the one based on the descriptions provided by the Comprehension convention involving the C-schema, works well in general, even if it turns out to fail in certain relatively localized cases. Thus, Russell's theory of types proposes a hierarchy of types of collections: the collections of individuals, the collections of collections of individuals, the collections of collections of collections of individuals, etc.; and it restricts comprehension to conditions defined over collections (or individuals) in types inferior to a given type. Other instances of comprehension are considered meaningless. Zermelo's set theory relaxes a bit the Russellian restriction, noting that in some natural non-paradoxical cases collections can contain elements from infinitely many types. This cannot be straightforwardly accommodated within a hierarchy of Russellian types, and Zermelo proposes a theory having one unique comprehension schema—and one single, untyped notion of membership. Nevertheless, contentful instances of comprehension are delimited in another way: they are restricted to those where the condition at stake is defined over an existing set—one whose existence is yielded by a succession of applications of the operations of set formation legislated by the set-theoretic axioms, starting with individuals and/or the empty set. Again, as in the semantic case, we have standard theories fitting a certain abstract pattern that theories of the collection paradoxes as paradoxes of reference fixing ought to fit.

It is thus not unreasonable, to say the least, to see in the abstract coincidences we have reviewed a correspondingly abstract diagnosis of the roots of paradox, at least for the main semantic and collection paradoxes. Certain procedures—the reliance on unrestricted T-biconditionals, or on unrestricted comprehension—are implicitly taken to determine truth conditions or referents for appropriate expressions—truth ascriptions, straightforward names for sets fixed by descriptions. It turns out that these procedures, while working well in many, or even a vast majority of cases, give problematic results in some—at least for truth ascriptions where the sentence to which truth is ascribed is ungrounded, and at least for some straightforward set-naming descriptions where the condition at stake is again "ungrounded" in a closely analogous sense. The functionality and even the intelligibility of the concepts of truth and collection is presumably to be explained at least in part by the non-problematic character of those reference-fixing procedures in a vast number of cases, and presumably in the central ones—the liar and truth-teller sentences, or the descriptions for Russell's collection or for the collection of all self-membered collections are presumably not central cases of the use of the concepts of truth and of collection. And yet the procedures do not fulfill perfectly the task assigned to them, as shown by the semantic and collection paradoxes. Can all these analogies be extended to the case of the sorites paradoxes?

1.3. The sorites paradoxes and reference fixing.

A typical sorites paradox arises as follows. The following set of sentences would all be taken as true in a typical situation involving the relevant people: (i) Alvar, now a nine-year old boy, is young now; (ii) Zellig, now a ninety-year old man, is not young now; and (iii) for all persons x and y, if x was born six months earlier than y and y is now young, x is also young now. (i) to (iii) may seem incontrovertible in typical situations, but, given that surely it is (empirically) true that (iv) there is an n and a series x_1, x_2,...,x_n of persons where Alvar is x_n, Zellig is x_1, and for every x_i ($i<n$), x_i was born six months earlier than x_{i+1}, (i) to (iii) (in view of (iv)) imply that (v) Zellig is young now, which contradicts (ii). Paradoxes similar

to this one for ‘young’ arise for most other predicates of natural language.

As it turns out, every paradox of this sort based on an empirical premise of the form of (iv) has an analog that requires only an *a priori* premise of the same form. Consider this: (i’) nine is an age in years whose possession makes a person young; (ii’) ninety is not an age in years whose possession makes a person young; (iii’) for all x, y, if y is an age in years whose possession makes a person young, and x is an age in years that is six months more than y, then x is an age in years whose possession makes a person young; (iv’) there is an n and a series x_1, x_2,...,x_n of ages in years where x_n is nine, x_1 is ninety, and for every x_i ($i<n$), x_i is an age in years that is six months more than x_{i+1}. Here (iv’) is an *a priori* truth, and together with (i’) to (iii’), it implies that (v’) ninety is an age in years whose possession makes a person young, which contradicts (ii’).

Now, if our diagnosis above is correct as a diagnosis of the roots of the main semantic and collection paradoxes, as I think it is, one might expect it to provide at least a prima facie sensible blueprint for a diagnosis of the sorites paradoxes as paradoxes of reference fixing. However, unlike in the case of the semantic and collection paradoxes, it is clear that the most standard theories of vagueness and the sorites are incompatible with any diagnosis of this kind, and so our third, weaker reason for the hypothesis in those cases cannot be given in the case of the sorites. To begin with, the most popular theories, the varieties of supervaluationism (as in Fine (1975)) and degree theories (as in Machina (1976)), postulate that sorites-susceptible concepts work only in tandem with a non-classical semantics for the logical expressions. Their diagnosis of what goes wrong in the sorites reasoning does not appeal to unforeseen necessary restrictions in the implicitly used mechanisms of content fixing, but to somewhat mysterious adaptations of the logic of natural language when vagueness is concerned (and which make relevant sentences of the form of (iii) untrue by changing the semantics of the quantifiers and connectives). On the other hand, when one focuses on standard theories that don’t abandon the classical semantics for the logical expressions, one never gets a diagnosis of our kind either. Epistemicists (such as Sorensen (1988) and Williamson (1994)) and most contextualists (such as Soames

(1999) and Fara (2000)) are led to postulate that the content-fixing mechanisms for sorites-susceptible concepts work well without restrictions (whether in a context-insensitive or in a context-sensitive way), making sentences of the form of (iii) classically false, but in some mysterious fashion that doesn't allow any normal speaker (or theorist) to know what are the precise cut-off points that bear witness to the falsehood of those universal quantifications. (There are also non-standard theories that postulate that those mechanisms just turn out not to work well at all: weak and strong nihilisms—theories that claim, respectively, that vague predicates have either a universal or an empty extension (as in Unger (1979)), or are just meaningless (as in Ludwig and Ray (2002) or Braun and Sider (2007)).)

Even if the most standard theories of vagueness and the sorites don't align themselves with the most widely accredited, standard theories of the semantic and collection paradoxes, our discussion in the preceding subsections suggests that a sensible theory of vagueness that did align itself with these theories might be desirable: given the hypothesis that the main semantic and collection paradoxes have their root in the localized failure of a content- or reference-fixing procedure, a failure that takes place only in a relatively restricted number of cases, it would be desirable at the very least to have available for consideration a theory of vagueness and the sorites that shared this feature.

Theories of this kind, however non-standard, do exist. According to a theory I have myself recommended elsewhere (see Gómez-Torrente (2010) and (forthcoming), which also contain discussion of kindred recent theories[8]), many predicates, especially degree adjectives like 'young', have associated with them a reference- or extension-fixing convention that relies on the contextual acceptability of sentences of the forms of (i) to (iii) above. To use a self-explanatory terminology from those works, call (i) a "positive paradigm" sentence for 'young', (ii) a "negative paradigm" sentence for 'young', and (iii) a "tolerance" sentence for 'young'. In essence, the convention, that appears to be embraced at least for degree adjectives like 'young' (and for derived predicates

[8] Some of these theories can be found in Manor (2006), Gaifman (2010), Pagin (2011), van Rooij (2011) and Rayo (2013).

like ‘is an age in years whose possession makes a person young’), can be stated in schematic form as follows:

> *Tolerance convention.* In a context of use of a degree adjective P, the extension of P is constituted by the things relevant in the context to which P is implied to apply by the contextually accepted positive paradigm sentences for P and the contextually accepted tolerance sentences for P. And the complement of the extension of P is constituted by the things relevant in the context to which the negation of P is implied to apply by the contextually accepted negative paradigm sentences for P and the contextually accepted tolerance sentences for P.

Said in a less precise but perhaps more intuitive way: the things to which a degree adjective P applies in context are the things (contextually) close in their P-ness to our paradigms for P, and the things to which “not-P” applies are the things (contextually) close in their non-P-ness to our paradigms for not-P. Arguably the Tolerance convention is an implicit content fixer underlying many (if not all) uses of degree adjectives. (For a critical comparison of this convention with the convention proposed by typical semantic theories of degree adjectives, which postulate idealized and unrealistic contextual “standards” or “norms” separating the extension of, say, “young” from its anti-extension, see Gómez-Torrente (forthcoming).)

The Tolerance convention yields very wrong results in cases where the things relevant in the context are such that they form a sorites chain for the predicate P, as in the examples above (recall (iv) and (iv’)), “linking” via tolerance sentences accepted in the context the positive paradigms and the negative paradigms. In such a case the convention implies that the extension of the predicate is just the universe of relevant things, and also that its complement is the very same universe. For any relevant object we care to focus on, the convention implies that the predicate P applies to it and also that the negation of P applies to it. I call such contexts *paradoxical.* One example would be a context in which all living people are relevant, and (i), (ii) and (iii) are accepted—(iv) is just plain empirically true here. Another would be an example in which all numbers are relevant, and (i'), (ii') and (iii') are accepted—(iv') is an always available *a priori* truth. In such contexts the sorites-susceptible predicate is not really assigned a referential content by the Tolerance convention, and typical sentences containing it presumably lack truth conditions. Paradoxical contexts are the analogs for the Tolerance convention of the paradoxical applications of the T-schema and the C-schema.

However, in many contexts—I would say in a vast majority—the Tolerance convention yields just what we would expect and wish. Many contexts are *contrastive*, in the sense that the things relevant in the context are appropriately few and form two clearly contrasted complementary sets. Generally, in these contexts these sets are obtainable via the Tolerance convention: one subset of the relevant things are "linked" via tolerance sentences accepted in the context to the positive paradigms, and another, disjoint and complementary subset, are "linked" via tolerance sentences accepted in the context to the negative paradigms. I call these contexts *regular*. One example of the simplest kind would be a context in which just Alvar and Zellig are relevant and (i), (ii) and (iii) are accepted—here nothing of the form of (iv) can intervene, as there is no appropriate sorites chain of people among the relevant things, even if (iii) and other tolerance sentences are accepted. (Another example would be a very simple context in which just nine and ninety are relevant and (i'), (ii') and (iii') are accepted—here nothing of the form of (iv') can intervene, as there is no appropriate sorites chain of numbers connecting nine and ninety, even if (iii') and other tolerance sentences are accepted.)[9] In these and other less simple contexts where one subset of the relevant things cluster around the positive paradigms and its complement cluster around the negative

[9] I cannot argue here that regular occasions of use provide a vast number of cases of use of sorites-susceptible predicates, and in fact the central ones. But in my mentioned earlier works I have argued that, while only a large empirical study can fully confirm this hypothesis, the hypothesis appears more than likely in view of the fact that a great number of uses of degree adjectives one comes across are essentially contrastive, and thus regular. In this way, there is an evidently strong case that the number of regular occasions of use of typical sorites-susceptible predicates is vast, and includes the central cases.

paradigms, the sorites-susceptible predicate is assigned a straightforward referential content, typical sentences containing it have normal extensional truth conditions, and intuitively true paradigm *and* tolerance sentences are simply true. Regular contexts are the analogs for the Tolerance convention of successful applications of the Truth and Comprehension conventions.

Also crucially, on the theory paradoxical contexts are only one kind of *irregular* contexts, the other being contexts where the things relevant in the context are appropriately few, but such that the subset of them that are "linked" via tolerance sentences accepted in the context to the positive paradigms, and the disjoint subset of them that are "linked" via tolerance sentences accepted in the context to the negative paradigms are not jointly exhaustive of the relevant things: a set of "isolated" things far from the extremes also exists in the context. These irregular but non-paradoxical contexts are thus contexts where the relevant predicate doesn't get a classical extension/anti-extension pair, and thus fails to get a classical referent, even if they don't lead to contradictions. Such contexts are the analogs for the Tolerance convention of the truth-teller and the collection of all self-membered collections, indeterminate but innocuous from a classical point of view.

I call this theory the "dual picture" of vagueness and the sorites, because of its postulation of a duality of kinds of contexts that distinguishes contexts where sorites-susceptible predicates get referents or extensions from contexts where they don't, including the paradoxical contexts. If our hypothesis concerning the roots of paradox for the semantic and the collection contradictions is compelling, this ought to give an impulse to the dual picture, for the picture certainly shares the relevant features with standard theories of the semantic and collection paradoxes.[10] To begin with, it is based on a particular description of a mechanism of reference fixing for typical sorites-susceptible predicates, postulated to underlie uses of these predicates. Furthermore, the picture describes how some occasions of use of these predicates will be fully successful from a

[10] In fact, I take this in itself to be an argument for the dual picture of vagueness and the sorites. For consideration of more arguments for and against the dual picture, see again Gómez-Torrente (2010) and (forthcoming).

referential point of view, and other such occasions of use will be either non-paradoxical but somehow irregular, or plain paradoxical—when the universe of discourse in the occasion of use provides a relevant sorites series. Given all this, it's clear that the dual picture carries along with it the two main reasons for seeing the sorites paradoxes as paradoxes of reference fixing, whose analogs we gave in the case of the semantic and collection paradoxes.

An important step in the diagnosis of the semantic and collection paradoxes was the idea that collections, as well as the truth conditions for truth ascriptions, had to be built up in a "grounded" way, without circularity or infinite regress—circularity or infinite regress were one component of what was wrong with the relevant instances of the T-schema and the C-schema, and made possible the inconsistency results that would not be possible without them. Ungroundedness, however, is not a component of a sorites paradox, and this is thus one respect in which semantic paradoxes are analogous to collection paradoxes but in which neither of them is analogous to sorites paradoxes. If we want to employ a metaphor related to that of ungroundedness to describe what is going on in a sorites paradox according to the dual picture, we should think more of the damage that a misconceived enlargement or extension can cause to a structure than of the damage that can result from a badly planned foundation. We might say that the Tolerance convention is designed to build an extension/anti-extension pair for a predicate, which is analogous to building a pair of widely separated towers. The building proceeds upon a non-circular, non-regressive foundation and method—the paradigm sentences and the tolerance sentences—but under the tacit assumption that the two towers are not going to be united eventually, by something like a thick skybridge—the towers must be kept separated, their foundations are too far away for them to be susceptible of eventually becoming united in any way. The sorites paradoxes indicate how an extension of the method of construction beyond reasonable bounds will occasionally lead to a union of the towers, and to a collapse of the whole structure.

Another relevant respect in which the sorites paradoxes are not entirely analogous to the semantic and collection paradoxes lies in the fact that typical sorites paradoxes do not establish from purely *a* *priori* premises that an appropriate extension/anti-extension pair is impossible in paradoxical contexts: in typical paradoxical contexts, an empirical premise like (iv) is required. However, since every such paradoxical context is associated with a similar paradoxical context in which only an *a priori* premise such as (iv') is used in the *a priori* derivation of an inconsistency result from paradigm and tolerance sentences, there is a clear sense in which sorites paradoxes do not relevantly arise from the fact that their premises presuppose or imply an empirical claim that contradicts an empirical truth. The semantic, collection and sorites paradoxes do all essentially emerge *a priori* from suitable problematic instances of the relevant reference-fixing principles.

Perhaps our hypothesis about the paradoxes and reference fixing will appear immediately more attractive about the semantic and collection paradoxes than about the sorites paradoxes *because* of the very fact that the dual picture and kindred theories are only very recent and non-standard, which might throw doubts on their claim to truth—how could a picture of this kind not have been developed earlier, if something of this sort is right? (Perhaps our third reason for our hypothesis in the case of the semantic and collection paradoxes was not so weak after all.) All this may well be so, but I suspect that there are more or less clear causes for the differences in theoretical situation between the semantic and collection paradoxes, on the one hand, and the sorites paradoxes, on the other, and that these causes have little to do with the ultimate substantive truth about these topics, even if they may be eventually responsible for any prejudice that the sorites paradoxes do not arise from something like the failed applications of an implicit Tolerance convention.

One of these causes lies in the fact that, in the semantic and set-theoretical cases, there was, from early in the twentieth century, a scientific need to use the concepts of truth and set in logic and mathematics, which gave a natural impulse to projects that sought to isolate well-behaved instances of the T-schema and the C-schema that would serve that scientific need. These projects had implicitly in their core, so to speak, the idea that the Truth and Comprehension conventions were partially successful and partially unsuccessful reference-fixing procedures. On the other hand, in the case of the sorites paradoxes, there was no such scientific need to isolate well-behaved instances of the Tolerance convention, basically because science deals almost without exception with what we have called regular contexts of use of sorites-susceptible predicates (and often deals with non-sorites-susceptible predicates). And, in any case, if the dual picture is right, ordinary use of sorites-susceptible predicates will also most often be unproblematic both from a referential point of view and from the point of view of reasoning, so not even here a need to isolate well-behaved instances of the Tolerance convention is likely to be felt. (In a way, the dual view predicts that people will not feel a need to develop it.)

Theories of the sorites from the late twentieth century on, by contrast with standard theories of sets and truth, most often do not respond to a need to isolate consistent principles or contexts for the use of sorites-paradoxical predicates in scientific pursuits, or even to a need to codify consistent principles that describe or could regulate ordinary reasoning with sorites-susceptible predicates. Rather, those thories often seem to respond to a desire to provide philosophically or logically conceivable ways in which tolerance sentences, generally seen as the culprits in sorites paradoxes, might be taken to be false or untrue. I think that many will agree with me that this gives the resulting standard theories of the sorites paradoxes a strong air of philosophical or logical artificiality, derived from their postulation of unknown or unintuitive mechanisms for vague predicates to work referentially and in reasoning. The dual picture, by contrast, places the sorites paradoxes on a comparatively boring terrain alongside the semantic and collection paradoxes, and postulates no comparable mysteries: it claims that the Tolerance convention typically works just fine and that tolerance sentences are typically simply true. This may take away some of the philosophical and logical fun, but I strongly conjecture that it must be closer to the truth than the nowadays standard theories of the sorites.[11]

[11] It may be worth noting that epistemicist and nihilist views about truth or sethood, as well as theories of truth and sets that appeal to a non-classical semantics or logic for the logical expressions, are far less popular in these domains than they are in the case of the sorites. (Some of these theories may have no actual defenders.) Presumably some of the reasons for this are related to the causes for the situation described in the text. I myself would think that these views are just as implausible in the semantic and set-theoretic case as they are in the sorites case, though in the latter this is

2. Consequences and non-consequences.

2.1. Unification.

The most obvious consequence of our hypothesis is that it provides a unified account of the semantic, collection and sorites paradoxes, in the sense that it locates a common root for all of them (even if they are of course not fully analogous in all philosophically interesting respects). I take this to be a good consequence of the hypothesis, because unification ought to be expected when several phenomena have at least some initial "air" of similarity, and ought to be desired when it can help explain common features of those phenomena. In the next subsection, 2.2, we will indicate how our hypothesis can help explain one basic and important feature common to the three kinds of paradoxes we have been concerned with, what we might call their "recalcitrance", the very strong feeling that the premises leading to the paradoxes "ought to be true", thus providing an indirect argument for unification as well. In the present subsection we will contrast in a critical way our hypothesis, and the sort of unification it provides, with two other unified accounts of the sorites paradoxes and other paradoxes. While views that seek to give unified accounts of the sorites paradoxes and other paradoxes

obscured, again in no small measure due to the factors described in the text. (To be a bit more explicit about epistemicist, nihilist and other non-standard views about truth and sethood: an epistemicist about truth would say that a liar sentence is not only completely meaningful, but has determinate classical truth conditions provided by whatever it is that fixes the content of truth ascriptions, only we don't or even can't know which conditions those are; an epistemicist about sethood would say that "the set of all sets that do not belong to themselves" has a determinate classical set as referent provided by whatever it is that fixes the content of straightforward set-naming descriptions, only we don't or even can't know which set that is. Weak nihilists about truth would say that nothing is true, weak nihilists about sethood that nothing is a set. Strong nihilism about truth (apparently the position in Armour-Garb and Unger (2017)) would say that no statement involving truth is meaningful, strong nihilism about sethood that every statement involving sethood is meaningless. Views of sethood and truth that appeal to a non-classical semantics or logic include those in Brady (2006) and Priest (2002).)

have been comparatively infrequent, there are by now a few examples of such views.

The first one that we will comment on can be found in Priest (2010). This extends to the sorites the idea, from Priest (1994), that the semantic and set-theoretic paradoxes all fit a certain schema called Inclosure:

> *Inclosure schema.* 1. There is a set S such that $S = \{x : \phi(x)\}$, and θ (S) (Existence)
> 2. If $X \subseteq S$ and θ (X),
> (a) $\delta(X) \notin X$ (Transcendence)
> (b) $\delta(X) \in S$ (Closure) (see e.g. Priest (2010), 70),

where the instances have predicates in place of 'ϕ' and 'θ' and the name of a function in place of 'δ'. The paradoxes arise when some instances of this kind strongly appear to be true.

To see an example of how this works, here is how Priest explains that the sorites paradoxes fit the Inclosure schema:

> In a sorites paradox there is a sequence of objects, $a_0,...,a_n$, and a vague predicate P such that Pa_0, and $\neg Pa_n$; but for successive members of the sequence there is very little difference between them with respect to their P-ness, so that if one satisfies P, so does the other—the principle of tolerance. For the Inclosure Schema, let $\phi(x)$ be Px, so $S = \{x : Px\}$; $\theta(X)$ is the vacuous condition. S is a subset of $A = \{a_0,...,a_n\}$—indeed, a proper subset, since a_n is not in it—and so we have Existence. If $X \subseteq S$ then, since X is a proper subset of A, there must be a first member of A not in it. Let this be $\delta(X)$. By definition, $\delta(X) \notin X$. So we have Transcendence. Now, either $\delta(X) = a_0$ (if $X = \varnothing$), and so $P\delta(X)$, or (if $X \neq \varnothing$) $\delta(X)$ comes immediately after something in $X \subseteq S$, so $P\delta(X)$, by tolerance. In either case, $\delta(X) \in S$, so we have Closure. The inclosure contradiction is of the form $\delta(S) \notin S \wedge \delta(X) \in S$. (Priest (2010), 70-1)

Now, it seems dubious to me that the idea that the paradoxes fit the Inclosure schema can be of much explanatory value. For one thing, the idea itself doesn't suggest any explanation of why the relevant instances of Existence, Transcendence and Closure "appear to be true" to us in the cases of the relevant semantic, collection, and sorites paradoxes, nor does Priest attempt to provide one, as far as I can see. And I take this to be the main puzzling aspect that a philosophical theory of a paradox, and even more a theory that unifies it with other paradoxes, ought to explain. But furthermore, and more importantly, it seems just too easy, in fact trivial, for a paradox to fit the Inclosure schema. After all, a paradox is just an argument that uses premises from a set S of (apparently) true (ϕ) sentences of some kind (θ), and which develops from those premises (and possibly some auxiliary ones) a peculiar conclusion ($\delta(S)$) that appears to be true as well ($\delta(S) \in S$) and that also appears not to be true ($\delta(S) \notin S$). Note that if a paradoxical argument does this, then it will also trivially fit the full part 2 of the Inclosure schema, which involves arbitrary subsets of S: if $X \subseteq S$ and $\theta(X)$, then we can take $\delta(X)=\delta(S)$ for both Transcendence and Closure. (In fact, the way Priest says that the Liar verifies the Inclosure schema proceeds in just this way.) All the ways in which Priest fits paradoxes into the Inclosure schema can be reduced to this structure, and his explanation for the sorites is no exception: here S is the set of sentences of the form Pa_i that ought to be true, and, given that there must be a k such that Pa_k is true and Pa_{k+1} is false, the peculiar conclusion $\delta(S)$ is the sentence Pa_{k+1}, which ought not to be true and yet also ought to be true (by tolerance).

I thus submit that the Inclosure schema gives the form of our paradoxes only in the trivial sense that it gives a form that any paradox must have. If so, then not much explanatory value can be expected to come out of it. Presumably the puzzling aspects of Zeno's paradoxes of motion, or of the prediction paradox,[12] do not

[12] This is the paradox that the following *a priori* refutation of determinism seems too easy, or just contradictory under the assumption, dear to many, that determinism is true: (a) if determinism is true, it is in principle possible to predict every future event; (b) if it is in principle possible to predict every future event, a person can in principle come to know a prediction about whether she will raise her hand within the ten minutes following her learning the prediction; (c) if a person can in principle come to know a prediction about whether she will raise her hand within the ten minutes following her learning the prediction, she can choose to falsify the prediction by doing the opposite; (d) if a person can choose to falsify a prediction by doing the

have the same root as either the sorites, the semantic, or the collection paradoxes. Their fitting the Inclosure schema can thus be of little help explaining those aspects, or the puzzling aspects of the sorites, the semantic, and the collection paradoxes. By contrast, as we will see in subsection 2.2, our hypothesis that these paradoxes are paradoxes of reference fixing can help explain at least a distinctive puzzling aspect they share, related precisely to the reason why the relevant premises do in all cases appear to us to be true.

The second way to unification that we will briefly comment on can be found, e.g., in Tappenden (1993) and Soames (1999) (McGee (1990) has a related view also), and sees a common aspect in the semantic and sorites paradoxes, leaving aside the collection paradoxes. (This leaving aside the collection paradoxes might be thought to be a weak point of the Soames-Tappenden view, but I will not push this line of criticism here.) The idea is that both the truth predicate and typical sorites-susceptible predicates are partially defined predicates that, in a given context, have a determinate extension and a determinate anti-extension which together do not exhaust the universe of contextually relevant things, but which can be extended as the context gets modified. I think something like this is indeed true for the truth predicate, as our remarks on the semantic paradoxes above may already have suggested: the picture is a common Kripkean one on which applications of the T-schema to 'true'-free sentences in a modified context yield the truth conditions of sentences with one occurrence of the truth predicate, applications of the T-schema to such sentences in a new context yield the truth conditions of sentences with two occurrences, etc., thus expanding progressively the extension and anti-extension of the truth predicate.

However, I don't think this can give more than a pleasing but falsified picture of vague predicates. In paradoxical contexts there is no such thing as a determinate extension and a determinate anti-extension for the relevant sorites-susceptible predicate; and there is also no determinate set of borderline cases for which the predicate is undefined. There are simply no precise boundaries for vague predicates, not between extension and anti-extension, not between extension and set of borderline cases or between set of borderline

opposite, it is not in principle possible to predict every future event; so (e) determinism is not true. (Compare Levin (1979), 258, and Clark, 2007, 164.)

cases and anti-extension; so if our theory postulates this, it's just wrong. Note that on the dual picture, by contrast, both in paradoxical contexts and in non-paradoxical irregular contexts there is simply no classical (mutually exclusive and jointly exhaustive) extension/anti-extension pair (nor set of borderline cases), while in regular contexts the extension and anti-extension form a classical pair of sets that are mutually exclusive and jointly exhaustive of the universe of relevant things.

It is also hard to see in what way the picture of 'true' and vague predicates as partially defined could account for the appearance of truth of the premises used in the derivation of the semantic and sorites paradoxes. If vague predicates are indeed partially defined in the mentioned determinate way, then, much as in epistemicist views of the sorites, there must be sharp cut-off points for vague predicates, though in the Soames-Tappenden case these must separate the extension from the set of borderline cases as well as the set of borderline cases from the anti-extension. And if so, this must happen via some mysterious mechanism of which typical speakers fond of tolerance intuitions have no understanding. Perhaps Soames or Tappenden might want to appeal to some of the strategies of epistemicists in order to explain our misguided tolerance intuitions (I'm thinking especially of Williamson's (1994) appeal to "margin for error principles"), but then, in any case, it's hard to see how the same strategies could be applied in order to explain the appearance of truth of instances of the T-schema.[13]

2.2. An explanation of "recalcitrance".

By contrast with the proposed unifications of Priest and Soames-Tappenden, our hypothesis can provide a certain explanation of the fact that the premises used in the derivation of the semantic, collection and sorites paradoxes, by contrast with the premises used in other paradoxes (such as Zeno's paradoxes of motion or the prediction paradox), all appear true to us in a very strong way—so

[13] Weak or strong nihilism about truth and/or sethood and/or sorites-susceptibility (see e.g. the above-cited Armour-Garb and Unger (2017)) provides another way of unifying (some of) the paradoxes we have been concerned with. The unattractive features of these positions are probably too well known to require a new exposition here.

strong that even in view of paradox almost all of us seem to retain a "this has to be true" feeling. There are two reasons for this, both relatively simple.

The first is that many, in fact a vast majority of instances of the T-schema and the C-schema, and many paradigm and tolerance sentences in contexts to which the Tolerance convention applies, *are* unproblematically true; so this may make it difficult for us to get a grasp of how other instances or the same sentences in other contexts are problematic and in fact truth-valueless.

The second, more important reason why the instances of the T-schema and the C-schema and the appropriate paradigm and tolerance sentences for sorites-susceptible predicates all seem so strongly true is that we implicitly view them as reference fixers, and we just have the feeling that if we want to use words in a certain fashion, virtually nothing could get in the way of that.

To be sure, if this is our implicit thought, then we are badly mistaken. For, even if normally we can make words *mean* whatever we want them to mean, it's not true that we can make them *refer* by our mere whim, let alone make them refer to whatever we would like them to refer to. Meaning and reference are different beasts, and this is one of the ways in which their differences are manifested. Part of the reason for the frequent temptation to conflate them presumably arises from the fact that reference-fixing principles are adopted by some kind of implicit convention. But this does not turn them into lexical definitions, incorrigible assignments of meaning to our expressions; rather, they are, and they are implicitly recognized to be, fallible attempts at providing our words with a referential content, however weak or difficult to make explicit our recognition of this may be.

Well, we all (or at least the sensible ones among us) will quickly and not weakly acknowledge that if an application of one of our reference-fixing conventions presupposes or implies something that has turned out to contradict an empirical truth (as in the case of 'miasma'), then the convention ought not to have been applied in that case. But in the semantic, collection and sorites paradoxes our reference-fixing conventions lead to contradictions that are essentially independent of the empirical truth of the matter, and, ironic though this may be, this independence from empirical facts probably makes it harder for us to recognize the paradoxical arguments as pure and simple failure of reference results. Perhaps in the back of our minds something tells us *Do logic, mathematics, and other kinds of alleged* a priori *reasoning really establish facts as hard and undeniable as the empirical facts?*

This points also to a difference not only between our paradoxes and other failures of reference, but to one between our paradoxes and other paradoxes which don't look like paradoxes of reference fixing but which may be seen (and are seen by some) as suggesting that certain claims that one might have thought of as *a priori* true, must in fact be empirically false. In the case of Zeno's paradoxes, for example, many think that what solves them is the rejection of the idea (wrongly thought to be *a priori*) that an infinite series of tasks cannot be accomplished in a finite determinate amount of time. In the case of the prediction paradox, those who have considered it appear to think that what solves it is the rejection of the idea (wrongly thought to be *a priori*) that if determinism were true, then predictions for all future events could in principle be given. Nothing like this route seems likely to become available for the semantic, collection and sorites paradoxes.

Of course, a strong, even "recalcitrant", appearance of truth of the premises in a paradox need not imply that some problem of reference fixing is behind that appearance. All that can be claimed is that applications of reference fixing principles, including paradoxical applications, will typically have a strong appearance of truth due to the reasons we have described. Other difficult and recalcitrant paradoxes need not arise as paradoxes of reference fixing.

That the explanation of the "recalcitrance" of the semantic, collection, and sorites paradoxes has to do with an intuition, that we ought to be able to make our words mean and refer to whatever we want them to mean or refer to, looks like a reasonable thing to believe. But it cannot be overemphasized that, reasonable though it appears, it's to say the least unclear how several theories of the paradoxes could entail something like this reasonable claim, including the unificatory theories discussed in the preceding subsection. The claim, however, follows straightforwardly from our hypothesis that the semantic, collection and sorites paradoxes are all paradoxes of reference fixing.

The claim that there is an explanation from our hypothesis for the "recalcitrance" of the semantic, collection and sorites paradoxes must not be conflated, of course, with the mistaken idea that the premises of these paradoxes, i.e. essentially the particular applications of the corresponding reference-fixing principles in the relevant cases, must after all be true and something else is the real culprit of the contradictions. My own view, as I have emphasized repeatedly, is that the applications of the Truth, Collection, and Tolerance conventions in the paradoxical cases is what is to blame for the contradictions, and that these establish that those applications cannot generate appropriate contents or referents in those cases.

2.3. Is there such a thing as "the" solution to the paradoxes?

The claim that our paradoxes are paradoxes of reference fixing might be thought to suggest that there can be no such thing as "the" solution to a paradox of this kind, much as it has often been hoped that a correct diagnosis of the paradoxes will point the way to "the" right solution for them. The idea would be that failures and indeterminacies of reference fixing may be patched (or disdained) in different, incompatible ways, as in general there are presumably no backup conventions as to what to do when the failures or indeterminacies come up. Thus, as we noted, although the predominant attitude toward 'miasma' nowadays seems to be as a term without reference, which is simply avoided, or which we will at best find in discussions of the old belief that there was a sickening substance underlying the foul odors of rotting materials (as we don't find relevant uses of 'Vulcan' save in discussions of the old belief that there was a planet with an orbit between Mercury and the Sun), there is also a less frequent and somewhat special use of 'miasma' as a purely descriptive term for foul odors. The situation with 'madness' appears to be similar: while some people disdain it as an obsolete term with no place in serious discussions (these people often disdain the notion of a psychiatric disorder as well), others see no problem using it as interchangeable with 'psychiatric disorder' as a term for an acceptable kind, and probably most people have no objection to 'madness' as a purely descriptive term for behaviors perceived as disordered. The thought would be that, analogously, there is nothing in the nature of things that prescribes a unique thing to do with 'true' (or 'false') as applied to problematic sentences, with names of paradoxical or otherwise ungrounded "collections", and with sorites-susceptible predicates as used in paradoxical or otherwise irregular contexts.

And the idea is that the thought would hold independently of the existence of a multitude of solutions of the paradoxes involving non-classical semantics or logics for the logical expressions—the thought would hold even within the space of what we might call "classical" solutions. So, for example, in the case of truth, one might view the existence of the Tarskian solution (which in effect banishes the problematic sentences and even 'true' as a unique predicate), the Kripkean solution (which keeps a unique truth predicate and assigns a truth value precisely to the grounded sentences of a given object language), the Gupta-Belnap solution (which declares some ungrounded non-paradoxical sentences true), and other solutions as a manifestation of this state of affairs. In the case of collections, one might think the same given the existence of Russellian type theory, Zermelian set theory, theories of non-well-founded sets, Bernays-Gödel set theory and its special-status names for proper classes, and other classical set theories. And in the case of vague predicates, solutions might range from theories that restrict meaningfulness to sorites-susceptible predicates as used in regular contexts (turning the dual theory into a prescriptive solution), through theories that (turning the contextualist wrong diagnosis into a prescriptive theory) prescribe in some way sharp cut-off points in paradoxical contexts between extension and set of indeterminate cases and between set of indeterminate cases and anti-extension for those predicates, to theories that (turning the epistemicist wrong diagnosis into a prescriptive theory) prescribe in some way sharp cut-off points in paradoxical contexts between extension and anti-extension for those predicates.

Attractive though the thought may be, I doubt that the hypothesis that our paradoxes are paradoxes of reference fixing necessarily implies it. I do think that if our perspective is a purely instrumentalist one, on which, in order for some theoretical construction to be an acceptable solution to a paradox it is enough that that construction satisfies some inspecific practical or theoretical needs one has, then there is certainly no such thing as "the" solution to a paradox, including our paradoxes. Surely any practical or theoretical requirement may be felt as a desideratum by someone or other. But an instrumentalist perspective need not be the only perspective compatible with our hypothesis.

In particular, we may have reasons to believe that a reference-fixing convention, that gives rise to indeterminacies or contradictions in some cases, nevertheless is somehow tracking a privileged kind of things. If so, one might argue that a solution to a paradox generated by that convention that would have a privileged status of some sort would be one that generated as referents for the relevant expressions precisely the things of that kind. For example, we might have reason to believe that there is a privileged notion of truth as correspondence that is naturally understood by means of the T-schema as restricted to the grounded sentences of the language we are interested in; if so, and if some technically useful construction declares as true precisely the grounded sentences determined to be true by the corresponding instances of the T-schema and the relevant facts, it will have a non-instrumental advantage over constructions that leave out some intuitively true grounded sentences, or over constructions that declare true some non-paradoxical ungrounded sentences, for example. That this might be the real situation is not excluded by our hypothesis in this paper. If the Truth convention doesn't work well in general, the possibility of reacting in a number of different ways as to our use of 'true' or of replacements for 'true' need not indicate that there is no sense in which there can be a privileged way of reacting.

Similar remarks hold for the problems with the Comprehension convention and the Tolerance convention. We might have reason to believe that the "grounded" sets created under some understanding of the cumulative hierarchy are precisely the existing sets, and that consequently the natural restriction on the C-schema is that it must operate on such sets when we want to use it to define and name a particular collection; if so, and if some technically useful construction declares as existing precisely those cumulatively created sets, and incorporates the ensuing restriction on the C-schema, it will have a non-instrumental advantage over constructions that leave out some cumulatively constructed sets, or over constructions that somehow name some non-paradoxical "ungrounded" collections or some (paradoxical) proper classes, for example. In the case of the Tolerance convention we might, for example, have reason to believe that reliance on tolerance principles forms a linguistically or psychologically essential component of the cases where a universe of relevant things is sorted out into two groups by means of a vague predicate; if so, there might be a non-instrumental advantage of a solution based on the dual theory over a solution based on epistemicism, for example. Again, that these might be the real situations in the case of the Comprehension and Tolerance conventions is not excluded by our hypothesis in this paper.

To sum up: If what I have argued in this paper is right, the hypothesis that the semantic, collection and sorites paradoxes are paradoxes of reference fixing is an idea strongly suggested by direct consideration of the paradoxes themselves, and one which provides a non-trivial and explanatory unification of them, but which is compatible with different attitudes toward the problem of finding solutions for the paradoxes. More substantive discussion of what might be the better attitude toward at least some of our paradoxes is something that must be left for another occasion.

References

Armour-Garb, B. (2017), "Introduction", in B. Armour-Garb (ed.), *Reflections on the Liar*, Oxford University Press, Oxford, 1–21.

Armour-Garb, B. and P. Unger (2017), "From no People to no Languages. A Nihilistic Response to the Liar Family of Semantic Paradoxes", in B. Armour-Garb (ed.), *Reflections on the Liar*, Oxford University Press, Oxford, 22–38.

Bacon, A. (2015), "Can the Classical Logician Avoid the Revenge Paradoxes?", *Philosophical Review* 124, 299–352.

Brady, R. (2006), *Universal Logic*, C.S.L.I., Stanford.

Braun, D. and T. Sider (2007), "Vague, so Untrue", *Noûs* 41, 133–156.

Chihara, C. (1979), "The Semantic Paradoxes: A Diagnostic Investigation", *Philosophical Review* 88, 590–618.

Clark, M. (2007), *Paradoxes from A to Z*, 2nd edn., Routledge, New York.

Cook, R. T. (2013), *Paradoxes*, Polity, Malden (Mass.).

Fara, D. G. (2000), "Shifting Sands: An Interest-Relative Theory of Vagueness", *Philosophical Topics* 28, 45–81. (Originally published under the name Delia Graff.)

Fine, K. (1975), "Vagueness, Truth and Logic', *Synthèse* 30, 265–300.

Gaifman, H. (2010), "Vagueness, Tolerance and Contextual Logic", *Synthèse* 174, 5–46.

Gödel, K. (1944), "Russell's Mathematical Logic", in P. Benacerraf and H. Putnam (eds.), *Philosophy of Mathematics. Selected Readings*, 2nd edn., Cambridge University Press, Cambridge, 1983, 447–469.

Goldstein, L. (1994), "A Yabloesque Paradox in Set Theory", *Analysis* 54, 223–227.

Gómez-Torrente, M. (2010), "The Sorites, Linguistic Preconceptions, and the Dual Picture of Vagueness", in R. Dietz and S. Moruzzi (eds.), *Cuts and Clouds. Vagueness, its Nature and its Logic*, Oxford University Press, Oxford, 228–253.

Gómez-Torrente, M. (2019), *Roads to Reference*, Oxford University Press, Oxford.

Gómez-Torrente, M. (forthcoming), "The Sorites, Content Fixing, and the Roots of Paradox", in O. Bueno and A. Abasnezhad (eds.), *On the Sorites Paradox*, Springer, Cham, forthcoming.

Gupta, A., and N. Belnap (1993), *The Revision Theory of Truth*, M.I.T. Press, Cambridge (Mass.).

Herzberger, H. G. (1970), "Paradoxes of Grounding in Semantics", *Journal of Philosophy* 67, 145–167.

Kripke, S. A. (1975), "Outline of a Theory of Truth", *Journal of Philosophy* 72, 690–716.

Levin, M. (1979), *Metaphysics and the Mind-Body Problem*, Clarendon Press, Oxford.

Ludwig, K. and G. Ray (2002), "Vagueness and the Sorites Paradox", *Philosophical Perspectives* 16, 419–461.

Machina, K. (1976), "Truth, Belief and Vagueness', *Journal of Philosophical Logic* 5, 47–78.

Manor, R. (2006), "Solving the Heap", *Synthèse* 153, 171–186.

McGee, V. (1990), *Truth, Vagueness, and Paradox: An Essay on the Logic of Truth*, Hackett, Indianapolis.

Pagin, P. (2011), "Vagueness and Domain Restriction", in P. Égré and N. Klinedinst (eds.), *Vagueness and Language Use*, Palgrave Macmillan, New York, 283–307.

Parsons, C. (1974), "The Liar Paradox", *Journal of Philosophical Logic* 3, 381–412.

Priest, G. (1994), "The Structure of the Paradoxes of Self-Reference," *Mind* 103, 25–34.

Priest, G. (2002), *Beyond the Limits of Thought*, 2nd edn., Oxford University Press, Oxford.

Priest, G. (2010), "Inclosure, Vagueness, and Self-Reference", *Notre Dame Journal of Formal Logic* 51, 69–84.

Rayo, A. (2013), "A Plea for Semantic Localism", *Noûs* 47, 647–679.

Soames, S. (1999), *Understanding Truth*, Oxford University Press, Oxford.

Sorensen, R. (1988), *Blindspots*, Oxford University Press, Oxford.

Tappenden, J. (1993), "The Liar and Sorites Paradoxes: Toward a Unified Treatment", *Journal of Philosophy* 90, 551–577.

Unger, P. (1979), "There Are No Ordinary Things", *Synthèse* 41, 117–154.

van Rooij, R. (2011), "Implicit versus Explicit Comparatives", in P. Égré and N. Klinedinst (eds.), *Vagueness and Language Use*, Palgrave Macmillan, New York, 51–72.

Williamson, T. (1994), *Vagueness*, Routledge, London.