Two Paradoxes of Satisfaction

Peter Eldridge-Smith
Australian National University
peter.eldridge-smith@anu.edu.au

There are two paradoxes of satisfaction, and they are of different kinds. The classic satisfaction paradox is a version of Grelling's: does 'does not satisfy itself' satisfy itself? The Unsatisfied paradox finds a predicate, P, such that Px if and only if x does not satisfy that predicate: paradox results for any x. The two are intuitively different as their predicates have different paradoxical extensions. Analysis reduces each paradoxical argument to differing rule sets, wherein their respective pathologies lie. Having different pathologies, they are paradoxes of different kinds. Furthermore, each of these satisfaction paradoxes has an analogue with the same pathology in set theory. Therefore, these analogues are respectively of the same two kinds. This level of abstraction is significant in that it tracks two related but different pathologies. Thus, not all paradoxes of semantics and set theory share the same pathology: there are at least two kinds of paradox cutting across the semantic and set-theoretic distinction.

1. A new paradox of satisfaction

Here is an example of *the Unsatisfied* paradox or, more specifically, *the Unsatisfied Predicate* paradox:

> My favourite predicate just happens to be 'does not satisfy my favourite predicate'. So, Crete satisfies my favourite predicate if and only if Crete does not satisfy my favourite predicate. Therefore, Crete satisfies my favourite predicate and does not satisfy my favourite predicate.

Here is essentially the same paradox using 'true of':

> My second favourite predicate just happens to be 'my second favourite predicate is not true of'. So, my second favourite predicate is true of Ariadne iff my second favourite predicate is not true of Ariadne. Therefore, my second favourite predicate is true of Ariadne and not true of Ariadne.

These informal examples make tacit use of a principle like the following:

> *Satisfaction Principle*:
> If something satisfies a predicate, then that thing is as the predicate says; and conversely, if something is as a predicate says, then that thing satisfies the predicate.

Mind, Vol. 124 . 493 . January 2015
doi:10.1093/mind/fzu183 Advance Access publication 24 January 2015

In giving a formal representation of these examples, I employ the following introduction and elimination rules for the satisfaction predicate:

Satisfaction Introduction (SatI): $\mathrm{P}z \vdash z \text{ Sat } \langle \mathrm{P}x \rangle$

Satisfaction Elimination (SatE): $z \text{ Sat } \langle \mathrm{P}x \rangle \vdash \mathrm{P}z$

where '*z*' is replaceable by any term, 'Sat' is a two-place satisfaction predicate, and 'P*x*' is replaceable by any open sentence with one free variable, loosely speaking, here and throughout, a one-place predicate. The angle brackets represent a canonical name-forming device, like quotes, such that the expression named can be effectively recovered from the name.

From these, we can derive a satisfaction schema:

(Sat-schema): $z \text{ Sat } \langle \mathrm{P}x \rangle \leftrightarrow \mathrm{P}z$

And here is a derivation (U1) of our paradox:

(U1) The Unsatisfied paradox[1]

(1) $p = \langle \neg(x \text{ Sat } p) \rangle$	Premiss
(2) $t \text{ Sat } \langle \neg(x \text{ Sat } p) \rangle \leftrightarrow \neg(t \text{ Sat } p)$	Sat-schema
(3) $t \text{ Sat } p \leftrightarrow \neg(t \text{ Sat } p)$	(1), (2) =E (substitution of identicals)[2]
(4) $t \text{ Sat } p \; \& \; \neg(t \text{ Sat } p)$	SL (Sentential Logic)

Consider how this represents the first example. Let *p* abbreviate 'my favourite predicate'. The predicate I so favour is 'does not satisfy my favourite predicate'. It seems the premiss is true and the argument has the above apparently valid proof. It does not matter what *t* is, a contradiction results for any *t*.

There was, by the way, already a contradiction at line 3 in the above proof, so there was actually no need to go to line 4 in presenting a paradox. A contradiction is just the negation of a theorem, any theorem, so it need not have the form A & ¬A. Line 3 has the form A ↔ ¬A. Such

[1] I devised this satisfaction paradox by analogy with the Eubulidean Liar and presented it at the Australian National University in March 2003. I also presented it in a paper at the 2005 Australasian Association of Philosophy conference. That paper is published as Ch. 4 in Eldridge-Smith 2008.

[2] This rule will feature in discussion. It is also known as 'substitutivity of identity' and 'substitution of co-referentials'. It is even sometimes known as 'Leibniz's law', although that is arguably a second-order principle and arguably restricted to identity of indiscernibles; whereas the closest second-order principle is the principle of indiscernibility of identicals, which might also be at issue in versions of this paradox.

biconditionals are contradictions in many sentential logics, as is the natural language biconditional 'Crete satisfies my favourite predicate if and only if Crete does not satisfy my favourite predicate'.

In any case, conditionals are not required:

(U2)	The Unsatisfied paradox: another derivation	
(1)	$p = \langle \neg(x \text{ Sat } p)\rangle$	Premiss
(2)	$t \text{ Sat } p$	Assumption
(3)	$t \text{ Sat } \langle \neg(x \text{ Sat } p)\rangle$	(2), (1) =E
(4)	$\neg(t \text{ Sat } p)$	(3) SatE
(5)	$\neg(t \text{ Sat } p)$	(2), (4) ¬I [2]
(6)	$t \text{ Sat } \langle \neg(x \text{ Sat } p)\rangle$	(5) SatI
(7)	$t \text{ Sat } p$	(6), (1) =E
(8)	$t \text{ Sat } p \ \& \ \neg(t \text{ Sat } p)$	(7), (5) &I

This derivation uses the inference rule set {=E, SatE, SatI, ¬I, &I} to reach a contradiction. The previous derivation only varied in its sentential logic: it relied on the rule set {=E, SatE, SatI, →I, ↔I}. (The Sat-schema is a derived rule abbreviating use of the rule set {SatI, SatI, →I, ↔I}.)[3]

There are more versions of the Unsatisfied paradox that involve a greater number of argument places.

> My favourite relation is 'do not satisfy my favourite relation'. So, Crete and Ariadne satisfy my favourite relation iff they do not satisfy my favourite relation.

To formalize these we need general forms of SatI and SatE for predicates with a finite number of free variables:

$$\text{(SatI)}\ \mathrm{P}^n(z_1, z_2, \ldots, z_n) \vdash (z_1, z_2, \ldots, z_n)\ \text{Sat}^{n+1}\langle \mathrm{P}^n(x_1, x_2, \ldots, x_n)\rangle$$

$$\text{(SatE)}\ (z_1, z_2, \ldots, z_k)\ \text{Sat}^{k+1}\ \langle \mathrm{P}^n(x_1, x_2, \ldots, x_n)\rangle \vdash \mathrm{P}^n(z_1/x_1, z_2/x_2, \ldots)$$

where the superscripts indicate the adicity of each predicate.

[3] This Sat-schema is derived as follows:

(1)	$\mathrm{P}(z)$	Assumption
(2)	$z \text{ Sat } \langle \mathrm{P}x\rangle$	SatI
(3)	$\mathrm{P}(z) \to z \text{ Sat } \langle \mathrm{P}x\rangle$	(1) – (2) →I [1]
(4)	$z \text{ Sat } \langle \mathrm{P}x\rangle$	Assumption
(5)	$\mathrm{P}(z)$	SatE
(6)	$z \text{ Sat } \langle \mathrm{P}x\rangle \to \mathrm{P}(z)$	(4)–(5) →I [4]
(7)	$z \text{ Sat } \langle \mathrm{P}x\rangle \leftrightarrow \mathrm{P}(z)$	(3), (6) ↔I

(The need for formal conditions on the adicity (number of arguments) of the predicates involved is explained in Sect. 6.3.) Here, again, a general Sat-schema can be introduced as a derived rule:

(Sat-schema) $(z_1, z_2, \ldots, z_n)\ \mathrm{Sat}^{n+1}\langle \mathrm{P}^n(x_1, x_2, \ldots, x_n)\rangle \leftrightarrow \mathrm{P}^n(z_1, z_2, \ldots, z_n)$

And an Unsatisfied paradox can be derived in the same way as (U1) or (U2), so that for any two objects, it is paradoxical whether they satisfy a predicate such as p_h:

(U3) The Unsatisfied paradox using general Sat-schema

(1)	$p_h = \langle \neg(x, y \text{ Sat } p_h)\rangle$	Premiss
(2)	$a, b \text{ Sat } \langle \neg(x, y \text{ Sat } p_h)\rangle$ iff $\neg(a, b \text{ Sat } p_h)$	Sat-schema
(3)	$a, b \text{ Sat } p_h$ iff $\neg(a, b \text{ Sat } p_h)$	(1), (2) = E

In contrast to the Unsatisfied Predicate paradox, consider the following argument for a version of Grelling's paradox.

(G1) Grelling's satisfaction paradox

(1)	$r = \langle \neg(x \text{ Sat } x)\rangle$	Abbreviation
(2)	$r \text{ Sat } \langle \neg(x \text{ Sat } x)\rangle \leftrightarrow \neg(r \text{ Sat } r)$	Sat-schema
(3)	$r \text{ Sat } r \leftrightarrow \neg(r \text{ Sat } r)$	(1), (2) = E

Suppose Grelling's favourite predicate is 'does not satisfy itself'. Various expressions satisfy themselves, and others do not; but Grelling's favourite predicate satisfies itself iff it does not.[4]

As satisfaction is the converse of the truth relation, it is natural to think there may be versions of each paradox using that relation. Indeed, here is a derivation of my new paradox using 'true of'.

(U4) The Untrue paradox

(1)	$s = \langle s$ is not true of $x\rangle$	Premiss
(2)	$\langle s$ is not true of $x\rangle$ is true of t iff s is not true of t	'true of' schema
(3)	s is true of t iff s is not true of t	(1), (2) = E

This version can be contrasted with a version of Grelling's using 'true of'.

(G2) Grelling's paradox: a proof using 'true of'

(1)	$g = \langle x$ is not true of $x\rangle$	Abbreviation

[4] A similar version of Grelling's appears in Quine 1962, Priest 1994, and Visser 2004.

(2) ⟨x is not true of x⟩ is true of g iff g is not true of g — 'true of' schema

(3) g is true of g iff g is not true of g — (1), (2) = E

I will try to make the intuitive difference between the two paradoxes explicit in the next section.

2. The obvious difference

The Unsatisfied paradox can use any term to derive a contradiction, whereas Grelling's is restricted to particular cases. Take Athens, Theophrastus, the Sphinx, 'short', or 'long': it is paradoxical whether any of these objects satisfies the predicate of the Unsatisfied paradox or not. In contrast, none of those particular cases are paradoxical for Grelling's predicate. 'Is not true of itself' does not even apply to Athens, Theophrastus, or the Sphinx; 'short' is true of itself, and 'long' is not, but Grelling's paradox just arises in considering specific reflexive cases, in particular, whether 'is not true of itself' is true of itself or not.

3. There are two paradoxes of the truth relation

The obvious difference between the Unsatisfied paradox and Grelling's is sufficient to assure that there are two paradoxes of the truth relation. They are intuitively different, and we can articulate how they are different: their predicates have different paradoxical extensions. This extensional difference guides identification of versions and variations of each satisfaction paradox.

Over the next five sections I present an analysis of why there are two paradoxes of the truth relation. An adequate analysis should explain the obvious difference and determine whether these paradoxes are just different manifestations of the same underlying pathological reasoning.

4. Paradox, type of paradox, and kind of paradox

In this section I define some terms which will be used in what follows.

Philosophers describe a number of things as paradoxical, among them arguments, conundrums, and sentences or what they seem to express (statements, propositions, beliefs, etc.).[5]

[5] Philosophers also describe inconsistent sets of individually plausible propositions, and some stories or scenarios, as paradoxical. The Barber and Grandfather paradoxes, for example, are usually introduced as stories or scenarios.

A *sentence* or *statement* with an apparently *overdetermined* semantic status is a paradox (see Quine 1962, p. 1; Sainsbury 2009, p. 1). Most commonly such statements seem true and not true (but they can even seem possible and impossible, as in the statement of a paradoxical time travel scenario). So 'Buridan satisfies my favourite predicate' is a paradox given that the predicate I so favour is 'does not satisfy my favourite predicate'.

A seemingly sound *argument* to an apparently false conclusion is also a paradox. An argument, if it seems any good, and if it is more involved than the most basic immediate inference, will have multiple derivations that differ in their inferential steps. In this article, *proofs* may be more or less formal, or enthymematic; every step of a *derivation* after the premisses is a purely formal inference. (In Sect. 6.4, I will argue that a subset of these derivations, which I refer to as *normalized*, are suitable for localizing and comparing paradoxical pathologies.)

A *conundrum* poses a paradox if seemingly logical answers lead individually or collectively to paradoxical statements or arguments (see Sorensen 2003, p. 4).

Types of paradox can be distinguished for all sorts of reasons, but are not necessarily *kinds*. One might say that herbivores, carnivores, and omnivores are different types of animals but not kinds, whereas reptiles and mammals are kinds and not just types. Some dogs are short types and others tall types, beagles and Dalmatians are not only different types but different kinds. Beagles and Dalmatians do not exhaust the kinds of dogs, nor are they necessarily exclusive: there are hybrids.[6]

Families are a type of paradox distinguished by the concept or sort of concept used. There is a family of satisfaction paradoxes, which belong to an extended family of semantic paradoxes. Among the family of satisfaction paradoxes, I have shown there are two paradoxes distinguished by the paradoxical extension of their predicates. With respect to this difference, the Unsatisfied paradox, *qua* type of paradox, is an *insatiable* type of satisfaction paradox and Grelling's satisfaction paradox is a *reflexive* type. Whether these are of the same kind or of different kinds is an open question under analysis.

An instance of a paradox is a *version* and versions can be grouped by fine-grained types, which I term *variations*, as I do in section 7. For

[6] *If* semantic, set-theoretic, and Curry paradoxes were three kinds, then set-theoretic Curry paradoxes and semantic Curry paradoxes would be hybrids.

example, some versions of the Unsatisfied paradox can be derived with a single use of the SatE and SatI rules, while others require multiple uses of these rules. These groupings are variations.

Individuating *kinds of paradox* aims at capturing a common *pathology*. As a Principle of Pathological Reasoning, paradoxes with the same pathology are the same kind of paradox.[7] A method is needed that avoids begging questions by assuming a particular theoretical explanation for various paradoxes in order to classify them (see Badici 2008, Weber 2010a, Priest 2010). Prior to settling on a solution, some paradoxes can, nevertheless, be classified if their pathology can be localized. Several ways of localizing the pathology of paradoxes are considered in section 8.

I pursue one particular strategy in detail through sections 6 to 8.1: if some seemingly sound derivation of a contradiction does not require any premisses, then the pathology of the paradox with this derivation is in those inferences.

5. Exclusion of explicitly circular definitions and mere stipulations

This section is an interlude in which I exclude and set aside some similar contradictions. First, the Unsatisfied paradox is not a case of explicit circular definition. Secondly, there are some dubious stipulations, which can be set aside. And thirdly, I am only analysing Grelling's paradox as a paradox of satisfaction.

First, I add a rider about the Unsatisfied paradox as a type of paradox, namely, it is an *insatiable* type of satisfaction paradox *that does not explicitly appeal to a circular definition*. There are other types of insatiable contradictions if one allows circular definitions. Because it is understood that such definitions can lead to contradictions, such contradictions are unsurprising and not very paradoxical.[8]

Some systems explicitly allow circular definitions; sometimes they are used unwittingly; nevertheless, in presenting the Unsatisfied

[7] For comparison, fallacies with the same pathology are the same kind of fallacy. For example, fallacies due to using terms ambiguously are fallacies of equivocation. This is not to say paradoxes are sure to be fallacies; a paradoxical pathology might result in dialetheias (true contradictions).

[8] They might rank between 2 and 5 on Sainsbury's (2009, pp. 1–2) Richter scale of paradoxes, above the Barber at 1 but well below the Liar and Russell's at 10. The Unsatisfied and Grelling's paradoxes rank, I think, at the higher end of the scale.

paradox I refrained from courting contradiction by deliberately allowing circular definitions.

For example, the following is *not* a version of the Unsatisfied paradox:

(S#) An insatiable contradiction due to circular definition

(1) z Sat $\langle Px\rangle \leftrightarrow P(z)$	Sat-schema
(2) $\forall z(z$ Sat $\langle Px\rangle \leftrightarrow P(z))$	(1) $\forall$I
(3) $\exists y\ \forall z(z$ Sat $y \leftrightarrow P(z))$	(2) $\exists$I
(4) $\exists y\ \forall z(z$ Sat $y \leftrightarrow \neg(z$ Sat $y))$	(3) with '$\neg(z$ Sat $y)$' for Pz

Line (4) contradicts a theorem of predicate logic. This looks like the Unsatisfied paradox, but the Unsatisfied paradox is derived without breaching a standard restriction against circularity. At step 3 above, we left off a standard constraint 'providing y does not occur free in P' (see Sect. 9.1). This standard constraint is designed to avoid circular definitions: it invalidates proof (S#).[9]

If one already accepts dialetheia, then one may well add contradictions from circular definitions, as does Routley (1980, p. 915). The *existence* of anything based on such contradictions may seem bizarre, but the *occurrence* of contradictions *given circular definitions* is unsurprising. Routley (1980) and Gupta and Belnap (1993) bravely make some sense beyond the bounds of circular definitions; nevertheless, those bounds are appropriate for delineating what counts as an Unsatisfied paradox. The above contradiction is clearly attributable to allowing an explicit circular definition, and, if it is a paradox, it is not the same type as the Unsatisfied paradox.

If one wants to engage in debate about what kinds of paradox there are, prior to making assumptions about how best to address them, as Priest (1994) does, then one ought to grant standard restrictions on merely circular definitions for sake of debate. And I ought to too.

Another source of circular definitions is to accept the Humpty Dumpty principle (when one uses a word, it means just what one chooses it to mean) as another thing one can do with words, so that one can merely stipulate 'p' means x does not satisfy p, put such words to work, and let 'p' line up to collect its dues in contradictions at the end of the week (Dodgson 1871, Ch. VI). In my

[9] More specifically, line (3#) below is an invalid use of $\exists$I:

(2*) $\forall z\ (z$ Sat $\langle\neg(x$ Sat $y)\rangle \leftrightarrow \neg(z$ Sat $y))$	As above
(3#) $\exists y\ \forall z\ (z$ Sat $y \leftrightarrow \neg(z$ Sat $y))$	(2*) $\exists$I

opinion, if a circular stipulation of meaning gives rise to contradiction, it is negated by a *reductio* argument. (When I introduced the Unsatisfied paradox in Sect. 1, I did not fix the reference of 'my favourite predicate' by definition.)

Secondly, I will set aside contradictions from dubious stipulations, such as:

> (U*) Let the Unsatisfied Predicate be 'does not satisfy the Unsatisfied Predicate'. Then anything satisfies the Unsatisfied Predicate iff it does not.

If it is merely assumed that the Unsatisfied predicate is 'does not satisfy the Unsatisfied predicate', then the resultant contradiction seems to overturn this assumption by *reductio*. If it is not being assumed, then what basis is there for this stipulation? So it is dubious whether (U*) is an example of the Unsatisfied paradox. It is not actually a circular definition; nevertheless, I set this example aside.[10]

Thirdly, although Grelling's was introduced as a paradox of denotation, I am only dealing with versions of Grelling's that depend on satisfaction. Grelling's paradox is also known as 'the paradox of Heterologicality'. An adjective is heterological iff it does not apply to itself. An adjective is autological just if it applies to itself. 'Short' is autological, and 'long' is heterological. Grelling asks whether 'heterological' is heterological or autological. These conundrums can be reduced to satisfaction paradoxes — which I am analysing — or otherwise, in my opinion, be treated as *reductios* of such definitions. Granted, there is still some puzzle about why such a definition should fail; nevertheless, such a definition is a mere stipulation unless it is underpinned by some stronger claim such as one based on use of the Satisfaction Principle.

6. The difference of inference

I take up the point made at the end of section 4 that for premissless derivations of a contradiction the pathology must be located in the inferences. In such a case, one or a number of inferences have led from truth to falsehood.[11] Some paradoxes are best resolved as surprising

[10] Personally, I think (U*) is an example of the Unsatisfied paradox, but I acknowledge its premiss *is* dubious and set it aside in favour of analysing clearer examples.

[11] This is so even if the contradictory conclusion is a dialetheia, in which case the inferences have *also* preserved truth.

reductios of erstwhile plausible premisses, but both the Unsatisfied paradox and Grelling's have premissless versions (at least without contingent premisses): neither of them is that kind of paradox. To eliminate the premiss from the Unsatisfied paradox, we can use somewhat informal proofs or syntactic functions with purely formal derivations, as I show respectively in sections 6.1 and 6.2. Grelling's seems simpler, because the premiss was just an abbreviation anyway; nevertheless, I argue there is a little more to formalizing Grelling's in section 6.3. Furthermore, if derivations of both paradoxes can be reduced to a common set of (purely formal) inferences, I will argue they have the same pathology. I return to these considerations in sections 6.4 through to 7, after giving premissless derivations of the two paradoxes.

6.1 Informal premissless versions of the Unsatisfied paradox

The usual means of informally shortening proofs include enthymemes and less than purely syntactic rules. I consider two ways this can be done for the Unsatisfied paradox, basically in order to deprecate such proofs as unsuitable for taxonomic purposes.

Restricting the satisfaction rules to use canonical naming has forced the explicit use of =E in proofs so far. By lifting this restriction, one can give a one-step derivation using an informal satisfaction schema:

(Non-canonical Satisfaction schema) $z \text{ Sat } \sigma \leftrightarrow \text{P}z$

where 'z' is replaced by any term, 'Sat' is a two-place satisfaction relation, 'Pz' is replaced by a closed sentence formed from any open sentence with its one free variable replaced by the term used in place of 'z', and 'σ' is replaced by a name of that open sentence.

Notice how this rule is not purely formal: when applying this rule, one has to *know* that 'σ' is replaced with a name for the open sentence P(x).

First, if such informality is acceptable — and it may be for some purposes — then let j be a name in the language for the open sentence 'x does not satisfy j' and here is a one-line proof of the Unsatisfied paradox:

(U5) The Unsatisfied paradox: proven in one step using non-canonical naming

(1) $t \text{ Sat } j \leftrightarrow \neg(t \text{ Sat } j)$ By the meaning of j and the Non-canonical Satisfaction schema

Although the denotation of j is not a stated premiss, it is required for soundness. Even so, one cannot give a purely formal proof by merely adding the unstated premiss, because that premiss has to be interpreted in order to meet the condition under which the Non-canonical Satisfaction schema can be used.[12] So (U6) is informal as well:

(U6) The Unsatisfied paradox: proven using non-canonical naming

(1) $j = \langle\neg(x \text{ Sat } j)\rangle$	Premiss
(2) $t \text{ Sat } j \leftrightarrow \neg(t \text{ Sat } j)$	from (1) by the denotation of j and the Non-canonical Satisfaction schema

Secondly, other sorts of indexical noun phrases avoid the need for a premiss, given the Non-canonical Satisfaction schema.

(U7) The Unsatisfied paradox: another proof in one step using non-canonical naming

(1)	*t* satisfies the predicate of the second component of this biconditional ↔ *t* does not satisfy the predicate of the second component of this biconditional	Non-canonical Satisfaction schema

Any seemingly sound proof using the non-canonical satisfaction schema, such as (U5), (U6), and (U7) can clearly be made formal by supplying unstated premisses, defining any syntactic functions, and using =E, SatE, SatI, and some rules of first-order logic: at a minimum →I, ↔I, or ¬I, &I. We do not need to prefer such derivations, *unless* we require that level of formality, standardization, and minimization of rules for some purpose (such as localizing logical pathologies and comparing derivations).

6.2 Some formal versions of the Unsatisfied paradox without premisses
Syntactic functions (and encodings) can be used to obviate the need for a contingent premiss in versions of the Unsatisfied paradox. I will present two examples.

First, let me adapt a device from Quine (1995a) and use a grammatical function, PSP(x), for partial self-predication. Quine's device is a function for self-predication, SP(x), such that:

$$\text{SP}(\langle \text{P}x\rangle) = \langle \text{P}(\langle \text{P}x\rangle)\rangle$$

[12] The same holds true for proofs of the Liar using a non-canonical T-schema, or non-canonical Truth introduction or elimination rules. I dare say this applies to many proofs in the literature. They too can be formalized by using canonical rules and =E.

The self-predication of 'x is long' is:

'x is long' is long

Partial self-predication is also a function, such that:

$\mathrm{PSP}(\langle x\mathrm{P}y\rangle) = \langle x\mathrm{P}\langle x\mathrm{P}y\rangle\rangle$

Given the canonical name of any open sentence with two free variables as an argument, the function substitutes a canonical name of that sentence for the second variable. The partial self-predication of 'x is longer than y' is:

x is longer than 'x is longer than y'

Here then is the wanted proof of the Unsatisfied paradox:

(U8)	The Unsatisfied paradox: a derivation using the PSP function	
(1)	$\mathrm{PSP}(\langle\neg(x \text{ Sat } \mathrm{PSP}(y))\rangle) = \langle\neg(x \text{ Sat } \mathrm{PSP}(\langle\neg(x \text{ Sat } \mathrm{PSP}(y))\rangle))\rangle$	PSP function
(2)	$t \text{ Sat } \langle\neg(x \text{ Sat } \mathrm{PSP}(\langle\neg(x \text{ Sat } \mathrm{PSP}(y))\rangle))\rangle \leftrightarrow \neg(t \text{ Sat } \mathrm{PSP}(\langle\neg(x \text{ Sat } \mathrm{PSP}(y))\rangle))$	Sat-schema
(3)	$t \text{ Sat } \mathrm{PSP}(\langle\neg(x \text{ Sat } \mathrm{PSP}(y))\rangle) \leftrightarrow \neg(t \text{ Sat } \mathrm{PSP}(\langle\neg(x \text{ Sat } \mathrm{PSP}(y))\rangle))$	(1), (2) =E

As a small cost for having no contingent premiss, this derivation uses the PSP function. So, a minimal rule set for deriving the Unsatisfied paradox without a premiss is {=E, SatE, SatI, SL} (using a syntactic function, like the PSP function).

Secondly, a version of the Unsatisfied paradox can be derived using an instance of Gödel's diagonal lemma instead. This lemma is provable for certain systems using various functions, and predicate logic with identity (McGee 1991, pp. 24–5).

By Gödel's lemma, one can find a formula for an argument of the expression '$\neg(x \text{ Sat } y)$' such that the following universal statement is a theorem:[13]

$\forall x(\mathrm{G}x \leftrightarrow \neg(x \text{ Sat } \underline{\ulcorner \mathrm{G}x \urcorner}))$

where the corner brackets signal a Gödel numbering, such that there is a numerical code $\ulcorner @ \urcorner$ for each formula $@$ and underlining represents a

[13] One can exactly match the form of the lemma in McGee 1991, p. 24, by using '$\neg(x$ true of $y)$'.

numeral, that is, $\underline{\ulcorner Gx \urcorner}$ is the numeric term for the number that encodes some formula G*x*.

So, by universal instantiation (∀E), for any closed term *t*, the following biconditional is also a theorem:

$$Gt \leftrightarrow \neg(t \text{ Sat } \underline{\ulcorner Gx \urcorner})$$

So a proof of the Unsatisfied paradox is summarized below.

(U9) The Unsatisfied paradox using Gödel's lemma[14]

(1) $Gt \leftrightarrow \neg(t \text{ Sat } \underline{\ulcorner Gx \urcorner})$	Various functions, =E, predicate logic
(2) $(t \text{ Sat } \underline{\ulcorner Gx \urcorner}) \leftrightarrow Gt$	Sat-schema
(3) $(t \text{ Sat } \underline{\ulcorner Gx \urcorner}) \leftrightarrow \neg(t \text{ Sat } \underline{\ulcorner Gx \urcorner})$	(1), (2) SL

I note that a full derivation of this proof depends on the validity of the minimal set of inferences used in (U8) as well as some additional inferences.

6.3 Fully formal versions of Grelling's

It seems a canonical name can be used in an instance of the satisfaction schema to prove a contradiction in one step from no premisses:

(G3) Grelling's paradox: a one-line proof

(1) $\langle\neg(x \text{ Sat } x)\rangle \text{ Sat } \langle\neg(x \text{ Sat } x)\rangle \leftrightarrow \neg(\langle\neg(x \text{ Sat } x)\rangle \text{ Sat } \langle\neg(x \text{ Sat } x)\rangle)$	Sat-schema

However, this and previous proofs have not been purely formal. They have tacitly relied on a distinctive syntactic rule. The rule is so trivial it is not usually worth making explicit. Nevertheless, an ambiguity is not discerned in this use of the Sat-schema. If first-order logic has a way of representing '*x* satisfies itself', it is by a repetition of the variable, as in '*x* Sat *x*', but then this expression seems ambiguous between the dyadic predicate '$x \text{ Sat}^2\ x$' and the monadic predicate '$x \text{ Sat}^1\ x$'. (In the latter case, the number of places comes apart from its adicity: such a 2-place predicate is also monadic; so this is not purely formal, at any rate I think it best to consider '$x \text{ Sat}^1\ x$' as somewhat informal, standing in a place where we need a representation of '*x* satisfies itself'.) The syntactic rule that takes a two-place open sentence with a repeated

[14] If the Sat-schema were merely an assumption, then this would be a proof of the indefinability of satisfaction theorem (Eldridge-Smith 2008, Ch. 5).

variable to a one-place open sentence with a reflexive predicate is Reflection, which Quine (1960) formalized using a predicate operator:

> '*Ref*', for 'reflexive', turns a 2-place predicate into a 1-place reflexive predicate; so '*Ref R*' means '… has *R* to itself'. The procedure is generalised to polyadic predicates and to compound predications … (Haack 1978, p. 48)

It seems first-order logic follows the conventions of natural language, in which the monadic reflexive predicate and dyadic predicate are ambiguously represented by the same word or symbol. Reflection purely formally distinguishes and relates them with the Ref predicate operator, and I will use this representation shortly after saying a little more about why such formal detail is required.

In first-order logic with identity there is only one predicate with any inference rules, and the use of =I and =E does not have any special restrictions. When logic is expanded with inference rules for predicates like 'Sat', the SatI and SatE rules need formal adicity restrictions, as set out in section 1. The absence of these restrictions would allow even more contradictions, for example:

(A*) The everything is identical to *a* paradox

(1) $a = a$	=I
(2) a Sat $\langle x = a \rangle$	(1) SatI (without its adicity restrictions)
(3) $\forall y$ (y Sat $\langle x = a \rangle$)	(2) ∀I
(4) t Sat $\langle x = a \rangle$	(3) ∀E
(5) $t = a$	(4) SatE
(6) $\forall x(x = a)$	(5) ∀I

Everything is identical to *a*, unless there is some restriction on SatI, as per the general form of SatI from section 1:

$$\text{(SatI)}\ \mathrm{P}^n(z_1, z_2, \ldots, z_n) \vdash (z_1, z_2, \ldots, z_n)\ \mathrm{Sat}^{n+1} \langle \mathrm{P}^n(x_1, x_2, \ldots, x_n) \rangle$$

Whether $a\mathrm{P}a$ entails (a Sat $\langle x\mathrm{P}a \rangle$) as well as ($a$, a Sat $\langle x\mathrm{P}x \rangle$) by SatI, depends on whether it is subject to adicity constraints or not; given its adicity constraints, line (2) above is invalid.

The adicity restrictions on SatI and SatE, and consequently the Sat-schema, are such that what looked like a one-line derivation of Grelling's contradiction, line (1) of (G3), actually has these adicities:

$$\langle \neg x \text{ Sat } x \rangle \text{ Sat}^2 \langle \neg x \text{ Sat}^1 x \rangle \leftrightarrow \neg(\langle \neg x \text{ Sat } x \rangle \text{ Sat}^1 \langle \neg x \text{ Sat } x \rangle)$$

The left-hand side is a binary predicate; whereas a monadic predicate is negated on the right-hand side. So, line (1) in proof (G3) is not a strictly formal contradiction. The same issue affects line (2) of proof (G1); so that line (3) of that proof has the following adicities and is not a purely formal contradiction:

$$r \text{ Sat}^2\ r \leftrightarrow \neg(r \text{ Sat}^1\ r)$$

Unfortunately, paradox is quickly regained using Reflection.

Formally, we have the following rules for the Ref predicate operator:

(RefI) $\mathrm{P}^{n+1}(x_1, \ldots, x_n, x_n) \vdash (\text{Ref P})^n(x_1, \ldots, x_n)$

(RefE) $(\text{Ref P})^n(x_1, \ldots, x_n) \vdash \mathrm{P}^{n+1}(x_1, \ldots, x_n, x_n)$

From these, we can derive Quine's (1960) Reflection schema:[15]

(Reflection) $(\text{Ref P})^n x, \ldots, x_n \leftrightarrow \mathrm{P}^{n+1} x, \ldots, x_n, x_n$

Reflection abbreviates use of {RefE, RefI, →I, ↔I}.

Note that (in connection with the above discussion of (A*), (G1), and (G3)):

$(\text{Ref P})\ a \vdash a \text{ Sat } \langle(\text{Ref P})\ x\rangle$ (by SatI)

and

$a \text{ Sat } \langle(\text{Ref P})x\rangle \vdash (\text{Ref P})\ a$ (by SatE)

Here is a formal proof of Grelling's using Reflection:

(G4)	Grelling's paradox: a proof using Reflection	
(1)	$(\text{Ref Sat})\ \langle\neg(\text{Ref Sat})(x)\rangle \leftrightarrow \langle\neg(\text{Ref Sat})(x)\rangle$ Sat $\langle\neg(\text{Ref Sat})(x)\rangle$	Reflection
(2)	$\langle\neg(\text{Ref Sat})(x)\rangle$ Sat $\langle\neg(\text{Ref Sat})(x)\rangle \leftrightarrow \neg(\text{Ref Sat})\ \langle\neg(\text{Ref Sat})(x)\rangle$	Sat-schema
(3)	$\langle\neg(\text{Ref Sat})(x)\rangle$ Sat $\langle\neg(\text{Ref Sat})(x)\rangle \leftrightarrow \neg(\langle\neg(\text{Ref Sat})(x)\rangle$ Sat $\langle\neg(\text{Ref Sat})(x)\rangle)$	(1), (2) SL

And here is another derivation:

(G5)	Grelling's paradox derived using RefE and RefI	
(1)	$\langle\neg(\text{Ref Sat})(x)\rangle$ Sat $\langle\neg(\text{Ref Sat})(x)\rangle$	Assumption

[15] It has been suggested to me in response to Eldridge-Smith 2008 that Reflection (or 'reflexivization' as I called it) can be achieved by introducing a definition: $\text{Sat}^1(x, x) =_{\text{df}} \exists y\, (\text{Sat}^2(x, y)\ \&\ y = x)$. The idea being that the definiens as a whole is a one-place open sentence. However, this definition is not reductive and does not eliminate the need for Reflection: in deriving Grelling's using this more complex expression, one will still be unable to formally meet the adicity restriction on SatI without first using RefI.

(2)	$\neg(\text{Ref Sat})(\langle\neg(\text{Ref Sat})(x)\rangle)$	(1) SatE
(3)	$\neg(\langle\neg(\text{Ref Sat})(x)\rangle \text{ Sat } \langle\neg(\text{Ref Sat})(x)\rangle)$	(2) RefE
(4)	$\neg(\langle\neg(\text{Ref Sat})(x)\rangle \text{ Sat } \langle\neg(\text{Ref Sat})(x)\rangle)$	(1), (3) ¬I [1]
(5)	$\neg(\text{Ref Sat})(\langle\neg(\text{Ref Sat})(x)\rangle)$	(4) RefI
(6)	$\langle\neg(\text{Ref Sat})(x)\rangle \text{ Sat } \langle\neg(\text{Ref Sat})(x)\rangle$	(5) SatI
(7)	$\langle\neg(\text{Ref Sat})(x)\rangle \text{ Sat } \langle\neg(\text{Ref Sat})(x)\rangle$ $\& \neg(\langle\neg(\text{Ref Sat})(x)\rangle \text{ Sat } \langle\neg(\text{Ref Sat})(x)\rangle)$	(6), (4) &I

(G4) relies on {RefE, RefI, SatE, SatI, →I, and ↔I} and (G5) using {RefE, RefI, SatE, SatI, ¬I, and &I}. In the next section I will argue that these particular rules are suitable for comparing the pathology of Grelling's (whatever that pathology may be) with that of the Unsatisfied paradox.

6.4 Normalized derivations

Each proof above, whether formal or informal, has been identifiable as a version of the Unsatisfied paradox or Grelling's based on the paradoxical extension of its predicate. Any such proof can be translated into a formal derivation using only strong introduction and elimination rules for atomic predicates, predicate operators, connectives, and quantifiers, such as =E, SatE, SatI, RefE, RefI, ¬I, and &I. The advantages of doing so are that minimal sets of these inferences in such normalized derivations effectively localize the paradoxical pathology of each paradox and their pathologies can be compared by comparing minimal sets of such rules. I will explain why in eight points below.

First, using only contingent or stipulated premisses in a normalized derivation exclusively separates any issues about the truth of the premisses from any other issues about validity or the principles of a theory. For example, a non-normalized proof could use an instance of the Sat-schema as a premiss; but in the normalized version of the proof it will be translated into inferences, based on the Satisfaction Principle: if there is an issue with the Sat-schema, it is not a matter of contingency.

Secondly, as the rules are purely formal, the Sat rules must use canonical names, and any denotation premisses must be stated. (The Non-canonical Satisfaction schema is not purely formal, as discussed in Sect. 6.1.)

Thirdly, for paradoxical taxonomy, the normalized rules should be strong enough that there is at most one introduction and one elimination rule for a predicate, predicate operator, connective, or quantifier. The SatI and SatE rules could be weakened in multiple ways.[16] However, weakening of these rules appears to generate multiple paradoxes where there need only be one. That is, there are parsimonious reasons not to distinguish paradoxes using weakened introduction and elimination rules. For example, consider the following pairs of weakened rules as alternatives for SatI and SatE:

(Sat-elim) z Sat $\langle \mathrm{P}x \rangle \rightarrow \mathrm{P}z$

(Necessitation) If Pz is a theorem, then z Sat $\langle \mathrm{P}x \rangle$ is a theorem

A version of the Unsatisfied paradox can be derived using these rules, and a different derivation can be given using the following:

(Sat-intro) $\mathrm{P}z \rightarrow z$ Sat $\langle \mathrm{P}x \rangle$

(Converse Necessitation) If z Sat $\langle \mathrm{P}x \rangle$ is a theorem, then Pz is a theorem

If these weaker rules were suitable for paradoxical taxonomy, it would appear as though we have even more paradoxes of satisfaction.

Fourthly, 'conflated' rules are unsuitable for paradoxical taxonomy. If they are relevant they can be derived. For example, the following rule conflates =E with the Sat-schema:

(Conflated =EwithSat-schema) If $a = \langle \mathrm{P}x \rangle$, then z Sat $a \leftrightarrow z$ Sat $\langle \mathrm{P}x \rangle$

Given the identity of my favourite predicate, the Unsatisfied paradox can be derived using this rule together with →E. However, this rule conflates any issues with =E with any issues with the Sat-schema. It is therefore unsuitable for a taxonomy that seeks to localize potential sources of pathology. (This is a latent issue for the Non-canonical Satisfaction schema, which would have this issue if its condition were changed to 'where $\sigma = \langle \mathrm{P}x \rangle$').

Fifthly, compound rules involving more than one predicate or connective, like conflated rules, should be decomposed into the normalized rules. For example, the following rule can be used to deduce both

[16] Friedman and Sheard 1987 investigated using multiple weakenings of the T-schema to derive versions of the Liar. Analysis of their work will show that one weakening of Truth Introduction is always used with another weakening of Truth Elimination in any proof of the Liar. The same general result can be found to hold if one analyses weakenings of SatI and SatE. Moreover, any proof of the Unsatisfied paradox using such weakenings can be mapped to a proof using (canonical) SatI, SatE and =E.

the Unsatisfactory paradox and Grelling's, but it is clearly just a conjunction of a number of rules.

(UnderTheSun) (If $a=b$, then Pa iff Pb) &((Ref P)$x \leftrightarrow x$Px) & (P$z \leftrightarrow z$ Sat $\langle$P$x\rangle$)

We can derive the paradoxes from this rule, but it gives no insight into localizing their pathology.

Any rules equivalent to the introduction and elimination rules I have preferred can clearly be translated into the normalized rules. Furthermore, formal equivalences may help abstract away from particular predicates or connectives. So, sixthly, if rules are equivalent or reducible to fewer rules for atomic predicates, connectives, or quantifiers, it makes sense to abstract away from surfeit or particular predicates, or reduce the rules to more primitive rules. (Some further reduction could be achieved along these lines, but is not required here. Here are two simple examples: there are surplus connectives, which are reducible in well-known ways, and 'Sat' is the converse of 'True of'.)

Seventhly, as a seemingly sound derivation to an apparently false conclusion is a paradox, if the conclusion is apparently contingently false, its negation can be added as a premiss to yield a seemingly sound argument to a contradiction. (For present purposes, this localizes any pathology to either the premisses or the inferences.)

Eighthly, where some inferential rules are not directly commensurable with the set of introduction and elimination rules I am using, but are able to derive something like the Unsatisfied paradox or Grelling's, then that paradox can be translated back into English and then derived using normalized introduction and elimination rules. (There may be more paradoxes in heaven and earth than are expressible in English, but they are beyond our current scope.)

6.5 Explaining the obvious difference: why the Unsatisfied paradox needs substitution of identicals

Any normalized derivation of the Unsatisfied paradox uses substitution of identicals because, being purely formal, it uses canonical naming with the Sat rules and then either (1) substitutes a co-referential term for that canonical name of the predicate, which is done using =E, or (2) proves there is a predicate Gx such that $\forall x(Gx \leftrightarrow \neg(x \text{ Sat } \langle Gx\rangle))$. All extant proofs of the latter, namely proofs of Gödel's lemma, rely on =E. In either case, the object assumed to satisfy these predicates is irrelevant to these key steps that generate the contradiction, which explains the paradoxical extension of such predicates. Thus, the reliance on =E is tied indirectly to the obvious difference between a version of the Unsatisfied paradox and Grelling's.

6.6 Explaining the obvious difference: Grelling's need for Reflection

Any normalized derivation of Grelling's paradox needs to use Reflection to be purely formal, because, as argued in section 6.3, on pain of other paradoxes, the SatI rule has adicity restrictions. These versions of Grelling's satisfaction paradox concern the conundrum of whether a reflexive predicate of non-satisfaction satisfies itself. To derive the paradox purely formally, Reflection or a similar rule is required to reduce the adicity of such a reflexive predicate for use with SatI (or the derived Sat-schema). So the use of Reflection and formal rules for the Satisfaction principle together with a reflexive predicate of non-satisfaction (or an equivalent expression) give an account of Grelling's characteristic reflexive paradoxical extension. Thus Reflection plays a part in an account of the obvious difference between the Unsatisfied paradox and Grelling's.

7. The difference of inference across variations

There are an open-ended number of versions of the satisfaction paradoxes, and one may want some reason to think the above claims about =E and the Unsatisfied paradox, and Reflection and Grelling's, apply to all normalized derivations. Toward this end, versions can be grouped as *variations* so that a possibly infinite number of versions can be considered as a finite number of variations. So I will choose some suitable variations, which plausibly exhaust all versions of these two types of satisfaction paradox, and show that the relevant claims hold for these variations.

7.1 Variations of the Unsatisfied Paradox

Minimal normalized derivations of the Unsatisfied Paradox vary in either (1) a *single* use of SatE and SatI rules, or (2) *multiple* uses of the SatE and SatI rules. These particular variations are exclusive and exhaustive of the versions of the Unsatisfied paradox. In this section, I lay out how normalized derivations of these varieties all have some reliance on =E.

The *singular* variations relate to either identities or a biconditional of the following forms, whether as premisses, functions, or theorems, where f is a suitable function:

Variation (1a) $p = \langle\neg(x \text{ Sat } p)\rangle$
Variation (1b) $f(p) = \langle\neg(x \text{ Sat } f(p))\rangle$
Variation (1c) $\forall x(\mathrm{P}x \leftrightarrow \neg(x \text{ Sat } \langle\mathrm{P}x\rangle))$

Those of type (1a) and (1b) can be used directly with an instance of the Sat-schema and =E towards a contradiction. In this respect, such contradictions seem clearly self-referential, regardless of how the self-reference was achieved.[17] All extant derivations of biconditionals of type (1c) rely on =E; in particular, derivations of Gödel's lemma depend on the validity of =E. Such biconditionals can be used with ∀E and an instance of the Sat-schema to derive a contradiction.[18]

The *multiple* variations of the Unsatisfied paradox extend *singular* variations (1a) and (1b) by using multiple identities.[19] Both can be extended in a *circular chain of reference* as in the following examples:[20]

Circular Unsatisfied paradoxes of type (variation 2a)	Circular Unsatisfied paradoxes of type (variation 2b)
$p_1 = \langle\neg(x \text{ Sat } p_2)\rangle$	$f(p_9) = \langle\neg(x \text{ Sat } f(p_{10}))\rangle$
$p_2 = \langle x \text{ Sat } p_1\rangle$	$f(p_{10}) = \langle x \text{ Sat } f(p_9)\rangle$
$p_3 = \langle\neg(x \text{ Sat } p_4)\rangle$	$f(p_{11}) = \langle\neg(x \text{ Sat } f(p_{12}))\rangle$
$p_4 = \langle\neg(x \text{ Sat } p_5)\rangle$	$f(p_{12}) = \langle\neg(x \text{ Sat } f(p_{13}))\rangle$
$p_5 = \langle\neg(x \text{ Sat } p_3)\rangle$	$f(p_{13}) = \langle\neg(x \text{ Sat } f(p_{11}))\rangle$
$p_6 = \langle x \text{ Sat } p_7\rangle$	$f(p_{14}) = \langle x \text{ Sat } f(p_{15})\rangle$
$p_7 = \langle\neg(x \text{ Sat } p_8)\rangle$	$f(p_{15}) = \langle\neg(x \text{ Sat } f(p_{16}))\rangle$
$p_8 = \langle x \text{ Sat } p_6\rangle$	$f(p_{16}) = \langle x \text{ Sat } f(p_{14})\rangle$

[17] Variations of type (1a) can be reduced to variations of type (1b) by supposing they use a vacuous identity function.

[18] It is unclear whether variation (1c) is really self-referential, or, more to the point, whether it is self-referential in the same sense as variations (1a) and (1b). But this issue is currently beside the point.

[19] It is conceivable that Gödel's lemma could be tweaked to prove circular variations of type (1c).

[20] This 'circular reference' is often considered a variety of self-reference; although, to my way of thinking, it is the converse: self-reference is a limit case of circular reference.

In the simplest example of a circular chain of reference, p_1 refers to p_2, which refers back to p_1. There are circular Unsatisfied paradoxes with any finite number of statements. (I note there are an odd number of negated statements in a circular Unsatisfied paradox, and any number of affirmative statements.)[21] For any object, a normalized derivation of a contradiction from a circular variation of type (2a) or (2b) requires iterative use of SatE, SatI and =E.

Finally, I think type (1b) can be extended by using an infinite number of identities in an *infinite chain of reference* using a suitable syntactic function that assures the truth of these identities:[22]

(U10) An Infinitely Unsatisfied paradox, a Yabloesque variation of the Unsatisfied paradox (variation 3b)[23]

Consider an infinite collection of predicates:
$f(u_1) = \langle \forall k > 1$
$\quad \neg(x \text{ Sat } f(u_k) \rangle$
…
$f(u_n) = \langle \forall k > n$
$\quad \neg(x \text{ Sat } f(u_k)) \rangle$
…

Outline of a proof:

By the Sat-schema and =E:
$t \text{ Sat } f(u_1) \leftrightarrow \forall k > 1\ \neg(t \text{ Sat } f(u_k))$
and (with some predicate logic):
$\neg(t \text{ Sat } f(u_2)) \leftrightarrow \exists k > 2\ (t \text{ Sat } f(u_k))$
Thus, if $t \text{ Sat } f(u_1)$, then
$\exists k > 2\ ((t \text{ Sat } f(u_k))\ \&\ \neg(t \text{ Sat } f(u_k)))$
So, $\neg(t \text{ Sat } f(u_1))$.
But then $\exists k > 1\ (t \text{ Sat } f(u_k))$
Suppose this occurs for $k = b$,
then repeat the above argument towards a contradiction

[21] I would call a case with an even number of negated statements a *hypodox*, viewing its semantic status as *underdetermined* (see the appendix). Alternatively, it could be construed as similar to a 'No No' paradox (Sorensen 2001, Ch. 11).

[22] It seems fair to me to call the chain of reference between the premisses *ungrounded*; although there is ongoing debate as to whether a similar sort of reference in Yablo's paradox (Yablo 1993) is any different from circular reference. Nevertheless, it seems fair to say that circular reference is a subset of ungrounded reference, even though it is debatable whether it is a proper subset.

[23] This particular paradox is similar to other infinite paradoxes already in the literature, for example, as an anonymous referee pointed out, Leitgeb's (2005) paradox of non-wellfounded definition. Leitgeb's uses an infinite sequence of definitions, the conjunction of which seems reduced to absurdity by the contradiction. I intend the above to use an infinite sequence of predicates, any one of which can be exhibited. Also, the above relies on satisfaction, whereas Leitgeb's does not (and Leitgeb cleverly avoids quantification). Leitgeb's is not an Unsatisfied paradox for these reasons. Leitgeb's seems to me to be what Quine 1962 called a *veridical* paradox. After one appreciates the paradox, one gives up any credence in the conjunction of the definitions stipulated as premisses. (U10) is intended to have premisses assured by definition of a function.

All the above variations rely on =E, SatI, SatE, and sentential logic, as well as some other rules for some variations. All versions of the Unsatisfied paradox can plausibly be mapped to these variations, so it is reasonable to claim that all their normalized derivations rely on {=E, SatI, SatE, SL}.

7.2 Variations of Grelling's paradox of satisfaction

One way in which normalized derivations of Grelling's paradox of satisfaction vary is that they concern either (1) whether a negative *atomic* reflexive satisfaction predicate satisfies itself (or not), or (2) whether a negative *compound* reflexive satisfaction predicate satisfies itself (or not). I characterize the first type as *atomic*, and the second as *circular*. In either case, other predicates may be defined as equivalent to such expressions (whether by definition, abbreviation, or synonymy). In this section, I show how Reflection is required to reduce the adicity of the reflexive predicate for use with SatI (or the derived Sat-schema) in normalized derivations for each of the above two varieties.

Derivation (G5) was a minimal version of the first variety, requiring only {RefE, RefI, SatE, SatI, SL}. The derivation (G6) below fully formalizes proof (G1). Here, the term r is just an optional abbreviational convenience, although using an abbreviation results in this derivation using =E and not being minimal. Had one not used an abbreviation, the use of =E would be unnecessary: so Grelling's does not rely on this rule. (Reflection, as a derived rule, also abbreviates use of {RefE, RefI, $\rightarrow$I, $\leftrightarrow$I}):

(G6) Grelling's satisfaction paradox

(1) $r = \langle\neg(\text{Ref Sat})(x)\rangle$	Abbreviation
(2) $r \text{ Sat } \langle\neg(\text{Ref Sat})(x)\rangle \leftrightarrow \neg(\text{Ref Sat})(r)$	Sat-schema
(3) $\neg(\text{Ref Sat})(r) \leftrightarrow \neg(r \text{ Sat } r)$	Reflection
(4) $r \text{ Sat } \langle\neg(\text{Ref Sat})(x)\rangle \leftrightarrow \neg(r \text{ Sat } r)$	(2), (3) SL
(5) $r \text{ Sat } r \leftrightarrow \neg(r \text{ Sat } r)$	(4), (1) =E

The classic paradox of Heterologicality, if based on satisfaction, is of this type:

(G7) Grelling's classic satisfaction paradox

(1)	$\text{Het}(x) \leftrightarrow \neg(\text{Ref Sat})(x)$	Abbreviational equivalence
(2)	$\langle\text{Het}(x)\rangle \text{ Sat } \langle\text{Het}(x)\rangle \leftrightarrow \text{Het}(\langle\text{Het}(x)\rangle)$	Sat-schema

(3)	$\langle \text{Het}(x)\rangle$ Sat $\langle \text{Het}(x)\rangle \leftrightarrow \neg(\text{Ref Sat})(\langle \text{Het}(x)\rangle)$	(2), (1) predicate logic
(4)	$\neg(\text{Ref Sat})(\langle \text{Het}(x)\rangle) \leftrightarrow \neg(\langle \text{Het}(x)\rangle$ Sat $\langle \text{Het}(x)\rangle)$	Reflection
(5)	$\langle \text{Het}(x)\rangle$ Sat $\langle \text{Het}(x)\rangle \leftrightarrow \neg(\langle \text{Het}(x)\rangle$ Sat $\langle \text{Het}(x)\rangle)$	(4), (3) SL

The second variety concerns compound predicates, as exemplified in the table below. Once again, to satisfy the adicity constraints on the Sat rules, {RefE, RefI} are required for a normalized derivation of a contradiction. To carry this out some other predicate operators are also required (such as 'X' for Cartesian multiplication and 'Inv' for swapping the first variable into the last variable position).

Grelling's paradoxes using compound predicates:	Equivalent monadic reflexive predicates using Reflection and other predicate operators:[24]
Does $\forall y\, \neg(x \text{ Sat } y \;\&\; y \text{ Sat } x)$ satisfy itself?[25]	$\forall y\, \neg(\text{Ref (Inv (Sat X Sat))}\; y, y, x) \leftrightarrow (\forall y\, \neg(x \text{ Sat } y \;\&\; y \text{ Sat } x))$
Does $\forall z\, \forall y\, \neg(x \text{ Sat } y \;\&\; y \text{ Sat } z \;\&\; z \text{ Sat } x)$ satisfy itself?	$\forall z\, \forall y\, \neg(\text{Ref(Inv((Sat X Sat) X Sat))}\; y, y, z, z, x) \leftrightarrow \forall z\, \forall y\, \neg(x \text{ Sat } y \;\&\; y \text{ Sat } z \;\&\; z \text{ Sat } x)$

Circular variations of Grelling's paradox can be derived from such predicates with any finite number of conjuncts. (The usual variation of Grelling's seems like a limit case with a single, simple reflexive 'conjunct', although it is then not really a conjunct at all.) However, I am not aware of an infinite version of Grelling's satisfaction paradox

[24] This equivalence indicates how Reflection can be used in a normalized derivation. The other predicate operators needed are set out as equivalences in Quine 1960. In particular:

(Cartesian multiplication)	$(\text{P X Q})x_1, \ldots, x_m, y_1, \ldots, y_n \leftrightarrow \text{P}x_{1 \ldots} x_m \;\&\; \text{Q}y_{1 \ldots} y_n$
(Major inversion)	$(\text{Inv P})x_1, \ldots, x_n \leftrightarrow \text{P}x_n, x_1, \ldots, x_{n-1}$

These schemata can be *normalized* for present purposes as introduction and elimination rules in obvious ways.

[25] A contradiction can be derived in a normalized way in about 20 lines.

that I would not deprecate as relying on premisses based on mere stipulation (see Sect. 5), and therefore premisses whose conjunction ought to be rejected.

Thus Grelling's paradox has atomic and circular reflexive variations, for each of which the paradox concerns whether the predicate satisfies itself. They are all *reflexive* in this sense.[26] The normalized derivations of these include the same minimal set of rules of inference. The atomic and compound variations above all include use of {RefE, RefI, SatE, SatI, SL} in their normalized derivations.[27]

8. Three methods of individuating kinds of paradox

Three methods for distinguishing kinds of paradox lead to the same result: the Unsatisfied paradox and Grelling's are examples of different kinds of paradox. I set these three arguments out in the following subsections.

8.1 Locating pathology in minimal normalized rule sets sufficient for contradiction

Remember that as a Principle of Localizing Pathology, a set of inference rules captures a pathology if these rules purely formally entail a contradiction from no premisses (using only syntactic functions if any). To facilitate comparison, rules have been restricted to a reference set of introduction and elimination rules. Such rules are not too weak,

[26] Thus my use of the term 'reflexive' differs from its historical use. Goldstein (2000, p. 53) points out that Russell used the term 'reflexivity' to characterize a property he believed all the paradoxes shared, but that Sorensen (1998) provides counterexamples to that view. Clearly, Russell used this term in a much broader sense than I am using 'reflexive'.

[27] Of course there are other ways of grouping versions of each paradox. These other overlapping types of paradox do not alter the above result. There are, for example, truth-functional variations (Eldridge-Smith 2008, Ch. 3). There are also hybrid paradoxes, in particular, Curry paradoxes, with features of both the Unsatisfied paradox and Curry's paradox or Grelling's paradox spliced with Curry's. Here is an example of a truth-functional variation, the Unsatisfied ESP paradox:

$$p_d = \langle \neg(x \text{ Sat } p_d) \leftrightarrow \text{Q} \rangle$$

Whether or not anything satisfies p_d, either way, ¬Q (any Q). Like Curry's one can prove all statements, one at a time. Choose a true Q, add it as a premiss, and prove a contradiction. If ¬Q is the case though, p_d is *hypodoxical* (a term explained in the appendix). Curry's, however, only requires the negation-free fragment of a sentential logic. And here is an example of a Curried Unsatisfied paradox (or an Unsatisfied Curry paradox):

$$p_c = \langle (x \text{ Sat } p_c) \rightarrow \text{Q} \rangle$$

Q can be derived in a number of ways, for example using {=E, SatE, SatI, →E, →E}.

not too strong, purely formal, and localized to a particular predicate, predicate operator, connective, or quantifier. Those paradoxes whose normalized derivations depend on the same minimal rule set to derive a contradiction have the same pathology. Therefore, we are looking for a minimal rule set in common for the satisfaction paradoxes if they are all of the same kind. Yet based on the results of sections 6 and 7, there are two minimal rule sets:

{=E, SatI, SatE, SL}

{RefE, RefI, SatI, SatE, SL}[28]

Sections 6.2, 6.4, and 6.5 explained why =E was required for normalized derivations of the Unsatisfied paradox; sections 6.3, 6.4, and 6.6 explained why Reflection was required for normalized derivations of Grelling's paradox.

Those paradoxes with different minimal contradictory rule sets have a different pathology. There is no obvious way of reducing {RefE, RefI} to {=E} or conversely reducing the latter rule to the former rules. Therefore, there are two kinds among the satisfaction paradoxes.

8.2 Locating pathology in structures

Set the above principles aside for the moment, and consider another point of view. Priest has advocated that pathology can be captured with abstract structures: paradoxes of the same kind have a structure in common. This is a two-part Principle of Common Structure:

> to convince ourselves two paradoxes are of the same kind we must convince ourselves (a) that there is a certain structure that produces contradiction and (b) that this structure is common to the paradoxes. (Priest 1994, p. 32)

Such a structure is abstracted to a certain level, according to some Principle of the Causal Level of Abstraction:

> [T]he appropriate level at which to analyse a phenomenon is the level which locates underlying causes … the correct level of abstraction for an analysis of the paradoxes … is … the level of the underlying structure that generates … contradictions. (Priest 2000, p. 125)

[28] Depending on one's sentential logic (SL), various combinations may be minimal, but {¬I, &I} and {→I, ↔I} have stood out in our discussion. While the latter depends on Excluded Middle for contradiction, the former, e.g. in derivation (U2), does not. Both paradoxes can be avoided by some very weak sentential logics, but that does not invalidate the above analysis of the inferences used to derive them.

At the right level of abstraction a paradoxical structure has maximal explanatory power without becoming enthymematic, that is, still having everything formally needed to entail a contradiction.

Furthermore, paradoxes of different kinds have a structural difference, in accord with a Principle of Structural Difference:

> If one wants to draw a fundamental distinction, this ought to be done in terms of the *structure* of the different paradoxes. (Priest 1994, p. 26)

So it is still the case that a set of rules, as a structure, captures a pathology if these rules entail a contradiction from no premisses while using only syntactically defined functions. To count as a suitable structure, such rules should be purely formal. Our introduction and elimination rules have that virtue. Granted, one wants to abstract away from particular predicates, like 'Sat', but this will not resolve the fundamental difference between {RefE, RefI} and {=E}, which are required for formal proof of a contradiction: there is no obvious way to merge them in an abstract formal structure that still entails a contradiction. Therefore, by the Principle of Structural Difference, there are two kinds among the paradoxes of satisfaction.

8.3 Locating pathology by minimum mutilation of our concepts and logic

From yet another point of view, localizing paradoxical pathology is largely a matter of the pragmatic Principle of Minimum Mutilation (PMM): when forced by paradox to revise our reasoning, we should do so in a way that does minimum mutilation to our pre-existing concepts or logical intuitions (as I adapt Quine 1991, p. 270).[29] Minimum mutilation recommends an order in searching for pathology: contingent premisses, concepts, then logic. This guidance aims at excising the minimum number of intuitively valid arguments. PMM has seemed to recommend modifying conceptual inferences or schemata (Quine 1962); however, to do this for the satisfaction paradoxes, our concept of satisfaction will have to be restricted or modified in two ways (to make minimum variance from our intuitions). Those two ways relate (in our regimented, formal way) to use of rules for the Satisfaction Principle in these paradoxes with =E and with

[29] I am using this pragmatic principle to localize pathology. However, I do not accept the following pragmatic argument by confirmation.

> Paradoxes of different kinds have different kinds of solution. Elegant, distinct solutions exist separately for set-theoretic and semantic paradoxes. Therefore, those are distinct kinds.

This argument affirms the consequent; there is little room for inductive support in this debate.

Reflection. Thus, PMM recommends modifying our logic in two ways to avoid two kinds of pathology. Therefore, there are two kinds of satisfaction paradox.

9. Russell's paradox and an unsatisfactory set-theoretic paradox

Are there set-theoretic analogues of the two satisfaction paradoxes? In section 9.1, I explore the possibilities for a set-theoretic analogue of my Unsatisfied paradox. In section 9.2, I revisit preconceptions about the analogy between Grelling's and Russell's paradox. I support this formal analogy: Russell's relies on Reflection as well!

We now use rules for introducing and eliminating membership:

Membership Introduction (∈I) $\quad \mathrm{P}z \vdash z \in \{x\colon \mathrm{P}x\}$

Membership Elimination (∈E) $\quad z \in \{x\colon \mathrm{P}x\} \vdash \mathrm{P}z$

where '$\{x\colon \mathrm{P}x\}$' represents a canonical name for a set such that the predicate, P, determining membership of the set, can be recovered from the name of the set.

From these, we derive the following schema (as an abbreviation for using the rules {∈E, ∈I, →I, and ↔I}):

(Abstraction schema) $\quad z \in \{x\colon \mathrm{P}x\} \leftrightarrow \mathrm{P}z$

9.1 The Unsatisfactory paradox

Here is the set-theoretic argument which is analogous to my Unsatisfied paradox.

(Q*) The Unsatisfactory paradox

(1) $q = \{x\colon x \notin q\}$	Premiss
(2) $t \in \{x\colon x \notin q\} \leftrightarrow t \notin q$	Abstraction schema
(3) $t \in q \leftrightarrow t \notin q$	(1), (2) =E

(Q*) relies on {=E, ∈I, ∈E, SL}. This is certainly distinct from Russell's paradox. An anonymous referee kindly pointed out that line (3) has some history as a contradiction in certain paraconsistent theories. These use an unrestricted comprehension axiom. We can derive comprehension from abstraction using ∀I and ∃I:

(Comprehension) $\quad \exists w \forall x (x \in w \leftrightarrow \mathrm{P}x)$ where w is not free in P

The restriction on free *w* avoids circular definitions, unsurprisingly avoiding some contradictions:

> Removing this substitutional restriction opens the way for the formation of further inconsistent sets, e.g. most simply a set Z with the property that $x \in Z \leftrightarrow \neg x \in Z$, upon writing $\neg x \in Z$ for [P in unrestricted Comprehension]. (Routley 1980, p. 915)[30]

Contradiction becomes paradox if we can maintain restricted comprehension and still derive this contradiction, as per argument (Q*) above.[31]

If we maintain the restriction against circular definition, as I have throughout this article (for reasons explained in Sect. 5), then it is very doubtful whether there is any contingent basis for the identity in line (1) of (Q*). I cannot exhibit my favourite set in the way I could my favourite predicate.

However, on analogy with the Unsatisfied paradox, it seems reasonable to use a function to eliminate the need for a premiss. Indeed, we can adapt partial self-predication and interpret the angle bracket expressions as canonically naming sets. With this in mind, here is our function again.

$$\mathrm{PSP}(\langle x\mathrm{P}y\rangle) = \langle x\mathrm{P}\langle x\mathrm{P}y\rangle\rangle$$

We can interpret $\{x\colon \mathrm{P}x\}$ as canonically naming the set of things that are, or satisfy the predicate P, so that:

$$\mathrm{PSP}(\{x\colon x\mathrm{P}y\}) = \{x\colon x\mathrm{P}\{x\colon x\mathrm{P}y\}\}$$

Here then is the wanted proof of the Unsatisfactory paradox:

(Q1) The Unsatisfactory paradox

(1) $\mathrm{PSP}(\{x\colon x \notin \mathrm{PSP}(y)\}) = \{x\colon x \notin \mathrm{PSP}(\{x\colon x \notin \mathrm{PSP}(y)\})\}$	PSP function
(2) $t \in \{x\colon x \notin \mathrm{PSP}(\{x\colon x \notin \mathrm{PSP}(y)\})\} \leftrightarrow t \notin \mathrm{PSP}(\{x\colon x \notin \mathrm{PSP}(y)\})$	Abstraction
(3) $t \in \mathrm{PSP}(\{x\colon x \notin \mathrm{PSP}(y)\}) \leftrightarrow t \notin \mathrm{PSP}(\{x\colon x \notin \mathrm{PSP}(y)\})$	(1), (2) =E

[30] See also Weber 2010b, who refers to Routley.

[31] I presented the Unsatisfactory paradox using argument (Q*) as the set-theoretic analogue of the Unsatisfied paradox in Eldridge-Smith 2008; I thought it really was unsatisfactory as I did not see how to make it seem sound. I now consider that functions provide a means of assuring truth of a premiss with this form.

(Q1) depends on {=E, ∈I, ∈E, SL}. Line (1) is true by the definition of the PSP function, and we have our Unsatisfactory paradox, notwithstanding the restriction on Comprehension.

9.2 Russell's paradox

Grelling's is directly analogous to Russell's paradox (Quine 1962). What I add to that analogy is the reliance on Reflection. Here is an informal proof using superscripting to represent the effect of using Reflection on adicity.

(R1) Russell's paradox[32]

(1) $\{x\colon x \notin^1 x\} \in^2 \{x\colon x \notin^1 x\} \leftrightarrow$ $\{x\colon x \notin^1 x\} \notin^1 \{x\colon x \notin^1 x\}$	Abstraction schema
(2) $\{x\colon x \notin^1 x\} \in^1 \{x\colon x \notin^1 x\} \leftrightarrow$ $\{x\colon x \notin^1 x\} \notin^1 \{x\colon x \notin^1 x\}$	(1) Reflection

A fully formal derivation would parallel (G5) in section 6.3, as does (R2):

(R2) Russell's paradox derived using ∈E and ∈I

(1) $\{x\colon (\mathrm{Ref} \notin)\ x\} \in \{x\colon (\mathrm{Ref} \notin)\ x\}$	Assumption
(2) $(\mathrm{Ref} \notin)(\{x\colon (\mathrm{Ref} \notin)\ x\})$	(1) ∈E
(3) $\{x\colon (\mathrm{Ref} \notin)\ x\} \notin \{x\colon (\mathrm{Ref} \notin)\ x\}$	(2) RefE
(4) $\{x\colon (\mathrm{Ref} \notin)\ x\} \notin \{x\colon (\mathrm{Ref} \notin)\ x\}$	(1), (3) ¬I [1]
(5) $(\mathrm{Ref} \notin)(\{x\colon (\mathrm{Ref} \notin)\ x\})$	(4) RefI
(6) $\{x\colon (\mathrm{Ref} \notin)\ x\} \in \{x\colon (\mathrm{Ref} \notin)\ x\}$	(5) ∈I
(7) $\{x\colon (\mathrm{Ref} \notin)\ x\} \in \{x\colon (\mathrm{Ref} \notin)\ x\}$ & $\{x\colon (\mathrm{Ref} \notin)\ x\} \notin \{x\colon (\mathrm{Ref} \notin)\ x\}$	(6), (4) &I

I note that (R2) depends on the minimal rule set {RefE, RefI, ∈E, ∈I, SL}.

One might try to obviate the need for Reflection by defining a predicate, x is ordinary iff $x \notin x$. But here, 'is ordinary' is monadic, so it can only be formally defined as equivalent to a monadic predicate, which is only given if Reflection is given. Such a definition relies on Reflection in that it only seems valid to define a monadic predicate expression in terms of a relation with identical arguments just because we assume Reflection is always valid. Purely formally, x is ordinary iff $(\mathrm{Ref} \notin)\ x$.

[32] There are also circular variations parallel to those for Grelling's paradox (Quine 1951, p. 129).

10. Two kinds of paradox

I presented a new paradox of satisfaction and contrasted it with Grelling's paradox. A predicate of the new paradox is paradoxical for any term whereas Grelling's is only paradoxical in certain cases. Each paradox has versions that do not require premisses, so the pathology of each paradox is not some fallacy related to its premisses. In normalized derivations, in addition to predicate rules for Satisfaction introduction and elimination, Grelling's relies on rules to introduce and eliminate Reflection, whereas the Unsatisfied paradox relies on substitution of identicals. As I explained in sections 6.5 and 6.6, this difference underpins the different paradoxical extensions of the two predicates. Also, as I argued in section 8, this difference tracks two pathologies, and so these satisfaction paradoxes are of different kinds.

Furthermore, these pathologies also affect set-theoretic paradoxes. The same pathologies are tracked. We can see this by abstracting from the particular introduction and elimination rules for satisfaction and membership. These rules are formally the same (see McDermott 1977, pp. 247–8). Moreover, the Sat-schema and the Abstraction schema entail that:

$$z \text{ Sat } \langle \text{P}x \rangle \leftrightarrow z \in \{x\text{: P}x\}$$

So, effectively, the normalized introduction and elimination rules for satisfaction and membership are equivalent. Therefore, the following two pairs of equivalent rule sets characterize the same two kinds of paradoxes among the satisfaction and set-theoretic paradoxes:

Insatiable: {{=E, SatI, SatE, SL}, {=E, ∈I, ∈E, SL}}

Reflexive: {{RefE, RefI, SatI, SatE, SL}, {RefE, RefI, ∈I, ∈E, SL}}

This result is not reduced by some further abstraction. That is, the =E and Reflection rules are not formally the same, so one cannot simply abstract away from them.

The Unsatisfactory paradox is the direct formal analogue of the Unsatisfied paradox, and Russell's is the formal analogue of Grelling's. Russell's not only has analogous proofs to Grelling's, it has the same pathology; a different pathology than that shared by the Unsatisfactory and the Unsatisfied paradoxes. There are then two irreducible pathologies cutting across the semantic and set-theoretic paradoxes. Thus, there are two kinds of paradox, insatiable and reflexive, cutting across the semantic and set-theoretic families of paradoxes.

11. Import of insatiable and reflexive kinds of paradoxes

As two kinds of paradox, the insatiable and reflexive paradoxes pose a clear counter-example to uniform theories of such paradoxes; also, as insatiable and reflexive kinds of paradox cut across the distinction between semantic and set-theoretic paradoxes — notwithstanding many kinds of solutions being aligned with that latter distinction — this new distinction runs counter to Ramsey's (1926) distinction.

A proponent of Ramsey's distinction between semantic and set-theoretic kinds of paradox could argue that there are four kinds, two kinds of satisfaction paradox and two kinds of membership paradox, but parsimony rules against it. More likely, a defence of Ramsey's distinction will seek to find the Unsatisfactory paradox true to its name. Then, the Ramsey distinction would be primary, but there would still be a secondary distinction between insatiable and reflexive semantic paradoxes, that is, between two kinds of satisfaction paradox.

Priest (1994, 2002) has argued that the semantic and set-theoretic paradoxes are uniformly *Inclosure* paradoxes (except for non-truth-functional Curry paradoxes). There are, however, two pathologies, and so these paradoxes are not of a uniform kind. Discrediting the Unsatisfactory paradox is not a sufficient defence; the Unsatisfied paradox would have to be discredited as well. Perhaps, a defence of a uniform structure will argue that the Unsatisfied paradox relies on a diagonal function (or premiss) that Grelling's does not require, and argue that Grelling's achieves diagonalization just by using inferential steps about satisfaction, such as SatE and SatI. However, as I argued in section 6, such rules do not purely formally entail a contradiction on their own: they do so in combination with Reflection. Therefore, either the Unsatisfied paradox is a different kind than Priest's Inclosure paradoxes, or the Inclosure paradoxes are not all uniform. Priest's own Principle of Uniform Solution suggests that the two kinds of paradox, insatiable and reflexive, have two kinds of solution:

> If two paradoxes are of different kinds, it is reasonable to expect them to have different kinds of solutions; on the other hand, if two paradoxes are of the same kind, then it is reasonable to expect them to have the same kind of solution. … *same kind of paradox, same kind of solution*. (Priest 1994, p. 32)

If this is correct, then there are two kinds of solution respectively for the insatiable and reflexive paradoxes, which are not aligned with Ramsey's distinction between the semantic and set-theoretic paradoxes. Hopefully, now knowing where to look will help find solutions.[33]

References

Badici, Emil 2008: 'The Liar Paradox and the Inclosure Schema'. *Australasian Journal of Philosophy*, 86, pp. 583–96.

Dodgson, Charles [Lewis Carroll] 1871: *Alice Through the Looking-glass*. Project Gutenberg, <http://www.gutenberg.org/ebooks/12>, accessed 23 August 2010.

Eldridge-Smith, Peter 2007: 'Paradoxes and Hypodoxes of Time Travel'. In Lloyd Jones, Campbell, and Wylie 2007, pp. 172–89.

——2008: *The Liar Paradox and its Relatives*, PhD dissertation: Australian National University, <https://digitalcollections.anu.edu.au/handle/1885/49284>.

——2012: 'A Hypodox! A Hypodox! A Disingenuous Hypodox!'. *The Reasoner*, 6, pp. 118–19.

Friedman, Harvey and Michael Sheard 1987: 'An Axiomatic Approach to Self-referential Truth'. *Annals of Pure and Applied Logic*, 33, pp. 1–21.

Gabbay, D. M. and F. Guenthner (eds) 2004: *Handbook of Philosophical Logic,* second edition, vol. 11. Dordrecht: Kluwer Academic Publishers.

Goldstein, Laurence 2000: 'A Unified Solution to Some Paradoxes'. *Proceedings of the Aristotelian Society*, 100, pp. 53–74.

Gupta, Anil and Nuel Belnap 1993: *The Revision Theory of Truth*. Cambridge, MA: MIT Press.

Haack, Susan 1978: *Philosophy of Logics*. Cambridge: Cambridge University Press.

Leitgeb, Hannes 2005: 'Paradox by (non-wellfounded) definition'. *Analysis*, 65, pp. 275–8.

Lloyd Jones, J., P. Campbell, and P. Wylie (eds) 2007: *Art and Time*. Melbourne: Australian Scholarly Publishing.

McDermott, Michael 1977: 'Sets as Open Sentences'. *American Philosophical Quarterly*, 14, pp. 247–53.

McGee, Vann 1991: *Truth, Vagueness, and Paradox: An Essay on the Logic of Truth*. Indianapolis: Hackett Pub. Co.

[33] I would like to thank three anonymous referees and the editor of *Mind* for corrections and insightful comments. Any unsatisfactory issues are my own.

Priest, Graham 1994: 'The Structure of the Paradoxes of Self-reference'. *Mind*, 103, pp. 25–34.

——2000: 'On the Principle of Uniform Solution: A Reply to Smith'. *Mind*, 109, pp. 123–26.

——2002: *Beyond the Limits of Thought*, second edition. Oxford: Oxford University Press.

——2010: 'Badici on Inclosures and the Liar Paradox'. *Australasian Journal of Philosophy*, 88, pp. 359–66.

Quine, Willard V. O. 1951: *Mathematical Logic*, revised edition. Page references to 1962 reprint, New York: Harper & Row.

——1960: 'Variables Explained Away'. *Proceedings of the American Philosophical Society*, 104, pp. 343–7.

——1962: 'The Ways of Paradox', in his 1976, pp. 1–18. Originally published in *Scientific American*, 1962.

——1976: *The Ways of Paradox and Other Essays*, second edition. Cambridge, MA: Harvard University Press.

——1991: 'Two Dogmas in Retrospect'. *Canadian Journal of Philosophy*. 21, pp. 265–74.

——1995a: 'Truth, Paradox and Gödel's Theorem', in his 1995b, pp. 236–41.

——1995b: *Selected Logic Papers,* second edition. Cambridge, MA: Harvard University Press.

Ramsey, Frank P. 1926: 'The Foundations of Mathematics'. *Proceedings of the London Mathematics Society*, 25, pp. 338–84.

Routley, Richard 1980: *Exploring Meinong's Jungle and Beyond*. Canberra: Australian National University.

Sainsbury, R. M. 2009: *Paradoxes*, third edition. Cambridge: Cambridge University Press.

Sorensen, Roy 1998: 'Yablo's Paradox and Kindred Infinite Liars'. *Mind*, 107, pp. 137–55.

——2001: *Vagueness and Contradiction*. Oxford: Clarendon Press.

——2003: *A Brief History of the Paradox*. Oxford: Oxford University Press.

Visser, Albert 2004: 'Semantics and the Liar Paradox'. In Gabbay and Guenthner 2004, pp. 149–240.

Weber, Zach 2010a: 'Explanation and Solution in the Inclosure Argument'. *Australasian Journal of Philosophy*, 88, pp. 353–7.

——2010b: 'Transfinite Numbers in Paraconsistent Set Theory'. *The Review of Symbolic Logic*, 3, pp. 71–92.

Yablo, Stephen 1993: 'Paradox without Self-reference'. *Analysis*, 53, pp. 251–2.

Appendix: Two hypodoxes of satisfaction and membership

Hypodoxes are consistent conundrums otherwise like paradoxes: the semantic value of a hypodox seems *underdetermined* where that of its intuitively related paradoxical statement seems *overdetermined*.[34] The Truth-teller is a paradigm hypodox, given that it seems to be true or false, and there seems no basis for determining which; it is intuitively paired with the Liar statement. There are two hypodoxes of satisfaction corresponding to the two paradoxes, and a small (less than obvious) difference provides some confirmation of the insatiable and reflexive distinction of their associated paradoxes.

I have summarized relevant hypodoxes in the table below. In the first row, the Satisfied hypodox is juxtaposed for partial contrast with Grelling's Autological hypodox. In the second row, the non-circular definition of a Satisfactory hypodox would also require a syntactic function: given such a function, the resultant expression can be compared and contrasted with the set of all self-membered sets.

Insatiable hypodoxes	*Reflexive hypodoxes*
The Satisfied hypodox: $p_e = \langle(x \text{ Sat } p_e)\rangle$ My third favourite predicate just happens to be 'satisfies my third favourite predicate'. Does Crete satisfy my third favourite predicate?	Grelling's Autological hypodox: $s = \langle(x \text{ Sat } x)\rangle$ Does s satisfy s?
The Satisfactory hypodox: $q_e = \{x: x \in q_e\}$ Although this cannot be a contingent identity, the partial self-predication function can be used to yield such an identity.	Russell's hypodoxical set: $v = \{x: x \in x\}$ Is v a member of v?

[34] For discussions of the concept and some comparisons with alternate intuitions see Eldridge-Smith 2007, 2008, or 2012. Some paradoxes have related hypodoxes and others do not seem to. A pairing of truth-functional variations of paradoxes and hypodoxes is set out in Eldridge-Smith 2008. Non-truth-functional Curry paradoxes do not seem to have hypodoxes, as a referee astutely pointed out.

There is a minor difference between insatiable and reflexive hypodoxes. The difference emerges in deriving formulae equivalent to instances of the Sat-schema or Abstraction without actually using those schemata (or the introduction of elimination rules for satisfaction or membership). Consider the following derivation:

(SH1) The Satisfied hypodox

(1) $p_e = \langle (x \text{ Sat } p_e) \rangle$	Premiss
(2) $t \text{ Sat } p_e \leftrightarrow t \text{ Sat } p_e$	Identity, a theorem of SL
(3) $t \text{ Sat } \langle (x \text{ Sat } p_e) \rangle \leftrightarrow t \text{ Sat } p_e$	(1), (2) =E

A similar derivation related to the Satisfactory hypodox and using =E concludes:

$$z \in \{x\colon x \in q_e\} \leftrightarrow z \in q_e$$

Yet for reflexive hypodoxes, the use of '*s*' or '*v*' is an abbreviational convenience. Reflection and sentential logic's theorem of Identity are sufficient to produce the corresponding instances of the Sat-schema for Grelling's Autological hypodox. The effect of Reflection can be illustrated informally using superscripting:

(GH1) Grelling's paradox: a proof illustrating Reflection

(1) $\langle x \text{ Sat}^1 x \rangle \text{ Sat}^1 \langle x \text{ Sat}^1 x \rangle \leftrightarrow (\langle x \text{ Sat}^1 x \rangle \text{ Sat}^1 \langle x \text{ Sat}^1 x \rangle)$	Theorem of Identity
(2) $\langle x \text{ Sat}^1 x \rangle \text{ Sat}^2 \langle x \text{ Sat}^1 x \rangle \leftrightarrow (\langle x \text{ Sat}^1 x \rangle \text{ Sat}^1 \langle x \text{ Sat}^1 x \rangle)$	(1) Reflection

Given the adicity restrictions required for the Sat-schema (as discussed in Sect. 6.3), line (2) above matches an instance of the Sat-schema, not line (1). A similar derivation concludes with a formula that is an instance of Abstraction for Russell's hypodoxical set.

On a final refrain, there is a lack of any principle to determine whether anything satisfies the hypodox related to my new paradox of satisfaction. Anything either satisfies p_e or does not, but there seems to be no principle that determines which. In contrast, the hypodoxical issue for 'autological' is localized to reflexive cases, such as whether 'autological' itself is autological.

