The Consistency of the Naive(?) Theory of Properties

Hartry Field*
New York University

June 11, 2003

1 Introduction

According to the Naive Theory of Properties, for every predicate $\Theta(x)$ there is a corresponding property $\lambda x\Theta(x)$. Moreover, this property $\lambda x\Theta(x)$ is instantiated by an object o if and only if $\Theta(o)$. More generally, the Naive Theory involves the following "Naive Comprehension Schema":

$$\text{(NC) } \forall u_1...\forall u_n\exists y[Property(y)\wedge\forall x(x\,\text{instantiates}\,y\leftrightarrow\Theta(x,u_1...u_n))].$$

This Naive Theory of Properties has many virtues, but it seems to have been shattered by (the property version of) Russell's Paradox.

"Seems to" have been shattered? There's no doubt that it *was* shattered, if we presuppose full classical logic. Let us use the symbol '$\in$' to mean "instantiates". The Russell Paradox involves the Russell property R corresponding to the predicate 'does not instantiate itself'. So according to the Naive Theory, $\forall x[x\in R\leftrightarrow\neg(x\in x)]$. Therefore in particular,

$$(*)\ R\in R\leftrightarrow\neg(R\in R).$$

But (*) is classically inconsistent.

There are two solution routes (routes for modifying the Naive Theory) within classical logic. The first says that for certain predicates, such as 'does not instantiate itself', there is no corresponding property. The second says that there is one, but it isn't instantiated by what you might think: there are either (i) cases where an object o has the property $\lambda x\Theta(x)$ even though $\neg\Theta(o)$, or (ii) cases where an object o doesn't have the property $\lambda x\Theta(x)$ even though $\Theta(o)$. In particular, when $\Theta(x)$ is 'does not instantiate itself', the Russell property is either of sort (i) or sort (ii). This second solution route subdivides into three variants. One variant commits itself to a solution of type (i): the Russell

*This is a preprint of an article accepted for publication in *The Philosophical Quarterly*, Jan. 2004, ©Basil Blackwell.

property instantiates itself, but nonetheless has the property of not instantiating itself. A second variant commits itself to a solution of type (ii): the Russell property doesn't instantiate itself, but nonetheless fails to have the property of not instantiating itself. A third variant hedges: it says that the Russell property is either of sort (i) or sort (ii), but refuses to say which.

These four classical theories–the three variants that admit the existence of the Russell property and the one that denies it–all seem to me problematic. (In the *prima facie* analogous case of sets, I take the approach that denies the existence of "the Russell set" to be quite *un*problematic. But I take properties to be very different from sets in this regard, for reasons to be discussed in the final section.) In my view we need a different sort of solution route, and it must inevitably involve a weakening of classical logic. It is the aim of this paper to provide one.

The idea of weakening logic to avoid the Russell Paradox is not new, but the proposal presented here is unlike many in that it saves the full Naive Comprehension Schema in the form stated above: it saves it not only from the Russell Paradox (which is relatively easy) but from far more virulent forms of paradox (such as the Curry Paradox and its many extensions). I know of no other ways of saving Naive Comprehension in as strong or as natural a logic.

2 Background

If we are going to weaken classical logic to get around the Russell paradox (along with others), it is useful to look at how it is that (*) leads to contradiction in classical logic; that way, we'll know which steps in the argument for contradiction might be denied. (Actually, one well-known approach accepts contradictions, in the sense of assertions of form $A \wedge \neg A$, and while I do not favor it, I want my initial discussion to recognize it as an alternative. For that reason, let me stipulate that a theory is to be called *inconsistent* if it implies, not just a contradiction in the above sense, but anything at all: the existence of Santa Claus, the omniscience of George Bush about matters of quantum field theory, you name it. So even those who accept "contradictions" won't want their theory to be "inconsistent", in the way I am now using these terms.) With this terminology in mind, here are the main steps in an obvious argument that (*) is inconsistent:

(1) (*) and $R \in R$ together imply the contradiction

(**) $(R \in R) \wedge \neg(R \in R)$,

since the first conjunct is one of the premises and the second conjunct follows by modus ponens.

(2) Analogously, (*) and $\neg(R \in R)$ together imply (**).

(3) So by disjunction elimination, (*) and $(R \in R) \vee \neg(R \in R)$ together imply the contradiction (**).

But (4) $(R \in R) \vee \neg(R \in R)$ is a logical truth (law of excluded middle), so (*) *all by itself* implies the contradiction (**).

(5) Anything that implies a contradiction implies anything whatever, and hence is inconsistent in the most obviously odious sense of the term.

That's the argument,[1] and obviously several different ways of restricting classical logic so as to avoid it are possible. (I take it that the argument that (NC) implies (*) involves nothing in the least controversial, so that anyone who wants to retain the Naive Comprehension will certainly want to retain (*).) I will simply state my preferred approach, without arguing that it is best: in my view, the most appealing way to weaken classical logic so as to evade the argument that (*) leads to inconsistency is to restrict the law of excluded middle, thereby undermining step (4). Disjunction elimination can be retained (even in the strong sense used in Step (3), i.e. allowing side formulas). So can "the odiousness of contradictions" assumed in Step (5).[2]

Unfortunately, restricting excluded middle falls far short of giving an adequate theory. In the first place, though restricting the law of excluded middle blocks *the above argument* for the inconsistency of $R \in R \leftrightarrow \neg(R \in R)$, it is by no means obvious that there is a satisfactory logic without unrestricted excluded middle in which that biconditional can be maintained. It is still less obvious that there is a satisfactory such logic in which the full Naive Theory of Properties can be maintained. Let me explain.

First of all, the most obvious ways to deal with the paradoxes in logics without excluded middle (e.g. the set-theoretic adaptation of the Kleene version of Kripke's [4] "fixed point" approach to the semantic paradoxes)[3] do not vindicate (NC), nor do they even vindicate its weak consequence (*). The reason is that they don't contain an appropriate conditional (or biconditional).

Indeed, the main issues involved in showing the consistency of the Naive

[1] I've been a bit sloppy about use and mention, since I've defined R to be a property, but appear to have spoken of a sentence (*) that contains it. There are several ways this could be made right. One is to work in a language where we have a property-abstraction operator, so that we could name R in the language; then that name would be used in (*). A second is to replace the 'R' in (*) with a free variable y: then the argument in the text goes over to an argument that formulas of form $y \in y \leftrightarrow \neg(y \in y)$ imply contradictions, so their existential generalizations do too, and (NC) implies such an existential generalization. A third involves the introduction of a convention of "parameterized formulas": pairs of formulas and assignments of objects to their free variables. Then (*) is simply a convenient notation for the pair of '$y \in y \leftrightarrow \neg(y \in y)$' and an assignment of R to 'y', and what appears in the text is a literally correct derivation involving parameterized formulas. Do things however you like.

[2] Of course, if you can evade the argument in a logic $£_1$ that contains the "odiousness of contradictions" rule $A \wedge \neg A \models B$, you can equally evade it in a logic $£_2$ just like $£_1$ but which is "paraconsistent" in that the rule $A \wedge \neg A \models B$ is dropped. But since classical laws like excluded middle that are absent from $£_1$ will be absent from $£_2$ as well, this has no evident advantages. What is potentially more interesting is using a paraconsistent logic in which we keep laws absent from $£_1$, such as excluded middle. For some skeptical remarks on the possibility of getting anything like Naive Property Theory in an interesting paraconsistent logic of this type, see [1].

[3] Such an adaptation of the Kleene variant of Kripke's approach is in effect given in Maddy [5], as a theory of proper classes. I will discuss the use of the theory to be given in this paper in connection with proper classes in the final section. In my opinion, the presence of (NC) is not only needed for property theory generally, it also makes for a more adequate theory of proper classes.

Theory center on the problem of finding an adequate treatment of the →. (And hence the ↔: I'll assume that $A \leftrightarrow B$ means $(A \rightarrow B) \wedge (B \rightarrow A)$.) Even if our goal were limited to the consistent assertion of the biconditional (*), that would pretty much rule out our defining $A \rightarrow B$ in terms of the other connectives in the manner familiar from classical logic, viz. $\neg A \vee B$. For on that "material conditional" reading of '→', (*) amounts to

$$[\neg(R \in R) \vee \neg(R \in R)] \wedge [\neg\neg(R \in R) \vee (R \in R)].$$

Assuming distributivity and a few other simple laws, this is equivalent to a disjunction of the classical inconsistencies $(R \in R) \wedge \neg(R \in R)$ and $\neg(R \in R) \wedge \neg\neg(R \in R)$. If we assume double-negation elimination, that's in effect just the simple contradiction $(R \in R) \wedge \neg(R \in R)$; and even if double-negation elimination isn't assumed, a disjunction of contradictions seems just as inconsistent as a single contradiction. So if we put aside the paraconsistent approaches mentioned in note 3, it's clear that we cannot in general interpret $A \rightarrow B$ as $\neg A \vee B$ if we want to retain even (*). And on the paraconsistent approaches the "material conditional" reading of → seems inappropriate on different grounds: that reading invalidates modus ponens.[4]

The first problem about getting a decent conditional, then, is licensing the assertion of (*). But there are plenty of "logics of →" that avoid that problem while still being inadequate to the Naive Theory, for the full Comprehension Schema (NC) is not consistently assertable in them. Indeed, many of these logics fail to handle a close analog of Russell's paradox due to Curry. The problem is this: (NC) implies the existence of a Curry property K, for which $\forall x[x \in K \leftrightarrow (x \in x \rightarrow \perp)]$, where $\perp$ is any absurdity you like. So $K \in K \leftrightarrow (K \in K \rightarrow \perp)$; that is,

(i) $K \in K \rightarrow (K \in K \rightarrow \perp)$

and

(ii) $(K \in K \rightarrow \perp) \rightarrow K \in K$.

But in many logics of → we have the contraction rule $A \rightarrow (A \rightarrow B) \models A \rightarrow B$, on which (i) implies

(i*) $K \in K \rightarrow \perp$.

But this with (ii) leads to $K \in K$ by modus ponens; and another application of modus ponens leads from that and (i*) to $\perp$.

Unless we restrict modus ponens (and it turns out a very drastic restriction of it would be required), we need to restrict the contraction rule. This requires further restrictions on the logic as well. For instance, given that we're keeping modus ponens in the form $A, A \rightarrow B \models B$, we certainly have $A, A \rightarrow (A \rightarrow B) \models B$ simply by using modus ponens twice; so to prevent contraction, we certainly can't have the generalized →-introduction meta-rule that allows passage from $\Gamma, A \models B$ to $\Gamma \models A \rightarrow B$. Indeed, even the weaker version which allows the

[4] Although it is important not to interpret → as the material conditional, the theory that I will advocate does posit a close relation between the two: while $(A \rightarrow B) \leftrightarrow (\neg A \vee B)$ is not a logical truth, it is a logical consequence of the premises $A \vee \neg A$ and $B \vee \neg B$. In other words, it is only in the context of a breakdown in the law of excluded middle that the divergence between the → and the material conditional emerges.

inference only when Γ is empty should be given up: it is the obvious culprit in an alternative form of the Curry paradox.

It turns out, though, that the difficulty in finding an adequate treatment of the '$\rightarrow$' is not insuperable, and that the Naive Comprehension Principle (NC) can be maintained; indeed, it can be maintained in a logic that, though not containing excluded middle or the contraction rule, is not altogether unnatural or hopelessly weak.[5] The aim of this paper is to show this.[6] Whether the theory should still count as "naive" when the logic is altered in this way is a question I leave to the reader.

It is worth emphasizing that though the law of excluded middle will need restriction, there is no need to give it up entirely: it can be retained in various restricted circumstances. For instance, the notion of property is normally employed in connection with a "base language" L that does not talk of properties; we then expand L to a language L^+ that allows for properties, including but not limited to properties of things talked about in L. (It is not limited to properties of things talked about in L because it will also include properties of properties: indeed, it is some of these that give rise to the apparent paradoxes.) It is within the ground language L that most of mathematics, physics, and so forth takes place; and the theory advocated here does not require any limitation of excluded middle in these domains, because as long as we restrict our quantifiers to the domain of the ground language we can retain full classical logic. We can also retain full classical logic in connection with those special ("rank 1") properties that are explicitly limited so as to apply to non-properties; and to those special ("rank 2") properties that are explicitly limited so as to apply to non-properties and rank 1 properties; and so on. Where excluded middle cannot be assumed is only in connection with certain properties that do not appear anywhere in such a rank hierarchy, like the Russell Property and the Curry Property (though for other such properties, e.g. those whose *complement* appears in the rank hierarchy, excluded middle is also unproblematic). Even for the "problematic" properties, there is no need to give up excluded middle for claims about property *identity*; it is only when it comes to claims about *instantiation* of problematic properties that excluded middle will not be able to be assumed in general.

I don't know if the theory here can be adapted to a theory of "naive sets", by adding an axiom or rule of extensionality; I will have a bit to say about this in the final section, including a discussion of why the matter is much less pressing for sets than for properties. But if it is possible to develop a theory of naive sets, it seems unlikely that we would be able to maintain excluded middle for identities between naive sets (e.g., between the empty set and $\{x|x = x \wedge K \in K\}$, where K is the "Curry set", defined in analogy with the Curry property). Because of

[5] For instance, the conditional obeys contraposition in the strong form $\models (\neg A \rightarrow \neg B) \rightarrow (B \rightarrow A)$. Also when $\models A \leftrightarrow B$ and C_B results from C_A by subsituting B for one or more occurrences of A, then $\models C_A \leftrightarrow C_B$; so (NC) yields that $y \in \lambda x \Theta(x)$ is everywhere intersubstitutable with $\Theta(y)$, even within the scope of a conditional.

[6] The approach I'll be giving is an adaptation of the approach to the semantic paradoxes developed in [2].

this, a "naive set theory", if possible at all, would have an importantly different character from the naive property theory about to be developed.

3 The Goal

I've said I want a consistent naive theory of properties, but actually what I want is a bit stronger than mere consistency. It's time to start being a bit more precise.

Let L be any first order language with identity. Since I won't want to identify $A \rightarrow B$ with $\neg A \vee B$, it is necessary to assume that $\rightarrow$ is a primitive connective, along with $\neg$, $\wedge$ and/or $\vee$, and $\forall$ and/or $\exists$. And to avoid annoying complications about how to extend function symbols when we add to the ontology, I'll assume that L contains no function symbols (except perhaps for 0-place ones, i.e. individual constants). L can be taken to be a language for mathematics, or physics, or whatever you like other than properties. (So it shouldn't contain the terms 'Property' or '$\in$' in the senses to be introduced. If it contains these terms in other senses–e.g., '$\in$' for membership among the iterative sets of standard set theory–then imagine these replaced by other terms.)

Let L^+ result from L by adding a new 1-place predicate '*Property*' and a new 2-place predicate '$\in$' meaning 'instantiates'. For any formula A of L, let A^L be the formula of L^+ obtained from A by restricting all bound occurrences of any variable z by the condition '$\neg Property(z)$'. Let T be any theory in the language L. "Naive Property Theory over T" is the theory T^+ that consists of the following non-logical axioms:

(I) A^L, for any A that is a closure of a formula that follows from T

(II) $\forall x \forall y[x \in y \rightarrow Property(y)]$

(III) $\forall u_1 ... \forall u_n \exists y[Property(y) \wedge \forall x(x \in y \leftrightarrow \Theta(x, u_1 ... u_n))]$, where $\Theta(x, u_1 ... u_n)$ is any formula of L^+ in which y is not free.

((III) is just (NC).) Then a minimal goal is to show that in a suitable logic, the theory T^+ consisting of (I)-(III) is always consistent as long as T itself is consistent. Note that if T is itself a classical theory, i.e. is closed under classical consequence, then "Naive Property Theory over T" effectively keeps classical logic among sentences of form A^L, even though its official logic is weaker: for if A_1,...,A_n are formulas of L that classically entail B, then $A_1 \wedge A_2 \wedge ... A_n \rightarrow B$ is in T, so $[(\forall u_1, ..., u_k)(A_1 \wedge A_2 \wedge ... A_n \rightarrow B)]^L$ is in T^+, and this is the same as $(\forall u_1, ..., u_k)[\neg Property(u_1) \wedge ... \wedge \neg Property(u_k) \rightarrow (A_1^L \wedge A_2^L \wedge ... A_n^L \rightarrow B^L)]$.

The *minimal* goal is to show that T^+ is consistent whenever T is, but I actually want something slightly stronger: I want to introduce a kind of multi-valued model for L^+ (infinite-valued, in fact), and then prove

> (G) For each classical model M of L, there is at least one model M^+ of L^+ that validates (II) and (III) and has M as its reduct;

where to say that M is the reduct of M^+ means roughly that when you restrict the domain of M^+ to the things that don't satisfy '*Property*' (and forget about the assignments to '*Property*' and to '$\in$') then what you are left with is just

M.[7] Since the connectives of L^+ will reduce to their classical counterparts on the reduct, the fact that M is the reduct of M^+ will guarantee the validity of Axiom Schema (I); so if M satisfies T, M^+ satisfies T^+.

There is good reason why (G) says 'at least one' rather than 'exactly one': we should expect that most or all models of T^+ can be extended to models that contain new properties but leave the property-less reduct unchanged. The proof that I will give yields the minimal M^+ for a given M, but extensions of the model with the same reduct could easily be given.

I will prove (G) in a classical set-theoretic metalanguage, so anyone who is willing to accept classical set-theory should be able to accept the coherence of the non-classical property theory to be introduced.

4 The Semantic Framework

The goal just enunciated calls for developing a model-theoretic semantics for L^+ in a classical set-theoretic metalanguage. The semantics will be multi-valued: in addition to (analogs of) the usual two truth values there will be others, infinitely many in fact.

4.1 The Space of Values

My approach to achieving the goal is an extension of the Kripke-style approach previously mentioned, but it needs to be substantially more complicated because of the need for a reasonable conditional.

One complication has to do with method of proof: the new conditional is not "monotonic" in the sense of Kripke, which means that we cannot make do merely with the sort of fixed point argument that is central to his approach (though such a fixed point argument will play an important role in this approach too).

The other complication is that the semantic framework itself should be generalized: whereas Kripke uses a 3-valued semantics, I will use a model theory in which sentences take on values in a subspace W^Π of the set F^Π of functions from $Pred(\Pi)$ to $\{0, \frac{1}{2}, 1\}$, where Π is an initial ordinal (ordinal with no predecessor of the same cardinality) that is greater than ω, and where $Pred(\Pi)$ is the set of its predecessors.[8] (I don't fix on a particular value of Π at this point, because I will later impose further minimum size requirements on it.)

[7] The reason for the 'roughly' in the definition of 'reduct' is that M is a classical model, whereas M^+ will be multi-valued; so its reduct will have to assign objects that live in the larger space of values. Nonetheless, the space of values will contain two rather special ones, to be denoted $\mathbf{1}$ and $\mathbf{0}$, and we can take 'A has value $\mathbf{1}$ in M^+' and 'A has value $\mathbf{0}$ in M^+' to correspond to 'A is true in M' and 'A is false in M', when A is in L. The reduct of M^+ won't strictly be M, but it will be the $\{\mathbf{0},\mathbf{1}\}$-valued model that corresponds to M in the obvious way.

[8] When I presented the analog of this for the semantic paradoxes in [2], I did not explicitly introduce this new space of semantic values (since I hadn't yet thought of the matter in that way); but the ideas seem to me clearer with this space of values made explicit.

Which subset of F^Π do I choose as my W^Π? If ρ is a non-zero ordinal less than Π, call a member f of F^Π *ρ-cyclic* if for all β and σ for which $\rho\cdot\beta+\sigma < \Pi$, $f(\rho\cdot\beta+\sigma) = f(\sigma)$; and call it *cyclic* if there is a non-zero ρ less than Π such that it is ρ-cyclic. Call it *regular* if in addition to being cyclic, it satisfies the condition that it is either one of the constant functions $\mathbf{0}$ and $\mathbf{1}$ (which map everything into 0 or map everything into 1) or else maps 0 into $\frac{1}{2}$. Then W^Π consists of the regular functions from $Pred(\Pi)$ to $\{0, \frac{1}{2}, 1\}$. (Once we've found a suitable method of assigning values in W^Π to sentences, then the valid inferences among sentences will be taken to be those inferences that are guaranteed to preserve the value $\mathbf{1}$.)

A few properties of W^Π are worth noting.

- It has a natural partial ordering: $f \preceq g$ iff $(\forall\alpha < \Pi)(f(\alpha) \leq g(\alpha))$. The ordering has a minimum $\mathbf{0}$ and maximum $\mathbf{1}$. And the ordering is not total: for instance, the constant function $\boldsymbol{\frac{1}{2}}$ is incomparable with the function that has value $\frac{1}{2}$ at limit ordinals, 0 at odd ordinals, and 1 at even successors.

- For each $f \in W^\Pi$ define $f^\bigstar$ to be the function for which $f^\bigstar(\alpha) = 1 - f(\alpha)$ for each α. Then $f^\bigstar$ will be in W^Π too. Moreover, the operation $^\bigstar$ is a symmetry that switches $\mathbf{0}$ with $\mathbf{1}$, leaving the constant function $\boldsymbol{\frac{1}{2}}$ fixed.

- For any nonempty subset S of W^Π that has cardinality less than that of Π, define $\curlywedge(S)$ to be the function whose value at each α is the minimum of $\{f(\alpha) | \alpha \in S\}$, and $\curlyvee(S)$ to be the function that analogously gives the pointwise maximum. Then $\curlywedge(S)$ and $\curlyvee(S)$ are in W^Π;[9] and clearly, they are the meet and join of S with respect to the partial ordering.

- For any S, $\curlyvee(S)$ is $\mathbf{1}$ only if $\mathbf{1} \in S$; that holds because if $\mathbf{1} \notin S$, then $f(0) < 1$ for each f in S. This is important: it will ensure that the logic that results will obey the meta-rules of $\vee$-elimination and $\exists$-elimination.

Observe also that if $f(0)$ is $\frac{1}{2}$ and $f \in W^\Pi$, then f assumes the value $\frac{1}{2}$ arbitrarily late, viz. at all right-multiples of ρ_f. (By ρ_f, I mean the smallest ρ for which f is ρ-cyclic.) Also, note that for any f and g in W^Π, there are $\rho < \Pi$ such that *both* f and g are ρ-cyclic: any common right-multiple of ρ_f and ρ_g will be one. One consequence of this is that if there are $\beta < \Pi$ for which $f(\beta) < g(\beta)$ (alternatively, $f(\beta) \leq g(\beta)$), then for any $\alpha < \Pi$, there are β in the open interval from α to Π for which $f(\beta) < g(\beta)$ (alternatively, $f(\beta) \leq g(\beta)$). And that implies that

- $f \preceq g$ is equivalent to the *prima facie* weaker claim that there is an α (less than Π) such that for all β greater than α (and less than Π), $f(\beta) \leq g(\beta)$.

[9] For $min(S)$, this is trivial if one of the members of S is $\mathbf{0}$ or if S is $\{\mathbf{1}\}$. Otherwise, consider the nonempty subset of members of W^Π that are of type (ii), and let ρ_S be the smallest ordinal that is a right-multiple of all the ρ_f for $f \in S$; by the cardinality restriction on S, this is less than Π (and at least 2). Moreover, all members of S ρ_S-cycle, so $min(S)$ ρ_S-cycles (and so $\rho_{min(S)} \leq \rho_S$).

Similarly, if we define a (quite strong) strict partial ordering $\prec\prec$ by $f \prec\prec g$ iff either ($f = \mathbf{0}$ and $g \succeq \mathbf{\frac{1}{2}}$) or ($f \preceq \mathbf{\frac{1}{2}}$ and $g = \mathbf{1}$), then

- $f \prec\prec g$ is equivalent to the *prima facie* weaker claim that there is an α (less than Π) such that for all β greater than α (and less than Π), $f(\beta) < g(\beta)$.

For if the consequent of this holds, then pick β to be a common right-multiple of ρ_f and ρ_g greater than α; since $f(\beta) < g(\beta)$, at least one of $f(\beta)$ and $g(\beta)$ isn't $\frac{1}{2}$, so at least one of $f(0)$ and $g(0)$ isn't $\frac{1}{2}$, so at least one of f and g is in $\{\mathbf{0}, \mathbf{1}\}$; and the rest is obvious.

The results just sketched are the keys to proving a final feature of the space W^{Π}, that it is that it is closed under the following operation $\implies$:

$(f \implies g)(0)$ is
1 if for some $\beta < \Pi$, and any γ such that $\beta \leq \gamma < \Pi$, $f(\gamma) \leq g(\gamma)$;
0 if for some $\beta < \Pi$, and any γ such that $\beta \leq \gamma < \Pi$, $f(\gamma) > g(\gamma)$;
$\frac{1}{2}$ otherwise;
and if $\alpha > 0$, $(f \implies g)(\alpha)$ is
1 if for some $\beta < \alpha$, and any γ such that $\beta \leq \gamma < \alpha$, $f(\gamma) \leq g(\gamma)$;
0 if for some $\beta < \alpha$, and any γ such that $\beta \leq \gamma < \alpha$, $f(\gamma) > g(\gamma)$;
$\frac{1}{2}$ otherwise.

(The value $\frac{1}{2}$ can occur only at 0 and at limits.) Note that the exceptional treatment of 0 in effect turns the domain of the functions in W^{Π} into a "transfinite circle", in which 0 is identified with Π. And we clearly have

- $f \implies g$ is $\mathbf{1}$ if and only if $f \preceq g$; and $f \implies g$ is $\mathbf{0}$ if and only if $f \succ\succ g$.

Why is W^{Π} closed under $\implies$? Since $\mathbf{1}$ and $\mathbf{0}$ are in W^{Π}, we need only show that when neither $f \preceq g$ nor $f \succ\succ g$ then $f \implies g$ is regular. Let ρ be the smallest non-zero ordinal for which both f and g ρ-cycle, and let ρ^* be $\rho \cdot \omega$. I claim that $f \implies g$ is ρ^*-cyclic, that is, for any $\sigma < \rho^*$, the value of $(f \implies g)(\rho^* \cdot \delta + \sigma)$ is independent of δ; and that when σ is 0, $(f \implies g)(\rho^* \cdot \delta + \sigma)$ is $\frac{1}{2}$. Case 1: $\sigma > 0$. Then $(f \implies g)(\rho^* \cdot \delta + \sigma) = 1$ iff $(\exists \beta < \rho^* \cdot \delta + \sigma)(\forall \gamma)(\beta \leq \gamma < \rho^* \cdot \delta + \sigma \supset f(\gamma) \leq g(\gamma))$ iff $(\exists \beta)(\rho^* \cdot \delta \leq \beta < \rho^* \cdot \delta + \sigma)(\forall \gamma)(\beta \leq \gamma < \rho^* \cdot \delta + \sigma \supset f(\gamma) \leq g(\gamma))$; but that's independent of δ since f and g are (ρ-cyclic and hence) ρ^*-cyclic. Similarly for the δ-independence of the condition for $(f \implies g)(\rho^* \cdot \delta + \sigma) = 0$. Case 2: $\sigma = 0$. We need that $(f \implies g)(\rho^* \cdot \delta) = \frac{1}{2}$ for all δ. The reason is that for any $\alpha < \rho^* \cdot \delta$ (i.e. $\alpha < \rho \cdot (\omega \cdot \delta)$), there is an ζ such that $\alpha < \rho \cdot \zeta < \rho \cdot (\zeta + 1) < \rho \cdot (\omega \cdot \delta)$; and (since neither $f \preceq g$ nor $f \succ\succ g$) there are bound to be β in the interval from $\rho \cdot\ \zeta$ to $\rho \cdot (\zeta + 1)$ (lower bound included) where $f(\beta) > g(\beta)$ and others where $f(\beta) \leq g(\beta)$, so $(f \implies g)(\rho^* \cdot \delta) = \frac{1}{2}$.

The operation $\implies$ just specified has some rather nice properties. I've already noted the conditions under which it takes the values $\mathbf{1}$ and $\mathbf{0}$. In addition:

- When f and g are in $\{\mathbf{0}, \mathbf{1}\}$ then $f \implies g$ is identical to the value of the material conditional, $\curlyvee\{f^{\bigstar}, g\}$.

Since I will be using $\Longrightarrow$ to evaluate the conditional, this will mean that the conditional reduces to the material conditional when excluded middle is assumed for antecedent and consequent.

It is beyond the present scope to investigate the laws governing $\Longrightarrow$ (though this is important since it will determine which inferences involving → are valid); for that, see [2].

It's worth making explicit that if $f \Longleftrightarrow g$ is defined in the obvious way (as $\curlywedge\{f \Longrightarrow g, g \Longrightarrow f\}$), then if $\alpha > 0$, $(f \Longleftrightarrow g)(\alpha)$ is

1 if for some $\beta < \alpha$, and any γ such that $\beta \leq \gamma < \alpha$, $f(\gamma) = g(\gamma)$;

0 if for some $\beta < \alpha$, and any γ such that $\beta \leq \gamma < \alpha$, $f(\gamma) \neq g(\gamma)$;

$\frac{1}{2}$ otherwise.

(And analogously for $(f \Longleftrightarrow g)(0)$: use Π in place of α on the right hand side.)

4.2 W^Π-Models

Having noted these features of the space W^Π of values, we can easily define models based on this space: W^Π-models. I will take a W^Π-*model* for a language to consist of a domain D *of cardinality less than* Π, an assignment to each individual constant c of a member $den(c)$ of D, and an assignment to each n-place predicate of a function p^* from D^n to W^Π (where D^n is the set of n-tuples of members of D). A W^Π-*valuation* for a language will consist of a W^Π-model together with a function s assigning objects in the domain of the model to the variables of the language. Given any valuation with assignment function s and any term t (individual constant or variable), let $den_s(t)$ be $den(t)$ if t is an individual constant, $s(t)$ if t is a variable.

Given a W^Π-valuation with assignment function s, we assign values in W^Π to formulas as follows:

$||p(t_1, ...t_n)||_s$ is $p^*(den_s(t_1), ..., den_s(t_n))$, which in the future I'll also write as $p^*_{den_s(t_1),...,den_s(t_n)}$;

$||\neg A||_s$ is $(||A||_s)^\bigstar$;

$||A \wedge B||_s$ is $\curlywedge\{||A||_s, ||B||_s\}$;

$||A \vee B||_s$ is $\curlyvee\{||A||_s, ||B||_s\}$;

$||\forall x A||_s$ is $\curlywedge\{||A||_{s'} \mid s'$ differs from s except perhaps in what is assigned to the variable $x\}$;

$||\exists x A||_s$ is $\curlyvee\{||A||_{s'} \mid s'$ differs from s except perhaps in what is assigned to the variable $x\}$;

$||A \rightarrow B||_s$ is $\Longrightarrow (||A||_s, ||B||_s)$.

Note that for the quantifier clause to make sense in general, it is essential that the domain of quantification have lower cardinality than Π. But this is no real restriction, it's simply that if you want to consider models of large cardinality you have to choose a large value of Π. (Recall the goal, (G): we want a strong form of consistency in which for any classical starting model M for the base language L, there is a non-classical model M^+ in L^+ that has M as its reduct. There is no reason why the space of values used for M^+ can't depend on the cardinality of M.) So I will stipulate that the non-classical model will be a

W^{Π}-model for some initial ordinal Π of cardinality greater than that of M (as well as being greater than ω). The M^+ shortly to be described will have a cardinality that is the maximum of the cardinalities of M and of ω, so this restriction will suffice for the quantifier clause to be well-defined.

I've written the valuation rules for ordinary formulas, but in the future I will adopt the convention of using parameterized formulas in which we combine the effect of the formula and the assignment function in our notation by plugging a metalinguistic name for an object assigned to a variable in for free occurrences of the variable in the displayed formula; that will allow me to drop the subscript s, and simplify the appearance of other clauses. For instance, the clauses for atomic formulas and universal quantifications become

$||p(o_1, ...o_n)||$ is the function $f_{p,o_1,...o_n}$ that takes any α into $p^*(o_1, ..., o_n)$;
$||\forall xA||$ is the function $\curlywedge\{||A(o)|| \mid o \in D\}$.

(Sometimes I'll make the parameters explicit, e.g.

For all $o_1, ..., o_n$, $||\forall xA(o_1, ..., o_n)||$ is the function $\curlywedge\{||A(o, o_1, ..., o_n)|| \mid o \in D\}$,

but the absence of explicit parameters should not be taken to imply that there are no parameters in the formula.)

5 A Model for Naive Property Theory

5.1 The Basics

The next step is to specify the particular model to be used for naive property theory. Recall that I'm imagining that we are given a model M for the base language L. We can assume without real loss of generality that $|M|$ (the domain of M) doesn't contain formulas of L^+, or n-tuples that include such formulas; for if the domain does contain such things, we can replace it with an isomorphic copy that doesn't. With this done, let E_0 be $|M|$. For each natural number k, we define a set E_{k+1} of *ersatz properties of level* $k+1$. A member of E_{k+1} is a triple consisting of a formula of L^+, a distinguished variable of L^+, and a function that assigns a member of $\cup\{E_j | j \leq k\}$ to each free variable of L^+ other than the distinguished one, meeting the condition that if $k > 0$ then at least one element of E_k is assigned.[10] If $\Theta(x, u_1, ..., u_n)$ is the formula and x the distinguished variable and $o_1, ..., o_n$ the objects assigned to $u_1, ..., u_n$ respectively, I'll use the notation $\lambda x\Theta(x, o_1, ..., o_n)$ for the ersatz property. Let E be the union of all the E_k for $k \geq 1$, and let $|M^+|$ be $|M| \cup E$. (The cardinality of $|M^+|$ is thus the same as that of $|M|$, when $|M|$ is infinite, and is $\aleph_0$ when M is finite.) The only terms of L^+ are the individual constants of L; they get the same values in M^+ as in M.

I hope it's clear that the fact that I'm taking the items in the domain to be constructed out of linguistic items does not commit me to viewing properties as linguistic constructions; the point of the model is simply to give a strong form

[10] The exception for $k = 0$ is needed only for formulas that contain no free variables beyond the distinguished one.

of consistency proof (i.e., to satisfy Goal (G)), and this is the most convenient way to do it.

Putting aside the unimportant issue of the nature of the entities in the domain, the domain does have a very special feature: all the properties in the model are ultimately generated (in an obvious sense I won't bother to make precise) from the entities in the ground model by the vocabulary of the ground model; so that the model contains the minimal number of properties that are possible, given the ground model. It is useful to consider such a special model for doing the consistency proof for naive property theory, but not all models of naive property theory will have this form (as is obvious simply from the fact that if we were to add new predicates to the ground model before starting the construction, we would generate new properties).

To complete the specification of M^+ we must specify an appropriate Π, and then assign to each n-place atomic predicate p an "W^Π-extension": a function p^* that takes n-tuples of members of $|M^+|$ into W^Π. I have already said that I would take Π to be the initial ordinal for a cardinal greater than the cardinality of $|M^+|$; a further stipulation will become necessary, but let us wait on that. As for the predicates, much of what we must say is obvious. If p is a predicate in L other than '$=$', and $o_1, ..., o_n$ are in M^+, we let $p^*_{o_1,....,o_n}$ be $\mathbf{1}$ if $\langle o_1, ..., o_n\rangle$ is in the M-extension of p, $\mathbf{0}$ otherwise. (So it's $\mathbf{0}$ if any of the o_i are in E.) We let $Property^*_o$ be $\mathbf{1}$ when o is in E, $\mathbf{0}$ otherwise. And we let $=_{o_1,o_2}$ be $\mathbf{1}$ when o_1 is the same object as o_2, and $\mathbf{0}$ otherwise. These stipulations obviously suffice to make M the reduct of M^+. Because of this, and the fact that the function assigned to each connective *including the conditional* reduces to its classical counterpart when confined to the set $\{\mathbf{0}, \mathbf{1}\}$, we get (by an obvious induction on complexity) that for any sentence A of L (or any formula A of L and any assignment function s that assigns only objects in $|M|$), the value of A^L in M^+ (relative to s) will be $\mathbf{1}$ when the value of A (relative to s) in M is 1, and will be $\mathbf{0}$ when the value of A (relative to s) in M is 0. Each instance of Axiom Schema (I) therefore gets value $\mathbf{1}$.

This leaves only '$\in$'. One desideratum should obviously be that when o_2 is in the ground model $|M|$, then $\in^*_{o_1,o_2}$ is $\mathbf{0}$. This will suffice for giving value $\mathbf{1}$ to Axiom (II).

The difficult matter, of course, is figuring out how to complete the specification of the W^Π-extension of '$\in$', in such a way as to validate Axiom Schema (III). This will be the subject of the next four subsections.

5.2 The Difficulty: How Do '$\in$' and '$\rightarrow$' Interact?

The main problem in constructing an interpretation for membership statements is due to the presence in the language of the conditional $\rightarrow$. Just to get a feeling for what might be involved here, consider a very simple case: the ordinary Curry property K. This is $\lambda x(x \in x \rightarrow \bot)$, where $\bot$ is some sentence with value $\mathbf{0}$, say $\exists y(y \neq y)$. What function $f_{K\in K}$ should serve as $\in^*_{K,K}$, and hence as $||K \in K||$, i.e. $||K \in \lambda x(x \in x \rightarrow \bot)||$? Since we want (III) to be valid, that had better be the same as $||K \in K \rightarrow \bot||$. That is, we want the function $f_{K\in K}$ to be identical to the function $f_{K\in K} \implies \mathbf{0}$.

But how do we get that to be the case? The first thing we want to know is, what is $f_{K\in K}(0)$? The rules tell us that it is

1 if $(\exists\beta < \Pi)(\forall\gamma)[\beta \leq \gamma < \Pi \supset f_{K\in K}(\gamma) = 0]$;
0 if $(\exists\beta < \Pi)(\forall\gamma)[\beta \leq \gamma < \Pi \supset f_{K\in K}(\gamma) > 0]$;
$\frac{1}{2}$ otherwise.

Evidently, we can't know the value of $f_{K\in K}(0)$ until we know the values of $f_{K\in K}(\alpha)$ for higher α. But finding out the values of $f_{K\in K}(\alpha)$ for each higher α seems to require already having the values for lower α. We seem to be involved in a vicious circle.

In fact, there is an easy way to find out what function $f_{K\in K}$ is. First, $f_{K\in K}(0)$ can't be 1; for the only function in W^{Π} that has value 1 at 0 is $\mathbf{1}$, and so $f_{K\in K}(1)$ would have to be 1; but $f_{K\in K}(1)$ can only be 1 if $f_{K\in K}(0)$ is 0. By a similar argument, $f_{K\in K}(0)$ can't be 0. It follows that $f_{K\in K}(0)$ must be $\frac{1}{2}$, and from that it is easy to successively obtain all the other values. (For the record, the value is $\frac{1}{2}$ at 0 and all limit ordinals, 0 at odd ordinals, and 1 at even successors.)

Other cases will not be so simple. For instance, consider a more general class of Curry-like properties, the properties of form $\lambda x(x \in x \rightarrow A(x; o_1, ..., o_n))$. Letting Q be the property for a specific choice of A and of $o_1, ..., o_n$, we want $|Q \in Q|$ to have the same value as $Q \in Q \rightarrow A(Q; o_1, ..., o_n))$. But A can be a formula of arbitrary complexity, itself containing $\in$ and $\rightarrow$, and the o_is can themselves be "odd" properties of various sorts. It isn't obvious how the reasoning that works for the simple Curry sentence will work more generally.

In many specific cases, actually, it is also easy to come up with a consistent value for the sentences involved; often a unique one, though in cases like the parameterized sentence '$\lambda x(x \in x) \in \lambda x(x \in x)$' it is far from unique unless further constraints are added. But it's one thing to figure out what the value would have to be in a lot of individual cases, another to come up with a general proof that values always can be consistently assigned. And it's still another thing to specify a method that determines a unique value for any formula relative to any assignment function. How are we to do these further things? The reasoning about the valuation of $K \in K$ suggests that for $\alpha > 0$ we might be able to figure out the function Z_α that assigns to each parameterized formula B the value $||B||(\alpha)$, if only we had the functions Z_β for $\beta < \alpha$; but getting the process going requires that we know Z_0, which includes the assignment to parameterized B in which includes arbitrarily high embeddings of the $\rightarrow$ in either the formula itself or the formulas involved in generating the parameters; this will depend on the Z_β for the very high values of β. The main problem, then, is to somehow break into the "transfinite circle".

I propose that we proceed by successive approximations. The main idea is to start *outside of* the space W^{Π}, so that we can treat the $\rightarrow$ in a way that mimics the behavior of the $\implies$ at ordinal values greater than 0 but abandons its rigid requirement about stage 0. We will start out by assigning the "0^{th} stage" of the evaluation of all conditionals artificially, and see what the later stages must be like as a result of this; it will turn out that by continuing far enough we will inevitably be led to an appropriate assignment Z_0 of values for the initial stage.

This must be combined with another idea, which is basically the one Kripke employed in his construction: we need to use a fixed point argument to construct the assignment to '$\in$' by approximations when the assignment to '$\rightarrow$' is given. And we need to somehow do these two approximation processes together; this is where most of the difficulties arise.

5.3 Constructing the Valuation of '$\in$': First Steps

OK, let's get down to business. The construction will assign values *in the set* $\{0, \frac{1}{2}, 1\}$ to formulas *relative to two ordinal parameters* α *and* σ (as well as to an assignment of values to the variables in the formula). α will initially be unrestricted; σ will be restricted to being no greater than Ω, the initial ordinal of the cardinality that immediately succeeds that of $|M^+|$. (Forget about Π for now; we will ultimately take it to be at least Ω, but it is not yet in the picture.) We order pairs $\langle\alpha, \sigma\rangle$ lexicographically, that is, $\langle\alpha, \sigma\rangle \preceq \langle\alpha', \sigma'\rangle$ iff either $\alpha < \alpha'$ or both $\alpha = \alpha'$ and $\sigma \leq \sigma'$; the reason for demanding that σ is restricted is so that this defines a genuine sequence. We will mostly be interested in the subsequence of pairs of form $\langle\alpha,\Omega\rangle$; values of σ smaller than Ω serve simply as auxiliaries toward producing the values at Ω. I will call the value of a sentence at the pair $\langle\alpha,\Omega\rangle$ its "value at stage α", and will often drop the Ω from the notation.

I now proceed to assign a value in the set $\{0, \frac{1}{2}, 1\}$ to each formula in L^+ relative to any choice of α, σ and s (the latter being a function assigning objects to the variables); except that as mentioned before I will drop the reference to s by understanding the formulas to be parameterized. I will use the single-bar notation $|A|_{\alpha,\sigma}$ instead of the double-bar notation $||A||$ used before, to emphasize that the value space is different. Eventually I will use the two-parameter sequence $|A|_{\alpha,\sigma}$ to recover $||A||$. (Just so you know where we're headed, the definition will be that $||A||$ is the function whose value at $\alpha < \Pi$ is $|A|_{\Delta+\alpha,\Omega}$; where Δ and Π are ordinals to be specified later. These ordinals will not depend on the particular A; and for all A, $|A|_{\Delta+\Pi,\Omega} = |A|_{\Delta,\Omega}$. Moreover, for all A, $|A|_{\Delta,\Omega}$ is 1 iff for all $\alpha > \Delta$, $|A|_{\alpha,\Omega}$ is 1; and analogously for 0, though not necessarily for $\frac{1}{2}$. These are the main conditions needed to ensure that $||A||$ meets the regularity conditions required for membership in the space W^Π, which in turn ensures that we get a reasonable logic.)

The single-bar assignment goes as follows:

1. $|o_1 = o_2|_{\alpha,\sigma}$ is 1 if $o_1 = o_2$; 0 otherwise.

2. If p is an atomic predicate of L other than '=', $|p(o_1, ..., o_n)|_{\alpha,\sigma}$ is 1 if $\langle o_1, ..., o_n\rangle$ is in the extension of p in M; 0 otherwise. (So it's 0 if any of the o_i are in E.)

3. $|Property(o)|_{\alpha,\sigma}$ is 1 if o is in E; 0 otherwise.

4. $|o_1 \in o_2|_{\alpha,\sigma}$ is 0 if o_2 is in the original domain $|M|$. Otherwise, o_2 is of form $\lambda x\Theta(x, b_1, ..., b_n)$ for some specific formula Θ and objects $b_1, ..., b_n$. In that case, $|o_1 \in o_2|_{\alpha,\sigma}$ is

1 if for some $\rho < \sigma$, $|\Theta(o, b_1...b_n)|_{\alpha,\rho} = 1$;
0 if for some $\rho < \sigma$, $|\Theta(o, b_1...b_n)|_{\alpha,\rho} = 0$;
$\frac{1}{2}$ otherwise.

5. $|\neg A|_{\alpha,\sigma}$ is $1 - |A|_{\alpha,\sigma}$
6. $|A \wedge B|_{\alpha,\sigma}$ is $\min\{|A|_{\alpha,\sigma}, |B|_{\alpha,\sigma}\}$
7. $|A \vee B|_{\alpha,\sigma}$ is $\max\{|A|_{\alpha,\sigma}, |B|_{\alpha,\sigma}\}$
8. $|\forall x A(x)|_{\alpha,\sigma}$ is $\min\{|A(o)|_{\alpha,\sigma} \mid o' \in |M^+|\}$
9. $|\exists x A(x)|_{\alpha,\sigma}$ is $\max\{|A(o)|_{\alpha,\sigma}| o' \in |M^+|\}$
10. $|A \rightarrow B|_{\alpha,\sigma}$ is

1 if for some $\beta < \alpha$, and any γ such that $\beta \leq \gamma < \alpha$, $|A|_{\alpha,\Omega} \leq |B|_{\alpha,\Omega}$;

0 if for some $\beta < \alpha$, and any γ such that $\beta \leq \gamma < \alpha$, $|A|_{\alpha,\Omega} > |B|_{\alpha,\Omega}$;

$\frac{1}{2}$ otherwise.

Note that when α is held fixed, the values of all atomic predications not involving '$\in$' (including those involving '$=$' and '*Property*'), and of all conditionals, is completely independent of σ: in the case of conditionals, that is because of the use of the specific ordinal Ω on the right hand side of 10. That means that for each fixed value of α we can perform the fixed point construction of [4]. (We perform it "transfinitely many times", once for each α.) More fully, for each α the construction is monotonic in σ: as σ increases with fixed α, the only possible switches in value are from $\frac{1}{2}$ to 0 and from $\frac{1}{2}$ to 1. So by the standard fixed-point argument, the construction must reach a fixed point at some ordinal of cardinality no greater than that of the domain; that is, at some ordinal less than Ω. And that means that we get the following consequence of 4:

(FP) For all α and all o and all Θ and all $b_1...b_n$,
$|o \in \lambda x\Theta(x, b_1...b_n)|_{\alpha,\Omega} = |\Theta(o, b_1...b_n)|_{\alpha,\Omega.}$

And by the rule for the biconditional that follows from 10 and 6 (together with the fact that an increase in σ stops having any effect by the time we've reached Ω) then implies that for any $\alpha \geq$ *1* (and any o, Θ and $b_1...b_n$),
$|o \in \lambda x\Theta(x, b_1...b_n) \leftrightarrow \Theta(o, b_1...b_n)|_{\alpha,\Omega} = 1$;
so (dropping Ω from the notation),

(FP-Cor1) For any Θ and $\alpha \geq 1$,
$|\forall u_1...\forall u_n \exists z \forall x[x \in z \leftrightarrow \Theta(x, u_1...u_n)]|_\alpha = 1.$

(FP-Cor1) looks superficially like the (III) that we require, but in fact it falls far short of it, for it says nothing about the double-bar semantic values that we need to guarantee a reasonable logic: nothing about regular functions from $Pred(\Pi)$ to $\{0, \frac{1}{2}, 1\}$. To do better, we need to explore what happens as we go to higher and higher values of α. That is the goal of the next subsection.

Before proceding to that, I note a substitutivity result:

> **(FP-Cor2)** If A is any parameterized formula, and A^* results from it by replacing an occurrence of $y \in \lambda x\Theta(x, o_1...o_n)$ by an occurrence of $\Theta(y, o_1...o_n)$, then for each α, $|A|_\alpha = |A^*|_\alpha$.

It's worth emphasizing that this holds even when the substitution is inside the scope of an $\rightarrow$. The proof (whose details I leave to the reader) is an induction on complexity, with a subinduction on α to handle the conditionals and the identity claims. (It is essential that the assignment of values to conditionals for $\alpha = 0$ didn't give conditionals different values when they differ by such a substitution; but it clearly didn't do that, since it gave all conditionals value $\frac{1}{2}$.)

5.4 The Fundamental Theorem

Is there a way to get from our single-bar semantic values relative to levels α to double-bar semantic values in a space W^Π? A naive thought might be to define $||o_1 \in o_2||$ as the function that maps each α into $|o_1 \in o_2|_\alpha$. But it should be obvious that this doesn't work: it doesn't meet the regularity condition that we need. (It does work in a few simple cases, like $f_{K \in K}$, but not in general.) The fact that all conditionals have value $\frac{1}{2}$ at $\alpha = 0$ is the most obvious indication of this.

But something like it will work: I will show that there are certain ordinals Δ, which I will call *acceptable ordinals*, with some nice properties. It turns out that if Δ is any acceptable ordinal and Π is any sufficiently larger acceptable ordinal that is also initial (so that it is equal to $\Delta + \Pi$), then we can use this Π for our value space W^Π, and we can define $||o_1 \in o_2||$ as the function that maps each $\alpha < \Pi$ into $|o_1 \in o_2|_{\Delta+\alpha}$. The conditions on acceptability will guarantee that the functions are regular. It will also turn out that even for complex formulas, $||A||$ is the function that maps each $\alpha < \Pi$ into $|A|_{\Delta+\alpha}$. And this will guarantee all of the laws that we need.[11]

The definition of acceptability that is easiest to use will require some preliminary explanation. To that end, I introduce a transfinite sequence of functions H_α. (These are the "single bar analogs of" the Z_α that I informally mentioned in Section 5.2.) H_α is defined as the function that assigns to each parameterized formula A the value $|A|_\alpha$ determined by the single-bar valuation rules. If $v = H_\alpha$, I say that α *represents* v. And if $H_\alpha = H_\beta$ I say that α is *equivalent to* β. I will make use of an intuitively obvious lemma that the reader can easily prove by induction on γ:

> **Lemma:** If α is equivalent to β then for any γ, $\alpha + \gamma$ is equivalent to $\beta + \gamma$.

Now let FINAL be the set of functions v that are represented arbitrarily late, i.e. such that $(\forall\alpha)(\exists\beta \geq \alpha)(v = H_\beta)$.

[11] In what follows, I use a slightly different definition of acceptability than in [2], though it is equivalent to the one there; the difference simplifies the proof somewhat.

Prop. 1: FINAL $\neq \emptyset$.

Proof: If it were empty, then for each function v from SENT to $\{0, \frac{1}{2}, 1\}$, there would be an α_v such that $(\forall\beta \geq \alpha_v)(v \neq H_\beta)$. Let θ be the supremum of all the α_v. Then for each function v from SENT to $\{0, \frac{1}{2}, 1\}$, $v \neq H_\theta$. Since H_θ itself is such a function, this is a contradiction. ■

Call an ordinal γ *ultimate* if it represents some v in FINAL; that is, if $(\forall\alpha)(\exists\beta \geq \alpha)(H_\gamma = H_\beta)$.

Prop. 2: If α is ultimate and $\alpha \leq \beta$ then β is ultimate.

Proof: If $\alpha \leq \beta$, then for some δ, $\beta = \alpha + \delta$. Suppose α is ultimate. Then for any μ, there is an $\eta_\mu \geq \mu$ which is equivalent to α. But then β, i.e. $\alpha + \delta$, is equivalent to $\eta_\mu + \delta$ by the Lemma, and $\eta_\mu + \delta \geq \mu$; so β is ultimate. ■

Call a parameterized formula A *ultimately good* if for every ultimate α, $|A|_\alpha = 1$; *ultimately bad* if for every ultimate α, $|A|_\alpha = 0$; and *ultimately indeterminate* if it is neither ultimately good nor ultimately bad. If Γ is a class of parameterized formulas, call an ordinal δ *correct for* Γ if

(ULT) For any $A \in \Gamma$, $|A|_\delta = 1$ iff A is ultimately good, and $|A|_\delta = 0$ iff A is ultimately bad.

(It follows that $|A|_\delta = \frac{1}{2}$ iff A is ultimately indeterminate. Also, if Γ is closed under negation then the clause for 0 follows from the clause for 1.) And call an ordinal *acceptable* if it is universally correct, that is, correct for the set of all parameterized formulas. (So if two ordinals are acceptable, they are equivalent, i.e. they assign the same values to every parameterized formula.)

Prop. 3: If δ is ultimate, then the following suffices for it to be correct for Γ:
for all $A \in \Gamma$, if A is ultimately indeterminate then $|A|_\delta = \frac{1}{2}$.

Proof: Since δ is ultimate, anything that is ultimately good or ultimately bad has the right value at δ, so only the ultimately indeterminate A have a chance of being treated incorrectly. ■

I now proceed to show that there are acceptable ordinals; indeed, arbitrarily large ones. Start with any ultimate ordinal τ, however large. Then every member of FINAL is represented by some ordinal $\geq \tau$; and since FINAL is a set rather than a proper class, and τ is ultimate, there must be a ρ such that $\tau + \rho$ is equivalent to τ and every member of FINAL is represented in the interval $[\tau, \tau + \rho)$. Finally, let Δ be $\tau + \rho \cdot \omega$. I will show that Δ is acceptable.

Prop. 4: For any n, every member of FINAL is represented in the interval $[\tau + \rho \cdot n, \tau + \rho \cdot (n + 1))$.

Proof: From the fact that $\tau + \rho$ is equivalent to τ, a trivial induction yields that for any finite n, $\tau + \rho \cdot n$ is equivalent to τ; so for any finite n and any $\alpha < \rho$, $\tau + \rho \cdot n + \alpha$ is equivalent to $\tau + \alpha$. So anything represented in the interval $[\tau, \tau + \rho)$ is represented in $[\tau + \rho \cdot n, \tau + \rho \cdot (n + 1))$. ■

Prop. 5: Δ is correct with respect to all conditionals.

Proof: Since Δ is ultimate, any ultimately good A has value 1 at Δ, and any ultimately bad A has value 0 at Δ. It remains to prove the converses.

Suppose $|B \rightarrow C|_\Delta = 1$. Then for some $\alpha < \tau + \rho \cdot \omega$, we have that $(\forall\beta \in [\alpha, \tau + \rho \cdot \omega))(|B|_\beta \leq |C|_\beta)$. Since $\alpha < \tau + \rho \cdot \omega$, there must be an n such that $\alpha < \tau + \rho \cdot n$. So $(\forall\beta \in [\tau + \rho \cdot n, \tau + \rho \cdot \omega))(|B|_\beta \leq |C|_\beta)$. But by Prop. 4, every member of FINAL is represented in $[\tau + \rho \cdot n, \tau + \rho \cdot \omega)$; so for every ultimate ordinal β, $|B|_\beta \leq |C|_\beta$. It follows by the valuation rules that for every ultimate β, $|B \rightarrow C|_\beta = 1$; that is, $B \rightarrow C$ is ultimately good. Similarly, if $|B \rightarrow C|_\Delta = 0$ then $B \rightarrow C$ is ultimately bad. ■

Fundamental Theorem: Δ is acceptable.

Proof: By Prop. 3, it suffices to show that if A is ultimately indeterminate then $|A|_\Delta = \frac{1}{2}$. Making the mini-stages explicit (and recalling that for any α, if a sentence has value $\frac{1}{2}$ at $\langle\alpha, \Omega\rangle$ then it has that value at all $\langle\alpha, \sigma\rangle$), the claim to be proved is that $(\forall A)(\forall\sigma)($if $||A|| = \frac{1}{2}$ then $|A|_{\Delta,\sigma} = \frac{1}{2})$. Or reversing the quantifiers, that $(\forall\sigma)(\forall A)($if $||A|| = \frac{1}{2}$ then $|A|_{\Delta,\sigma} = \frac{1}{2})$. Suppose this fails; let σ_0 be the smallest ordinal at which it fails. We get a contradiction by proving by induction on the complexity of A that

(*) $(\forall A)($ if A is ultimately indeterminate then $|A|_{\Delta,\sigma_0} = \frac{1}{2})$.

If A is atomic with predicate other than ’$\in$’, then A is not ultimately indeterminate, so the claim is vacuous. Similarly if A is $o_1 \in o_2$ where o_2 is not in E.

Suppose A is $o_1 \in o_2$ where $o_2 \in E$. Then o_2 is $\{x|\Theta(x, b_1, ..., b_n)\}$, for some $\Theta(x, b_1, ..., b_n)$. So if A is ultimately indeterminate, $\exists x[x = o \wedge \Theta(x, b_1...b_n)]$ must be too, since it has the same value as A at each stage. So by choice of σ_0, $|\exists x[x = o \wedge \Theta(x, b_1...b_n)]|_{\Delta,\sigma} = \frac{1}{2}$ for all $\sigma < \sigma_0$. But then by the valuation rules, $|o_1 \in o_2|_{\Delta,\sigma_0} = \frac{1}{2}$.

If A is a conditional, then by the valuation rules $|A|_{\Delta,\sigma_0}$ is $|A|_{\Delta,\Omega}$, i.e. $|A|_\Delta$, which (when A is ultimately indeterminate) is $\frac{1}{2}$ by Prop. 5.

The other cases use the claim that (*) holds for simpler sentences, and are fairly routine. E.g., if A is $\forall x A$, then if A is ultimately indeterminate, there is a t_0 such that $A(t_0/x)$ is ultimately indeterminate and for no t is $A(t/x)$ ultimately bad. But for any t for which $A(t/x)$ is ultimately indeterminate, including t_0, the induction hypothesis gives that $|A(t_0/x)|_{\Delta,\sigma_0} = \frac{1}{2}$; and for any t for which $A(t/x)$ is ultimately good, $|A(t/x)|_{\Delta,\Omega}$ is 1 and so $|A(t/x)|_{\Delta,\sigma_0} \in \{\frac{1}{2}, 1\}$. So by the valuation rules for $\forall$, $|\forall x A|_{\Delta,\sigma_0} = \frac{1}{2}$. ■

5.5 The Valuation of '$\in$' Concluded

We are now ready to choose the value of Π for our space W^{Π}, and to choose a W^{Π}-extension for '$\in$'.

Recall that the acceptable ordinal Δ just constructed was chosen to be bigger than an arbitrarily big τ; so the fundamental theorem gives that acceptable ordinals occur arbitrarily late. Let Δ_0 be the first acceptable ordinal and $\Delta_0 + \delta$ be the second; then an ordinal is acceptable iff it is of form $\Delta_0 + \delta \cdot \beta$.

If I hadn't already imposed stringent requirements on the space of semantic values (so as to be able to develop the semantics generally with as little bother as possible), I could now simply let $\in^*_{o_1 o_2}$ be the function that maps each $\alpha < \delta$ into $|o_1 \in o_2|_{\Delta_0+\alpha}$, and let the set of semantic values be the set of such functions for the different pairs $\langle o_1, o_2 \rangle$. But given that I have imposed the stringent requirements, this won't work: I need an acceptable $\Delta_0 + \Pi$ for which Π is an initial ordinal $\geq \Omega$. Also, if I don't insist that Π is strictly greater than δ I will need to prove that for each parameterized formula A there is a ρ_A smaller than δ such that the function $|A|_{\Delta_0+\alpha}$ is ρ_A-cyclic; I imagine that's so, but to avoid taking the trouble to prove it, I will construct Π to be strictly greater than δ, so that we can use δ as a common cycle for all the A.[12]

So let Π be any initial ordinal that is greater than $\Delta_0 + \delta$ and no less than Ω. Since Π is initial, and greater than $\Delta_0 + \delta$, it is identical to $\Delta_0 + \delta \cdot \Pi$, so it is acceptable. And (since $\Delta_0 + \Pi$ is also just Π), we can carry out the above idea using Π in place of δ:

> **(E)** For each σ_1 and σ_2, $\in^*_{o_1 o_2}$ is the function that assigns to each ordinal $\alpha < \Pi$ the value $|o_1 \in o_2|_{\Delta_0+\alpha}$.

Then every value $||o_1 \in o_2||$ is δ-cyclic.

The last thing that must be shown, to show that (E) does in fact succeed in assigning a W^{Π}-extension to '$\in$' for the Π recently chosen, is that each $||o_1 \in o_2||$ is regular. But that's clear: if it maps 0 into either 0 or 1, then $|o_1 \in o_2|_{\Delta_0}$ is 0 or 1, so by acceptability, $o_1 \in o_2$ is either ultimately bad or ultimately good, and so $|o_1 \in o_2|_{\Delta_0+\alpha}$ is either 0 for all α or 1 for all α; so $||o_1 \in o_2||$ is either $\mathbf{0}$ or $\mathbf{1}$.

So we have a W^{Π}-model. All that now remains of the consistency proof is to verify that the model validates Axiom Schema (III). This requires the following:

> **Theorem:** For each parameterized formula A, $||A||$ is the function that assigns to each ordinal $\alpha < \Pi$ the value $|A|_{\Delta_0+\alpha}$.

Proof: by induction on the complexity of A. It's true by stipulation for membership statements, and trivial for other atomic statements; and the clauses for

[12] Actually I could avoid a separate stipulation that $\Delta_0 + \delta < \Pi$ by proving this from the stipulation that Π is an initial ordinal, and that is an obvious consequence of what I assume to be a fact, that $\delta < \Delta_0$. But again, there's no need to take the trouble to prove this when an alternative stipulation of the value of Π will obviate the need.

quantifiers and for connectives other than → are completely transparent because the functions assigned these connectives in W^{Π}-models behave pointwise just like the corresponding connectives behave in the single-bar assignments. This is true for → too, except for the behavior at 0. So all we need to verify is the following:

> If $||A||$ and $||B||$ are the functions that assign to each ordinal $\alpha < \Pi$ the values $|A|_{\Delta_0+\alpha}$ and $|B|_{\Delta_0+\alpha}$ respectively, then $||A \to B||(0)$ assigns the value $|A \to B|_{\Delta_0}$.

But $||A \to B||(0)$ is by stipulation $(||A|| \implies ||B||)(0)$; that is,
1 if for some $\beta < \Pi$, and any γ such that $\beta \leq \gamma < \Pi$, $||A||(\gamma) \leq ||B||(\gamma)$;
0 if for some $\beta < \Pi$, and any γ such that $\beta \leq \gamma < \Pi$, $||A||(\gamma) > ||B||(\gamma)$;
$\frac{1}{2}$ otherwise.
But $||A||(\gamma)$ is by hypothesis $|A|_{\Delta_0+\gamma}$, and likewise for B, so these conditions are just the same as the corresponding conditions for $|A \to B|_{\Delta_0+\Pi}$. In other words, we've shown that $||A \to B||(0)$ is $|A \to B|_{\Delta_0+\Pi}$. And since acceptable ordinals are equivalent, that is just $|A \to B|_{\Delta_0}$, as required. ■

Corollary: Each instance of Axiom Schema (III) gets value **1**.

Proof: We need that for any $o, o_1, ..., o_n$, $||o \in \lambda x\Theta(x, o_1...o_n)|| = ||\Theta(o, o_1...o_n)||$. But by the Theorem, this reduces to the claim that for each α, $|o \in \lambda x\Theta(x, o_1...o_n)|_{\Delta_0+\alpha} = |\Theta(o, o_1...o_n)|_{\Delta_0+\alpha}$, and that is just a special case of the fixed point result (FP) proved in Section 5.3. ■

6 Satisfaction, Sets, and Proper Classes

Without too much trouble, the above construction could be generalized from properties to (non-extensional) n-place relations, for each natural number n. (Properties are the $n = 1$ case. We can include propositions as the $n = 0$ case.) There is a weak way to do this and a strong way. The weak way is to introduce, for each n, the unary predicate 'Rel^n' ('is an n-ary relation') and the $(n+1)$-place predicate '$\in^n$' (with $x_1, ..., x_n \in^n y$ meaning "y is an n-place relation and $\langle x_1, ..., x_n \rangle$ instantiates it"); also a single unary predicate 'REL' which each of the 'Rel^n' entail (we need this for restricting the variables to things that aren't relations). The strong way, which requires that the ground language L and ground theory T be adequate to arithmetic and the theory of finite sequences, is to introduce a single binary predicate '$Rel(n, z)$' meaning that z is an n-place relation ('REL' can obviously then be *defined*), and a single binary predicate '∈', with '$\in (s, z)$' meaning "for some n, z is an n-place relation and s is an n-place sequence that instantiates z". The details of both the weak and the strong generalization are, as far as I can see, routine. We can also easily build into the language an abstraction symbol that, when applied to any formula $\Theta(x_1, ..., x_n, u_1...u_k)$ of the language and any k-tuple of entities $o_1...o_k$, denotes the n-place relation $\lambda x_1, ..., x_n\Theta(x_1, ..., x_n, o_1...o_k)$; and we can introduce a predicate that applies only to such canonical relations. (In the model we used to prove consistency, all the relations were canonical, but this needn't be so in general.)

From such a generalized theory in the strong form, we could also obtain a consistent theory of expressions and of their satisfaction, a theory that validates the naive schema

$\langle x_1, ...x_n\rangle$ satisfies $\ulcorner\Theta(v_1, ..., v_n)\urcorner \leftrightarrow \Theta(x_1, ..., x_n)$.

The basic idea is obvious: identify the formulas of a language that contains a satisfaction predicate with canonical relations, and identify satisfaction with instantiation. Satisfaction claims thus would get values in the space W^{Π}, and excluded middle could not in general be assumed for them. It would be worth being more explicit about the details were it not for the fact that such a theory of satisfaction was given more directly in [2].

A more difficult question is whether we can generalize the above construction to a naive theory of *extensional* relations; or, to stick to the $n = 1$ case, a naive theory of *sets.* Here there do seem to be some difficulties. The matter is a bit complicated because there are several different ways one might propose to treat identity, and there are questions about whether one wants certain laws involving it to hold in full conditional form or only in the form of rules. But the main problem seems to be independent of these issues, for it doesn't involve identity: the issue is, how can we secure the rule

$Set(x) \wedge Set(y) \wedge \forall w(w \in x \leftrightarrow w \in y) \models \forall z(x \in z \leftrightarrow y \in z)$,

and preferably also the "reverse negated" rule

$\neg\forall z(x \in z \leftrightarrow y \in z) \models \neg\forall w(w \in x \leftrightarrow w \in y)$,

without any weakening of the Naive Comprehension Schema (III)? The natural way to try to secure these rules is to modify the treatment of '$\in$' so that what the fixed point construction ensures is not the (FP) of Section 5.3, but rather, (FP) only for the special case $\alpha = 0$, supplemented with

(FP-Mod) For all $\alpha > 0$ and all o and all Θ and all $b_1...b_n$, $|o \in \lambda x\Theta(x, b_1...b_n)|_\alpha = |\exists x[x \equiv o \wedge \Theta(x, b_1...b_n)]|_\alpha$,

where '$x \equiv y$' abbreviates '$[\neg Set(x) \wedge x = y] \vee [Set(x) \wedge Set(y) \wedge \forall z(z \in x \leftrightarrow z \in y)]$'. (Notice that $|x \equiv y|_\alpha$ depends only on the single-bar values of membership claims for $\beta < \alpha$, given the valuation rules for the biconditional; so there is no threat of circularity.) If we introduce the double-bar values on the basis of the single-bar ones as before, this would yield

$||o \in \lambda x\Theta(x, b_1...b_n)|| = ||\exists x[x \equiv o \wedge \Theta(x, b_1...b_n)]||$.

Since $||o \equiv o|| = \mathbf{1}$ for any o (given that $\equiv$ was defined via $\leftrightarrow$ rather than the material biconditional, and that the single-bar value at $\alpha = 0$ drops out by the time you get to the double-bar values), this would in turn yield

$||o \in \lambda x\Theta(x, b_1...b_n)|| \succeq ||\Theta(o, b_1...b_n)]||$,

which would ensure the validity of

(A) $\Theta(o, u_1...u_n) \rightarrow o \in \lambda x\Theta(x, u_1...u_n)$.

We also get a limited converse, viz. the rule

(B_1) $o \in \lambda x\Theta(x, u_1...u_n) \models \Theta(o, u_1...u_n)$.

But this is a significant lessening of naive comprehension: indeed not only do we not get the validity of the conditional

$o \in \lambda x\Theta(x, u_1...u_n) \rightarrow \Theta(o, u_1...u_n)$,

we don't even get the "reverse negation" of (B_1), viz.

(B_2) $\neg\Theta(o, u_1...u_n) \models o \notin \lambda x\Theta(x, u_1...u_n)$.[13]

This does not seem to me enough to count as Naive Set Theory. I don't rule out that we might be able to do better by a more clever construction, but it doesn't look easy.

But why do we need a naive theory of sets (or other extensional relations) anyway? We have a very nice non-naive theory of sets, namely the Zermelo-Fraenkel theory; and it can be extended to a naive theory of extensional relations either artificially, by defining extensional relations within it by the usual trick, or by a notationally messy but conceptually obvious generalization of ZF that treats multiplace extensional relations autonomously. (Formulations of ZF in terms of a relation of "having no greater rank than" greatly facilitate this generalization.)

It is true that the absence of proper classes in ZF is sometimes awkward. It is also true that adding proper classes in the usual ways (either predicative classes as in Gödel-Bernays, or impredicative ones as in Morse-Kelley) is conceptually unsettling: in each case (and especially in the more convenient Morse-Kelley case) they "look too much like just another level of sets", and the fact that there is no entity that captures the extension of predicates true of proper classes suggests the introduction of still further entities ("super-classes" that can have proper classes as members), and so on *ad infinitum.* But once we have properties (and non-extensional relations more generally), this difficulty is overcome: properties can serve the function that proper classes have traditionally served. The rules they obey are so different from the rules for iterative sets (for instance, they can apply to themselves) that there is no danger of their appearing as "just another level of sets". And since every predicate of properties itself has a corresponding property, there is no fear that the motivation for the introduction of properties will also motivate the introduction of further entities ("super-properties").[14]

Of course, in standard proper class theories, proper classes are extensional; whereas properties are not. Does this show that the properties won't serve the purposes that proper classes have been used for? No. I doubt that extensionality among proper classes plays much role anyway, but without getting

[13] For a counterexample, let o_1 be $\{w|w \equiv w\}$, o_2 be $\{w|w \equiv w \wedge K \in K\}$ (where K is the Curry set), and o_3 be $\{w|\neg(w \equiv w)\}$; and let $\Theta(x, o_3)$ be '$x \equiv o_3$'. $||\neg\Theta(o_1, o_3)|| = \mathbf{1}$; but $||o_1 \notin \lambda x\Theta(x, o_3)||$ is $\mathbf{1} - \curlyvee\{||o \equiv o_1 \wedge \Theta(o, o_3)||\}$, which is $\preceq \mathbf{1} - ||o_2 \equiv o_1 \wedge \Theta(o_2, o_3)||$, i.e. $\preceq \mathbf{1} - ||o_2 \equiv o_1 \wedge o_2 \equiv o_3)||$. But $o_2 \equiv o_1$ has the value $K \in K \rightarrow \top$ and $o_2 \equiv o_3$ has the value $K \in K \rightarrow \bot$; and both assume value $\frac{1}{2}$ at limit ordinals, so $||o_2 \equiv o_1 \wedge o_2 \equiv o_3)||$ is not $\mathbf{0}$, so $||o_1 \notin \lambda x\Theta(x, o_3)||$ is not $\mathbf{1}$.

[14] The general philosophical view here is quite similar to that in [5], though the theory of properties on offer here is much stronger because of the presence of a serious conditional.

into that, one could always use the surrogate $\equiv$ as a "pseudo-identity" among properties that is bound to be adequate in all traditional applications of proper classes; and an extensionality law stated in terms of $\equiv$ rather than $=$ is trivially true. Of course, $\equiv$ is very bad at imitating identity among properties generally: if it weren't, the problem of getting an extensional analog of naive property theory would be easy. But when we confine our attention to those properties that correspond to the proper classes of Gödel-Bernays or Morse-Kelley–in both cases, properties that hold only of things that aren't themselves properties but rather are sets– then $\equiv$ is a very good surrogate for identity: for instance, *over this restricted domain*, excluded middle and all the usual substitutivity principles hold of $\equiv$. Consequently, we have a guarantee that properties will serve all the traditional purposes of proper classes (even in the impredicative Morse-Kelley theory).

My claim, then, is (i) that if we have a naive theory of properties in the background, we have all the advantages of proper classes without the need of any "set-like entities" beyond ordinary sets; (ii) that given this naive theory of properties, ordinary iterative set theory (ZF) is a highly satisfactory theory; and (iii) there is no obvious need for any *additional* theory of "naive sets".

But if there is no need of a naive theory of sets, why is there a need for a naive theory of properties, and for a naive theory of satisfaction? Was this paper a wasted effort?

In fact, the case of properties (on at least one conception of them) and of satisfaction are totally different from the case of sets. For the way we solve the paradoxes of naive set theory in ZF is to deny the existence of the alleged set: for instance, there simply is no set of all sets that don't have themselves as members. The analogous paradox in the case of the theory of satisfaction involves the expression 'is not true of itself', and if we were to try to solve the paradox on strictly analogous lines we would have to deny the existence of the expression! That would be absurd: after all, I just exhibited the expression. We could of course say "Sure, there's an expression 'is not true of itself', but it doesn't have the features one would naively think it has, such as being true of just those things that are true of themselves". This would be admitting that the expression exists, but denying the naive satisfaction theory. That is certainly a possible way to go, but it isn't at all like the solution in the ZF case. There are reasons why I don't think it is a *good* way to go: see [3]. But without getting into that here, let me simply say that the cost of violating such schemas is high, and is quite unlike anything that is done in ZF (where we deny the existence of the set, instead of saying that it exists but has different members than you might have thought).

The case of properties is slightly more complicated, because there is I believe more than one notion of property. There is, first, the notion of *natural property*, as discussed for instance in [6]. Here we do not want anything like Naive Comprehension: it is central to the idea of natural properties that it is up to science to tell us which natural properties there are. (It is also doubtful that we want natural properties of natural properties. Even if we do, it seems likely that we should adopt a picture which is "ZF-like" in that each natural property has a rank and applies only to non-properties and to properties of lower rank. But there is no need to decide these issues here.) But in addition to the notion of natural property, there is also a conception of property that is useful in semantics. And it is the *raison d'etre* of such "semantically conceived properties" (*sc-properties* for short) that every meaningful open sentence (in a given context) corresponds to one.[15] (Open sentences in the language of sc-properties are themselves meaningful, so they must correspond to sc-properties too.) Again, a ZF-like solution in which the existence of the properties is denied goes against the whole point of the notion.

In a theory of semantically conceived properties, then, it is unsatisfactory to say that for a meaningful formula $\Theta(x)$, there is no such thing as $\lambda x\Theta(x)$. It also seems unsatisfactory to say that though $\lambda x\Theta(x)$ exists, the things that instantiate it are not the o for which $\Theta(o)$. In classical logic, those are the only two options, but what I've shown in this paper is how to develop a third option in which we weaken classical logic. If we do that, then we can retain the naive theory of (sc-)properties, and that has an important payoff that has no analogue in the case of sets. At the very least, the value of a naive set theory is unobvious; but the value of a naive theory of satisfaction is overwhelmingly clear, and it is almost as clear that we ought to want a naive theory of sc-properties if we are going to posit sc-properties at all.

You may still want a naive theory of sets, for whatever reason; but what you need is a naive theory of properties and a naive theory of satisfaction. I suspect that you can't get what you want; but you get what you need.

References

[1] Hartry Field. Is the liar sentence both true and false? In JC Beall and Brad Armour-Garb, editors, *Deflationism and Paradox*. Oxford University Press, 2003.

[2] Hartry Field. A revenge-immune solution to the semantic paradoxes. *Journal of Philosophical Logic*, 32, 2003.

[3] Hartry Field. The semantic paradoxes and the paradoxes of vagueness. In JC Beall, editor, *Liars and Heaps*. Oxford University Press, 2003.

[4] Saul Kripke. Outline of a theory of truth. *Journal of Philosophy*, 72:690–716, 1975.

[5] Penelope Maddy. Proper classes. *Journal of Symbolic Logic*, 48:113–39, 1983.

[15] Or rather, every meaningful open sentence with a distinguished free variable corresponds to a sc-property relative to any assignment of entities, including possibly sc-properties, to the other free variables.

[6] Hilary Putnam. On properties. In Hilary Putnam, editor, *Mathematics, Matter and Method:Philosophical Papers, vol. 1*. Cambridge University, 1975.