Forthcoming in *The Routledge Handbook of Propositions*, Chris Tillman (ed.)

The Propositional Benacerraf Problem

Jesse J. Fitts
University of Nevada Las Vegas
jesse.fitts@unlv.edu

Abstract

Writers in the propositions literature consider the Benacerraf objection serious, often decisive. The objection figures heavily in dismissing standard theories of propositions of the past, notably set-theoretic theories. I argue that the situation is more complicated. After explicating the propositional Benacerraf problem, I focus on a classic set-theoretic theory of propositions, the possible worlds theory, and argue that methodological considerations influence the objection's success.

Contents

1 Introduction **1**

2 The Benacerraf Problem **3**
2.1 Benacerraf 1965 . . . 3
2.2 The Propositional Benacerraf Problem: The Classic Version . . . 4
2.3 The Propositional Benacerraf Objection: Further Thoughts and Another Version . . . 4
2.4 Example: King's account . . . 6
2.5 M1 gives rise to a novel source of indeterminacy . . . 9

3 A defense of a worlds theory **13**

1. Introduction

How can any given propositional candidate *be* the proposition that p when there are equally good candidates? Call this the *propositional Benacerraf problem*.[1] This problem originates in the philosophy of mathematics, targeting set-theoretic reductions of numbers, but generalizes to other entities, including some views of propositions. This problem has played a major role in shaping the current landscape of the propositions literature, mostly in setting aside traditional set-theoretic theories of propositions to make way for contemporary theories. For example, Jeffrey King's theory was partially constructed to be Benacerraf immune (as he

[1] Benacerraf (1965)

was reacting against theories he thought succumbed to the problem).[2] Scott Soames leveled a Benacerraf objection against his own "deflationary account" of propositions, which lead to his current "cognitive realist theory."[3] Others have drawn dire conclusions in the face of arbitrariness worries like the Benacerraf problem. Moore (1999) and Bealer (1998) argue from the Benacerraf objection to the conclusion that propositions are *sui generis* entities; Armour-Garb & Woodbridge (2012) employ the problem to argue for a fictionalist account of propositions; and Jubien (2001) presents a dilemma, a horn of which is the Benacerraf objection, to argue that propositions don't exist. The reaction has been that the Benacerraf problem is not merely a demerit for a theory but a theory ender, and the possibility of a general Benacerraf problem has lead some to reject propositions all together.

Given the problem's importance, we should properly understand it and evaluate whether its targets in fact fall to it. This understanding and evaluation correspond to the two goals of this paper, presented in §2 and §3, respectively. For the first goal, I argue that the introduction of an empirical aspect to propositional theorizing complicates the problem. The propositions literature has recently taken an empirical turn with the work of King and Soames. Below, I call this empirical turn the *mind-first* movement. According to theories falling within this movement, the representational capacity of propositions derives from the representational capacity of the mind. I contrast this with traditional *proposition-first* theories according to which propositional representation is fundamental. According to mind-first theories, the nature of propositions is in some way tied to the actual workings of the minds of representing beings. This empirical aspect of the theory introduces complications for the propositional Benacerraf problem that aren't present in the original mathematical version.

This empirical turn is partly the result of mind-first authors—both King and Soames, among others—finding some traditional theories of propositions lacking in part because they fall to the propositional Benacerraf problem. In particular, both authors, and others, have lodged the propositional Benacerraf problem against traditional set-theoretic theories of propositions, which tend to come in two varieties: tuple theories and worlds theories. In fact, set theoretic theories are supposed to be the paradigmatic example of theories that fall to the objection. For the second goal, I consider a case study of the problem with a version of the possible worlds theory of propositions and argue that this theory, on one plausible way of thinking of it, can be defended against the problem. In particular, I argue that the problem has incorrectly targeted

[2]See King (2007) for a book-length treatment. See King's papers in King et al. (2014) for further developments in his theory.

[3]See Soames (2010, ch.5) for the objection. See Soames (2015) for a more recent book-length treatment as well as his papers in King et al. (2014).

models of propositions rather than the target of those models, where the problem properly lives. In order to be concrete, I will defend one understanding of one set-theoretic theory, and this being the case, the second goal of the paper will have a limited scope.

2. The Benacerraf Problem

2.1. Benacerraf 1965

Benacerraf raised a problem for the reduction of numbers to sets. For such a purported reduction, various progressions of sets work equally well. Two conspicuous contenders are von Neumann's ordinals and Zermelo's, each progression which goes, respectively, as follows:

(1) $\varnothing, \{\varnothing\}, \{\varnothing, \{\varnothing\}\}, ...$

(2) $\varnothing, \{\varnothing\}, \{\{\varnothing\}\}, ...$

These two progressions may serve equally well as *representations* of the natural numbers, but the question is, *Are they the natural numbers*? Benacerraf says no. Benacerraf presents the objection in parable form in which different logician parents bring up two children, Johnny (for John von Neumann) and Ernie (for Ernst Zermelo), teaching them the opposite progressions—Johnny learning Zermelo's progression; Ernie learning von Neumann's—instead of usual counting and mathematical instruction. Upon the set-theoretic instruction's completion, the children quarrel: Ernie insists that 3 is a member of 17, and Johnny disagrees. This disagreement brings out the first step of the Benacerraf objection: the candidate reduction bases are *numerically distinct*. The second step is the arbitrariness step. The naturals can't be both progressions since they're non-equivalent by step one, so the naturals must be one or the other (or some yet other progression). If the naturals are, say, the von Neumann progression rather than Zermelo's, then there must be an argument for that choice. Yet, no argument is forthcoming—what could the reasoning possibly be since (the argument goes) both work equally well foundations-wise? The final step is to conclude that the purported reduction/identification fails—in this case, that numbers aren't sets.

Here is a first attempt at generalizing the problem:

Benacerraf problem—provisional statement If there are multiple, non-equivalent classes of entities that contend for an identification (or reduction) base for some class of entities E, and there's no non-arbitrary reason to choose one reduction/identification base over another, then E in fact isn't (or can't be reduced to) any of those bases.

2.2. The Propositional Benacerraf Problem: The Classic Version

The objection applies straightforwardly to worlds theories. Step one: Worlds theorists employ both sets of worlds and functions from worlds to truth values. We may have the set of worlds at which snow is white—

$$\{w_1, w_2, w_3, ...\}$$

—and the function that pairs those worlds with 1 (or true, or whatever) and the others with 0—

$$\{\langle w_1, 1\rangle, \langle w_2, 1\rangle, \langle w_3, 1\rangle, ..., \langle w_n, 0\rangle, ...\}.$$

And it's clear that, ontologically,

$$\{w_1, w_2, w_3, ...\} \neq \{\langle w_1, 1\rangle, \langle w_2, 1\rangle, \langle w_3, 1\rangle, ..., \langle w_n, 0\rangle, ...\}.$$

Thus, the two are numerically distinct. Step two: Both sets and their characteristic functions work equally well by the world theorist's lights. Conclusion: Propositions are neither sets of worlds nor their characteristic functions. Call this the *classic propositional Benacerraf problem*, or just the *classic problem* since we'll focus mostly on propositions, and I'll flag the mathematical case below.

2.3. The Propositional Benacerraf Objection: Further Thoughts and Another Version

The key feature of a Benacerraf problem of any type is untoward arbitrariness. About that arbitrariness, I raise two points that I discuss at once. First, the *nature* of the arbitrariness is important, for it gives rise to different kinds of Benacerraf problems the strengths of which may depend on the source of that arbitrariness. Second, different *kinds* of potential constraints may eliminate the arbitrariness.

The arbitrariness in the original mathematical Benacerraf problem involved, as Benacerraf himself presented it, the epistemic idea of *cogent reasons*: If one doesn't have a *cogent reason* to favor that $a = b$ rather than that $a = c$ (where $b \neq c$), or vice versa, then it's arbitrary whether $a = b$ or $a = c$. And if it's arbitrary whether $a = b$ or $a = c$, then $a \neq b \neq c$. But it can't be merely that if we lack cogent reasons, at present, whether $a = b$ or $a = c$, then a Benacerraf problem arises. If either the butler or the gardener solely committed the murder, and we don't have a cogent reason to identify one or the other as the murderer, that may give us reason not to accuse or indict either—that would be arbitrary. But it's not arbitrary, holding fixed the disjunction, which *is* the murderer despite our ignorance. It mustn't be that we merely lack a cogent reason, but that, for some reason, no cogent reason is forthcoming.

Why is no cogent reason forthcoming? In the original mathematical Benacerraf problem and in the classic propositional Benacerraf problem, it seems that not only is no cogent reason forthcoming, but no reason could *ever possibly* be forthcoming. In no possible world could we obtain cogent reasons that would favor choosing one set progression over another in the mathematical problem. In the classic propositional case, if we grant that there are roles for coarse grained propositions, then, as in the mathematical case, there is no possible world in which we could have cogent reasons to favor the set version versus the function version. In both the mathematical and classic propositional Benacerraf problems, we're choosing among candidates according to role suitability—whether, respectively, some entities behave like natural numbers (in that we can define arithmetical operations on them, e.g.) or whether some entities behave like coarse-grained content (in that factually equivalent propositions are identical, logical relations are definable, etc.). And given that the candidates in both reduction/identification projects are formally equivalent, there can be no *possible* cogent reasons to favor one candidate over another.

We solve a Benacerraf problem when we resolve this arbitrariness, and we resolve the arbitrariness on the basis of some kind of criteria or other. Yet there are different kinds of criteria, and recent developments in the propositions literature give rise to a new kind of criterion that can affect the modal profile of the just-mentioned cogent reasons. We're familiar with the two typical kinds of criteria on which we judge candidates in an identification/reduction. The first I'll just call *role-filling* criteria. These criteria are simply whether the candidate class of entities really satisfies the recognized *desiderata* for a theory. In the original Benacerraf problem, if we couldn't define the successor relation on either the von Neumann or Zermelo progressions, then the arbitrariness would obviously resolve. For propositions, any entities that we identify with propositions have to play whatever propositional roles we intend those propositions to play. It may be that one class of entities can play all of the propositional roles, or perhaps only some roles if "[t]he conception we associate with the word 'proposition,"' as Lewis (1986, p.54) says, is "a jumble of conflicting *desiderata*."

Whatever the case may be, if we set out to give a theory of, e.g., the objects that enter into logical relations, then we need to be able to define logical relations on those objects. If one class of entities satisfies this criteria and another doesn't, then that favors the role-filling class of entities. I may be stating the obvious—that *desiderata*-satisfaction favors one class of entities over another—but this arbitrariness resolver contrasts with another: *instrumental* criteria, such as simplicity, economy, etc. While such criteria can perhaps be called upon to break ties—simplicity in particular has in the propositions literature—such criteria are applied *after* role-filling criteria. If we're presenting a theory of the objects of logical relations, and we can define such relations on one class of entities and not another, then it doesn't matter how simple, e.g., the latter entities are.

The propositions literature has recently taken an empirical turn. Traditionally, theorists saw propositions as the ultimate source of representation from which mental representation derived. This contrasts with views according to which mental representation is fundamental. These traditions are important enough to set off and name:

Proposition-First (P1) The thesis that propositions exist, and have their nature, independently of minds, and propositions are the ultimate source of representation while mental representation is derivative.

Mind-First (M1) The thesis that propositions depend for their existence and nature on the mind, and minds are the ultimate source of representation while propositional representation is derivative.

On an M1 theory, the *basis of content*—the workings of the mind that give rise to content—constrain the M1 theorist's proposed propositions. What this means is that in addition to the propositions filling their various roles, propositions must somehow be related to the actual, empirical workings of the mind. Thus, if we have an apparent tie between two candidates both of which satisfy their proposed roles equally well—and both are equally simple, yet one *actually* derives from the basis of content while the other is somehow in error in its understanding of the empirical facts that give rise to content—then the tie is broken in favor of the empirically informed entities. For a quick example to which we'll return, King employs syntactic relations in his propositions. Now suppose we focus on two propositional candidates for the proposition that p, one which includes binary branching syntax only while the other includes some ternary branching in its syntax. The candidates may serve equally well as the objects of the attitudes, or as the content of assertive sentences, or whatever. Further suppose that linguists inform us that the binary branching hypothesis is true. Then, since one candidate gets the empirical facts about the basis of content wrong, we break the tie.

2.4. Example: King's account

Let's apply this material to a contemporary example: Caplan & Tillman's (2013) Benacerraf objection against King's theory. I will only review as much of King's theory as I need and assume some familiarity with the theory.

King's propositions are *facts*—an object instantiating a property, or n objects standing in an n-place relation. Consider the simple sentence "Rachel reads." In that sentence, two words stand in a sentential relation, call it R, familiar from syntactic tree representations:

Rachel reads

R is syntactic concatenation, which the branching structure represents. The lexical items "Rachel" and "reads" have as their semantic values the person Rachel (assuming direct reference with King) and the property of being a reader, respectively. We'll use an asterisks to denote the semantic value of a lexical item—so, e.g., "Rachel*" is the semantic value of the lexical item "Rachel":

Rachel reads

| |

Rachel* reads*

Propositions are language-independent entities, so King existentially quantifies over the language (in this case English) and the lexical items (in this case "Rachel" and "reads"), represented by stars:

★ ★

| |

Rachel* reads*

This is a partial visual representation of King's initial account[4] of the proposition that Rachel reads. He later added material about contextually sensitive lexical items and assignment functions for variables as well as the semantic significance of the sentential relation R. Agents interpret syntactic concatenation in a handful of ways, those ways described by formal semanticists, e.g., Heim & Kratzer (1998). Without this information in the account, it is possible that syntactic concatenation might encode the anti-instantiation function, so that the proposition that Rachel reads is true just in case Rachel does not possess the property of being a reader. In English, however, agents interpret R to ascribe the semantic value of "reads"—the property of being a reader—to the semantic value of "Rachel"—the person Rachel. The proposition that Rachel reads, on King's account, is Rachel, the person, standing in the following relation to the property of being a reader, i.e., $(Pr_2)(Rachel, reads)$:

(Pr_2) There is a context c and assignment f such that ____ is the semantic value relative to c and f of a lexical item e of some language L and ____ is the semantic value relative to c and f of a lexical item e'

[4] King (1995).

of L such that e occurs at the left terminal node of the sentential relation R that in L encodes ascription and e' occurs at R's right terminal node.

At different times, as King (2016) himself notes, he has written as though interpreting R and the propositional relation itself, that is Pr_2, are different things and that Pr_2 itself encodes instantiation. For example, in King (2009, p.264), he says that "there is no reason to think" $(Pr_2)(Rachel, reads)$ " has truth conditions and so is either true or false." In order for that to be the case, King, in the past, claimed that the *propositional relation itself*, i.e., (Pr_2), encodes instantiation. We thus have two separate facts-qua-propositions as the candidates for the proposition that Rachel reads:

(3) $(Pr_2)(Rachel, reads)$

(4) $(Pr_2)(Rachel, reads)$ and (Pr_2) encodes instantiation.

Caplan and Tillman claim that we can't break the tie between (3) and (4). They claim that both have truth conditions, but that (4) has them intrinsically while (3) has them extrinsically. Yet, they claim, (3) is simpler while (4) is more complex. *If* Caplan and Tillman were right in (4) alone securing intrinsic truth conditions, and *if* (4) were King's official account of the proposition that Rachel reads, we would not have a genuine tie. Intrinsic truth conditions is a *role-filling* criteria while simplicity is an *instrumental* criteria and the former take priority over the latter.[5] However, as of King (2016), King's official account of the propositional relation is Pr_2 and that propositions involving Pr_2 have truth conditions essentially and intrinsically. As he says:

> [I]nterpreting the propositional relation in a certain way just is interpreting the relevant sentential relation in that way [i.e. composing semantic values at syntactic concatenation points]. Once you are clear about the fact that interpreting the sentential relation is just a matter of composing semantic values (relative to contexts) in a certain way and that semantic values only get composed in the definition of a proposition being true at a world on my view, you see that interpreting the propositional relation in the relevant way in this definition just is composing semantic values in the way dictated by the interpretation of the relevant sentential relation. . . . [I]t would be metaphysically impossible [for the sentence "Rachel reads"] to exist with its sentential relation interpreted as encoding ascription and the propositional relation . . . not to be so interpreted.

On the next page, however, King is somewhat ambivalent about whether it's possible for R to encode

[5] Fletcher (2013, p.8) makes this point originally (without my language of criteria and their order of importance).

ascription without the propositional relation encoding ascription. In the end, King chooses instrumental criteria to resolve the indeterminacy in favor of (3) because it's more "minimal" than the "unnecessarily large" (4).

Both (3) and (4) fill the propositional roles equally well, but empirical criteria may break the possible tie here. Recall that as an M1 theory, King's propositions are grounded in the actual psychological processes of humans. This leaves open two responses to this alleged Benacerraf problem. If propositions are based on the actual (and possible) relations that humans bear to King's facts-qua-propositions, then, either there will be a fact of the matter regarding the relations agents bear toward the proposition that Rachel reads—either agents relate to (3) or (4)—or there won't be a fact of the matter. We take the former attitude if we are thoroughgoing realists about the objects that computations are defined over in syntax and formal semantics. If agents really are composing the semantic values of some fact when they compute the meaning of a sentence, they will be computing some fact or another, and these empirical facts will resolve the indeterminacy without recourse to instrumental criteria. On the other hand, if one is inclined to take formal semanticists and syntacticians to be providing something akin to psychological *models*, in the sense familiar from the philosophy of science literature on scientific modeling, then it may be possible that the psychological facts underdetermine which proposition between (3) and (4) is *the* proposition that Rachel reads. In the next sub-section, I suggest this indeterminacy is benign and doesn't give rise to a Benacerraf problem.

2.5. M1 gives rise to a novel source of indeterminacy

It may seem like this latter option just is a Benacerraf problem for King—the possibility of the empirical facts underdetermining our choice of (3) versus (4). The empirical turn in the propositions gives rise to this kind of indeterminacy, previously foreign in the propositions literature, though familiar elsewhere, that can crop up amidst any empirical investigation, and I don't think of this indeterminacy as giving rise to a Benacerraf problem. We're familiar with indeterminacy arising from vagueness, future contingents, translation, reference, etc. The list goes on and isn't restricted to such philosophical mainstays: For example, given the historical city-planning facts, it's indeterminate whether Princeton is the borough of Princeton or the borough in addition to the surrounding township.[6] Why is this relevant to us? Because that's what the Benacerraf problem involves in other words—irresolvable indeterminacy. But the kind of indeterminacy that often arises in metaphysical investigations is different from the kind that arises in empirical investigations,

[6] This example is from Lewis (1988, p.129).

especially investigations that employ modeling methodology.[7] And before moving on, I want to be clear that the indeterminacy that I'm interested in here is *not* epistemic but metaphysical.

I have in mind scenarios involving a non-theoretical level of description underdetermining facts about a theoretical level of description whose purpose is to account for that non-theoretical level. Here are a few examples:

- Consider formal epistemological facts about someone's subjective confidence in some proposition—say whether it will rain tomorrow. There are, on the one hand, the facts about the person's brain and her behavior (say betting behavior), and on the other, her subjective confidence in rain, say it's .5. There may be no fact of the matter whether her credence in rain is .5 or .50000...1 (add in as many zeros as you like). This is because the one level of description—the facts about her brain, her behavior—whatever goes into determining whether and to what degree someone believes something—underdetermines what her credence is.

- Consider the Lewisian view about the relationship between languages and language in Lewis (1983b). On the one hand, we have language as a convention-governed social phenomenon—conventions that Lewis accounts for game theoretically; and on the other, we have language as a formal semantic function from strings to meanings, where meanings determine extensions in possible worlds. Given that there will be some indeterminacy in the regularities that govern linguistic conventions, which in turn determines which language some population speaks, there will be indeterminacy in the relation between a given population's language as a social phenomenon and the formal semantic account of that language.

- Suppose at the end of syntactic inquiry, there is no evidence to distinguish between whether every branching node is binary or whether some branches are *n*-ary. Further suppose that syntacticians model actual human psychological processes, as many take themselves to be doing. Of course one reason that such a situation may arise is that our science will just never be good enough. But it may be that human physical and behavioral facts underdetermine fine-grained theoretical distinctions at the level of the syntactic model.

The *level* at which this indeterminacy resides is, respectively, a given agent's subjective confidence, the correct semantics for a natural language, or human syntactic knowledge. I'm *not* interested in the models of

[7] "Modeling" here means the methodology discussed in philosophy of science; for example, see Weisberg (2013) for a contemporary book-length discussion.

these—respectively, candidate credence functions, set-theoretic formal semantics, and tree-theoretic syntax. As we'll see, Benacerraf objections don't apply at the level of a theoretical model.

The above scenario types look similar to the classic mathematical and propositional versions: We have multiple theoretical candidates to choose from among which no candidate is better than others in terms of cogent reasons. For the classic propositional Benacerraf problem, only role-filling and instrumental criteria could break the tie. Like I said, the lack of cogent reasons is necessary and the indeterminacy is necessarily irresolvable. Yet there is a feeling of contingency once we introduce empirical investigation into theories of propositions. It may very well turn out that certain facts will remain unresolved not merely as a matter of our clumsy science but even at the end of science. In these cases, it seems to me that it's *possible* that some evidence could distinguish the candidates. It's just that (or at least we've stipulated) that there is no *actual* evidence to distinguish the cases. If we can distinguish between .9 and .1 credences with some confidence, then it seems possible that for some very fine-grained distinction, there could be some behavioral or brain fact that distinguishes them. Similar considerations go for the other two examples: there will always be indeterminacy in the relation between Lewis' two notions of language, but for two given formal semantics, we can produce a world where *that* indeterminacy resolves. And if we have constituency tests that give us cogent reasons to identify obvious constituents, we can imagine speech and brain facts that would resolve the indeterminacy between binary and *n*-ary branching. That we're left with these indeterminacies and not some others at the end of science just seems to be a contingent fact about our world.

I want to be clear with what I mean about "contingent" indeterminacy. When I say that familiar cases of indeterminacy, for example vagueness, are necessarily irresolvable, what I mean is this: Once we fix upon *our* concept of, e.g., baldness, and we fix upon some borderline bald person, it doesn't matter how other facts vary from world to world—that person will be a borderline case in every world. Once we fix our concept of what a proposition is and fix that a given proposition *p*'s identity is indeterminate between a set of worlds and a characteristic function thereof, we can vary other facts and the indeterminacy will remain irresolvable. I'm contrasting this case with a different kind of indeterminacy, and this different kind of indeterminacy is relevant in our current propositions literature. This difference, in the propositional case, arises from theorists introducing empirical considerations into their theories. For example, King employs syntax in his facts that are his propositions. Syntax is, of course, an ongoing field of investigation, and King employs syntax in his propositions *however, with some constraint,*[8] *syntax turns out in the end*. Now, in our world, it may turn out

[8] For example, syntax not providing instructions to the semantics to perform the relevant compositions may doom King's program.

that, with respect to some syntactic facts, those facts don't turn out one way or another. There may be some irresolvable *actual* indeterminacy between the facts about our brains and speech behavior and syntax. But how things turn out *at some other world* may be different—such indeterminacies may resolve. And this is what I mean by the indeterminacy being contingent.

This matters for us because which candidate proposition is *the* proposition that p may depend on the resolution of some actual indeterminacy that may, in our world, be irresolvable. The binary/n-ary branching example is just one example. Thus we may end up with multiple propositional candidates equivalent in role and instrumental suitability, and because of empirical indeterminacy, they may be equally empirically well-suited. It may seem, then, in such cases that we have a Benacerraf problem on our hands, but I think this is the wrong conclusion to draw. In this set up, where empirical investigation stalls out, I think the indeterminacy is benign and should be chalked up to the complexities of the empirical world. If a Benacerraf problem were to threaten *here*, at the propositional level, then we would have an explosion of Benacerraf problems wherever there is indeterminacy at the interface between philosophy and science. To put this point another way, as Clarke-Doane (2013) notes, casual chains help to resolve indeterminacy in the empirical realm. But there may be situations, such as those just outlined, where such chains do not, in our world, resolve the indeterminacy. This differs, Clarke-Doane notes, from a purely metaphysical realm, in his case the realm of numbers. In such realms, causal chains won't be relevant. Here, it is our concepts that will resolve indeterminacies if they are resolvable.

In light of the new empirical aspect of the mind-first movement, where do we stand with the propositional Benacerraf problem? We need to amend our provisional statement so that indeterminacies of the kind just mentioned don't lead to a Benacerraf problem. It's important to note here that my goal isn't a general discussion of the Benacerraf problem in all of its contexts. Rather, the purpose of this section is to synthesize the developments in the propositions literature with the propositional Benacerraf problem. We'll add to the provisional statement of the propositional Benacerraf problem these alterations: If the reason for lack of non-arbitrary reasons is an actual, metaphysical indeterminacy resulting from the empirical aspect of the theory in question—a kind of indeterminacy familiar from empirical modeling—then there is no propositional Benacerraf problem. If the indeterminacy is necessary and not the result of any empirical aspect—such as arises in the choice between the set and function versions of the worlds theory—then there is a propositional Benacerraf problem.

Having brought out the complexities of the propositional Benacerraf problem itself, let us turn to the complexities in its application.

3. A DEFENSE OF A WORLDS THEORY

One kind of worlds theorist employs sets of worlds (or characteristic functions thereof) to *represent* or *model* propositions. Of course, one properly lodges the Benacerraf problem at the level of the *target* of a model. If the Benacerraf problem fails at the level of the modeler's target, then this kind of worlds theorist may escape the Benacerraf problem. But the cost of such a view may be that the modeler isn't really giving us any insight into the *metaphysics* of proposition—she "merely" provides representations, as the attitude goes. I want to suggest that we can plausibly combine the worlds theory and modeling methodology, and that the target of such models are Benacerraf resistant, at least on some ways of understanding the approach. The plausibility of the approach, then, will rest upon whether the targets really are Benacerraf immune and whether the attitude of "merely" providing representations can be resisted. To show that the latter part is plausible, it needs to be shown that theorizing about propositions obliquely via a model is a fruitful approach to propositions. To show as much, I think we need to appreciate two things. First, that modeling in metaphysics isn't isolated to just propositions but constitutes a mainstream methodology. Second, that when we look at what propositions *do*, even on contemporary, mind-first theories—ones which have taken to investigate the metaphysics of propositions "directly"—what propositions *do* on such theories is only clearly revealed at the level of the model, just as with the world-theoretic modeler.

If sets of worlds represent or model propositions, then what, at the level of fundamental ontology, are propositions? There are two general options: The target of the propositional models may be *sui generis* entities or not. It's not always clear what a given philosopher means by *sui generis*. The idea is supposed to be that some entity is of a type that we can't reduce to some class of entities we already recognize. In the propositional context, we may mean that propositions can't be reduced or identified with some other entities such as sets of worlds, event types, syntactic-semantic facts, etc. So, in one sense, propositions, if *sui generis*, form their own ontological "category" in that we count them among the basic entities in the universe that we don't identify with or reduce to something more basic. The individuation of what counts as an ontological category is somewhat unclear, though, so in another sense, we may still assimilate propositions to this or that ontological "category" in the sense that we may think they are objects or properties, abstract or concrete, etc. For example, according to Schiffer's theory,[9] propositions are *sui generis*, yet they are things rather than, say, properties, and they are abstract rather than concrete.

Many may find the *sui generis* view implausible, and I count myself in that camp. And if you find

[9]See Schiffer (2003).

primitive representational powers difficult to understand, you will also find it difficult to understand why one *sui generis* proposition pairs with one proposition model over another. Before moving to the non-*sui generis* option, I want to register what I think is an under-appreciated fact—that Lewis' argument against magical ersatzism in *On the Plurality of Worlds* targets such views of propositions as well.[10] Here is an extremely compressed version of the argument: *Sui generis* propositions will stand in the *selecting* relation to this concrete world just in case a proposition represents it. This relation is either internal—depending only on the intrinsic natures of the elements—or external—depending on the intrinsic natures of the *relata* taken together. If it's internal, then there must be some differences among the *sui generis* propositions to explain why the snow-is-white proposition stands in the relation while the snow-is-green proposition doesn't. Being a vast collection of abstract, *sui generis* entities, the only difference between the two propositions would seem to be that one represents this world while the other doesn't. This leads to a vicious circularity: A proposition p is selected because it represents the world as thus and so. But why does it represent the world as thus and so? Because it is selected if and only if the world is thus and so. If the relation is external, then we have an unacceptable necessary connection between various propositions (those that represent this concrete world) and this concrete world. But the goings on at this concrete world and the relations that it enters into with abstract *sui generis* propositions seem related, if at all, only contingently and not necessarily. This is Lewis' argument adapted for *sui generis* propositions.

I don't want to characterize Lewis' argument as decisive—it's complicated, as usual. van Inwagen (1986) argues that he isn't sure what, exactly, is wrong with the argument, but if the argument does work, then it will threaten the relation of set membership, which may count as a kind of reductio of Lewis' argument. And Jubien (1991) provides a comprehensive overview of Lewis' argument, weighing in against, though Jubien (1991, p.266) thinks structureless, primitively representational *sui generis* entities may serve as models themselves, but not the targets of models. I am, before moving onto the non-*sui genris* option, registering a problem with the *sui genris* option. This problem, however, is not a Benacerraf problem. If one identifies *sui generis* entities with propositions at the level of fundamental ontology and models them with world-theoretic devices, whatever problems such a view may incur, the view will avoid the Benacerraf problem. The problem fails at step one since there isn't a surfeit of candidates by stipulation.

The more plausible option is that sets of worlds model propositions that, at the level of fundamental ontology, are non-*sui generis* entities. What kind of entity? *Properties* are a natural answer. First this squares

[10] See Lewis (1986, §3.4).

hermeneutically with the ways that some theorists[11] describe possible worlds and set-theoretic devices—that the former are "ways the world might be" and that the latter *explain* or *explicate* such ways of the world's being. And second, there are explicit views of propositions in the literature according to which propositions are properties.

Property views fit into three broad categories. First, there is the view of Lewis (1983a) according to which propositions are property self ascriptions.[12] Second, we have Jeff Speaks' recent view according to which the proposition that p is the property (or 1-adic relation) of *being an* x *such that* p.[13] According to the final view that I want to focus on, propositions are 0-adic properties (or relations).[14] We canonically represent an n-adic relation as a sentence with n places missing, e.g. the two-place relation of *loves* as *_loves_*. When we put some particular object, say Bob, in one of the slots, we get the property *Bob loves_*. Then if we add another object, say Mary, we get the zero-adic property *Bob loves Mary*, which is the proposition that Bob loves Mary. These properties that are propositions aren't (typically) maximal in that they don't settle every aspect of a possible world; they only settle one if instantiated—that it is a p world. One nice feature of this account is that it provides a clean demarcation between properties that are propositions and those that aren't in terms of adicy. Let's call such propositions *world properties*, and the target of our modeler's models will be such properties.

Our modeler, then, will understand, at the level of ontology, p's being the case in terms of property instantiation. And there is good reason to think that at this level, at the level of properties, the Benacerraf objection fails. However the details of a property view play out, there is good reason to think that the Benacerraf problem doesn't apply to properties. If I say that the world is a certain way, there won't be multiple candidate ways—just as if I say some particular object is some way, say that a ball is red, there won't be multiple candidates to choose from in accounting for the ball's redness. In other words we either won't have a Benacerraf problem or if we do, it will explode to include all properties.

The world property view has the added benefit of avoiding Lewis' argument against ersatzism refashioned to target *sui generis* propositions. Propositions, on this view, don't represent ways the world could be but *are* ways the world could be—these propositions don't *have* or *determine* truth conditions but *are* truth conditions.[15] The second horn of Lewis' dilemma arose from an untoward necessary connection between a

[11]Stalnaker often says that possible worlds *explicate* or *explain* propositions, or that propositions *determine* a set of worlds/function from worlds to truth values: see (Stalnaker, 1999, p.32, p.33) and (Stalnaker, 1978, p.79).

[12]These propositions play a limited role for Lewis.

[13]See Speaks' papers in King et al. (2014).

[14]van Inwagen (2004, fn.17) endorses the zero-adic view, and Swoyer & Orilia (2014, §7.5) explicate such a view.

[15]This is Stalnaker's view. See Stalnaker (2011, ch1 & ch5).

representation and some state of this concrete world: Why does this *sui generis* proposition get paired with this state of the actual world rather than this other one? If propositions are world properties, the question, then, isn't why the proposition p represents the world as being p—a legitimate question—but rather, why does a p world instantiate the property of being p rather than some other property?—an unintelligible question.[16]

Like the *sui generis* view, the Benacerraf problem doesn't apply, but unlike the *sui generis* view, the property view is more plausibly combined with world-theoretic devices as models. The *sui generis* option and modeling methodology clashed because it was difficult to see why any given set of worlds (or characteristic function) would pair with one proposition rather than another, at least if the *sui generis* view is something like Schiffer's. The property view has enough structure and detail to set up a fruitful isomorphism. Stalnaker (1987, p.9) captures the idea that I'm after with an example concerning the relation between certain physical properties and numbers. Why is it that we represent certain physical properties, such as height, in terms of numerical properties relative to a scale—say, being 5′9″? It's because physical properties such as height and weight have a structure in common with (positive) real numbers such that we can pick out these properties with numbers. In doing so, we understand having a certain physical property in terms of an object relating to a number given a unit. Thus we have a natural intermediary to approach physical properties via numerical properties given a unit, and we give a rigorous account of such properties via measurement theory, which we cast in terms of numerical quantities. In this number–height case, there is a natural pairing between the elements of the model and elements of the target and likewise for world theoretic devices. We pair the world-property p with the set of worlds that have one thing in common—that they instantiate p—but otherwise vary in every other possible way. In the number-height case, there is a systematic relationship between the similarities and differences in the model and in the target. The same goes for world-theoretic devices and world properties. Such a relationship allows us to explore relationships among the members of the model, which often lead to discoveries, and transfer those insights to the target.

The move that I'm considering—claiming that what were taken to be propositions are models of some other Benacerraf-immune entities—isn't clearly available to other theorists. The other common target of the problem is the tuple theory, according to which propositions are n-tuples containing constituents related to the sentence expressing the related proposition and whose order mirrors that sentence's syntax. On one way of expressing the theory, the proposition p is the tuple $\langle \langle e_1, ..., e_n \rangle, R^n \rangle$ where $\langle e_1, ..., e_n \rangle$ is an n-ary tuple of objects and R^n is an n-ary relation, and p is true just in case the entities stand in the relation. The theory is an

[16]Stalnaker makes his point in the context of worlds, but the point applies to propositions.

obvious Benacerraf target since, ontologically,

$$\langle \langle e_1, ..., e_n \rangle, R^n \rangle \neq \langle R^n, \langle e_1, ..., e_n \rangle \rangle,$$

and since both tuples work equally well by the relevant theorist's lights. In particular, as Schiffer (2016, pp.2553-2554) notes, working well for this theory means making clear what the *truth conditions* of a given proposition are—as we've just laid out—and the identity conditions, which depend, some say to the theory's credit, not only on the factual truth conditions (i.e., possible worlds truth conditions) but also on the proposition's constituents and their order. But, as Schiffer further notes, if we take tuples as models of propositions, it's hard to see what the target is. In other words, there won't be some non-*sui generis* entity that answers to those two defining features, for they can't be any kind of set-theoretic entity since the Benacerraf problem will remain, and no other sort of target entity is obviously forthcoming. Of course, the tuple theorist can just posit entities with those two features to serve as the target of her tuples, but such a move would seem more like a wish list for a theory rather than a theory.

Construing world-theoretic devices as models isn't an eccentricity to avoid the Benacerraf problem. Indeed, Paul (2012) argues that modeling is *the* main methodology of metaphysics, the result being that both science and metaphysics share methodology but differ on subject matter. According to the mainstream view of modeling in the philosophy of science literature, the *semantic view*, a theory consists in a family of structures—models—that themselves consist of entities and relations, usually mathematical, along with an interpretation of those structures. Then, the theory is true if, and only if, one of its models is relevantly isomorphic to its target.[17] According to Paul, both science and metaphysics engage in modeling, and we understand the difference between the two subjects not in terms of methodology but in terms of subject matter.

The view of this subsection fits the mold of metaphysical modeling that Paul describes. Paul (2012, p.12) considers as an example a mereological composition theory according to which some xs compose a y if but only if the x activity constitute a life. The models in this case are abstract objects that stand in part–whole relations and these structures allegedly represent parts composing wholes that constitute a life. The metaphysical theory is the class of abstract structures and the target of the model are the parts and wholes that those models represent. In the propositional modeler case, the abstract family of structures are sets of possible worlds or functions from worlds to truth values. The target, as we've seen, are world properties.

[17] Paul (2012, §2).

The interpretation connects the set structures to the instantiation of these properties. One structure will be a set that collects worlds that have one thing in common—the instantiation of p. The interpretation is especially conspicuous on the function version—for it doesn't really matter *what* we pair the p and non-p worlds with, so long as we have two distinct elements of the model. True and false are common, but we could use 1 and 0, or, to use Soames' example meant to show that such "propositions" don't really have truth conditions, a dog and a cat. These structures are then isomorphic to various world properties, and we study the world properties and their relations obliquely via the set-theory-powered model. That the interpretation component is essential brings out the absurdity of thinking that world-theoretic devices are themselves intrinsically representational—a point that has sometimes gone unrecognized. For example, we find Field (1986, p.101), in a review of Stalnaker's *Inquiry*, saying that

> [I]f we take as the object of the belief-state [that Caesar crossed the Rubicon] the set of possible worlds in which Caesar crossed the Rubicon. . . that too is a conception of objects of belief as intrinsically representational.

The reason being is that the set "can be construed as representing the world as such that Caesar crossed the Rubicon." It *can* be construed as he says, but it needn't be. Stalnaker never took his set-theoretic devices to be intrinsically representational.

To recap, we can think of proposition theory construction as proceeding in two steps.[18] In step one we find some entities, *sui generis* or otherwise, whose existence we accept. Properties, one might think, are less controversial than propositions, so reducing propositions to properties counts as progress since accepting properties into one's ontology is commonplace. In step two, we give good reason to think that those entities are propositions. Whereas historically a lot of the debate was on step one—and perhaps still is for *sui generis* theorists—I think, nowadays, step two is where all the action is, and this is the step where the Benacerraf problem arises. If we are on step two and decide that some entity e isn't the proposition that p, then we'll think that because e doesn't fill the right propositional roles, e isn't instrumentally acceptable, or e doesn't satisfy empirical criteria if the proposition theory is mind first. If I've been successful in this subsection, then this particular worlds theorist escapes at least the Benacerraf problem.

Finally, we should resist the attitude that such a theorist "merely" models/represents propositions instead of telling us what they *really* are. We've said, in this case, what propositions really are: they are properties. One may find this view lacking in detail, but that shouldn't be surprising. The reason that one employs models in an explanation is because much of the theoretical payoff occurs at the level of the model.

[18]This is how King (2007, p.25) describes giving an account of propositional structure.

The properties view, at the level of ontology, translates naturally to the worlds view, which provides an elegant understanding of these basic propositional relations, all in terms of set theory, which illuminates relations among property instantiations. So, e.g., p will entail q, at the level of ontology, just in case p can't be instantiated without q also being instantiated—or, at the level of the model, just in case p is a subset of q.

Now consider basic propositional relations on contemporary theories. First, consider Soames' theory, which we haven't yet explained. According to this account, to entertain a proposition is to engage in a cognitive activity—the mental act of predication, which for Soames is a primitive mental act, basic among others. When an agent sees o as red, there is an event token of her predicating redness to the object o. Now of course this event token can't be the proposition that o is red. Instead, Soames takes the cognitive event type of predicating redness to o to be the proposition that o is red. Now suppose we take any two of King's propositions—complex syntactic-semantics facts—or two of Somes' propositions—mental-action-involving event types—and ask whether one is logically stronger than the other, or one entails the other, etc. At this level, the level of ontology, I don't know the answer. I would know the answer, however, if I could model these propositions' truth conditions. But this is the same situation the modeler of this section is in.

Let us quickly address the issue of whether properties themselves might be susceptible to a Benacerraf problem. Have we not moved the bump around in the rug? First, were we to take properties as *sui generis*, as I find plausible, then, as we've noted, this option for properties has the benefit of avoiding Lewis' argument from *Plurality*. Second, it's better, it seems to me, to reduce propositions to properties and take properties as *sui generis* than to take both propositions and properties as two distinct classes of *sui generis* entities. Last, if there is a Benacerraf problem for properties, then the problem isn't a problem for the propositions theorist *in particular* and targets a class of entities that a lot of philosophers accept.

References

Armour-Garb, B. & Woodbridge, J. A. (2012). The story about propositions. *Noûs*, 46(4), 635–674.

Bealer, G. (1998). Propositions. *Mind*, 107(425), 1–32.

Benacerraf, P. (1965). What numbers could not be. *The Philosophical Review*, 74(1), 47–73.

Caplan, B. & Tillman, C. (2013). Benacerraf's revenge. *Philosophical Studies*, 166(1), 111–129.

Clarke-Doane, J. (2013). What is absolute undecidability? *Noûs*, 47(3), 467–481.

Field, H. (1986). Stalnaker on intentionality: On robert stalnaker's inquiry. *Pacific Philosophical Quarterly*, 67, 98–112.

Fletcher, L. (2013). Why it isn't syntax that unifies the proposition. *Canadian Journal of Philosophy*, 43(5-6), 590–611.

Heim, I. & Kratzer, A. (1998). *Semantics in Generative Grammar*. Oxford: Blackwell.

Jubien, M. (1991). Could this be magic. *Philosophical Review*, 100(2), 249–267.

Jubien, M. (2001). Propositions and the objects of thought. *Philosophical Studies*, 104(1), 47–62.

King, J. C. (1995). Structured proposition and complex predicates. *Noûs*, 29(4), 516–535.

King, J. C. (2007). *The Nature and Structure of Content*. Oxford: Oxford University Press.

King, J. C. (2009). Questions of unity. *Proceedings of the Aristotelian Society*, 109(cix, pt. 3), 257–277.

King, J. C. (2016). On propositions and fineness of grain (again!). *Synthese*, (pp. 1–25).

King, J. C., Soames, S., & Speaks, J. (2014). *New Thinking About Propositions*. Oxford: Oxford University Press.

Lewis, D. (1975/1983b). Languages and language. In *Philosophical Papers*, volume 1 chapter 11, (pp. 163–188). Oxford: Oxford University Press.

Lewis, D. (1979/1983a). Attitudes de dicto and de se. In *Philosophical Papers*, volume I chapter 10. Oxford: Oxford University Press.

Lewis, D. (1986). *On the Plurality of Worlds*. Oxford: Blackwell.

Lewis, D. (1988). Vague identity: Evans misunderstood. *Analysis*, 48(3), 128–130.

Moore, J. G. (1999). Numbers, and the problem of arbitrary identification. *Synthese*, 120(2), 229–263.

Paul, L. A. (2012). Metaphysics as modeling: The handmaiden's tale. *Philosophical Studies*, 160(1), 1–29.

Schiffer, S. (2003). *The Things We Mean*. Oxford: Oxford University Press.

Schiffer, S. (2016). Cognitive propositions. *Philosophical Studies*, 173(9), 2551–2563.

Soames, S. (2010). *What is Meaning?* Soochow University Lectures in Philosophy. Princeton: Princeton University Press.

Soames, S. (2015). *Rethinking Language, Mind, and Meaning*. Princeton: Princeton University Press.

Stalnaker, R. (1970/1999). Pragmatics. In *Context and Content: Essays on Intentionality in Speech and Thought*, Oxford Cognitive Science Series chapter 1, (pp. 31–46). Oxford: Oxford University Press.

Stalnaker, R. (1987). *Inquiry*. Cambridge: The MIT Press.

Stalnaker, R. (1999/1978). Assertion. In *Context and Content* chapter 4. Oxford: Oxford University Press.

Stalnaker, R. (2011). *Mere Possibilities: Metaphysical Foundations of Modal Semantics*. Princeton: Princeton University Press.

Swoyer, C. & Orilia, F. (2014). Properties. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Fall 2014 edition.

van Inwagen, P. (1986). Two concepts of possible worlds. *Midwest Studies in Philosophy*, 11(1), 185–213.

van Inwagen, P. (2004). A theory of properties. In D. W. Zimmerman (Ed.), *Oxford Studies in Metaphysics*, volume 1 (pp. 107–138). Oxford: Oxford University Press.

Weisberg, M. (2013). *Simulation and Similarity: Using Models to Understand the World*. Oxford University Press.