
Jago, Mark (2013) Impossible worlds. Noûs, 2013 . pp. 1-16. ISSN 0029-4624

NOÛS

NOÛS 00:0 (2013) 1–16
doi: 10.1111/nous.12051

# *Impossible Worlds*

Mark Jago
Department of Philosophy
University of Nottingham

Impossible worlds are representations of impossible things and impossible happenings. They earn their keep in a semantic or metaphysical theory if they do the right theoretical work for us. As it happens, a worlds-based account provides the best philosophical story about semantic content, knowledge and belief states, cognitive significance and cognitive information, and informative deductive reasoning. A worlds-based story may also provide the best semantics for counterfactuals. But to function well, all these accounts need use of impossible and as well as possible worlds. So what are impossible worlds? Graham Priest claims that any of the usual stories about possible worlds can be told about impossible worlds, too. But far from it. I'll argue that impossible worlds cannot be genuine worlds, of the kind proposed by Lewis, McDaniel or Yagisawa. Nor can they be ersatz worlds on the model proposed by Melia or Sider. Constructing impossible worlds, it turns out, requires novel metaphysical resources.

## 1. Impossible Worlds

Impossible worlds are worlds according to which impossible things happen. If a world represents that such-and-such, and it is impossible that such-and-such, then that is an impossible world. Impossible worlds (just like possible worlds) earn their keep in a theory based on the theoretical work they do for us. If impossible worlds are set-theoretic constitutions from actual entities (as I will argue here), then they exist whether we like it or not. So the interesting question is why it is philosophically worthwhile to discuss their logical and metaphysical nature. A worlds-based account provides the best account of epistemic and doxastic notions of content, including knowledge and belief states (Hintikka 1962), cognitive significance and information (Chalmers 2010; Jago 2009), and the content of informative deduction (Jago 2013b). As these are all *hyperintensional* notions, drawing distinctions between logically equivalent contents, those stories must include impossible as well as possible worlds.

Similarly, a worlds-based account provides the best semantics for counterfactuals (Lewis 1973; Stalnaker 1968). Yet to make good sense of counter-possible conditionals, whose antecedents could not have been true, we require impossible as

well as possible worlds to be part of the story (Brogaard and Salerno 2008; Nolan 1997; Read 1995; Routley 1989). Some counter-possible conditionals, such as

(1) If Linear Logic had been the One True Logic, then the TONK rules would have been valid rules of inference;
(2) If Fermat's Last Theorem had been false, then I would have been a lemon;

are clearly false (not trivially true, as the Stalnaker-Lewis approach tells us). On a worlds-based approach, such truths require worlds where the antecedent is true and the consequent is false. Such worlds are, of course, impossible. Hence the need for impossible worlds in a non-trivial theory of counter-possibles.

(Just how we should think of similarity between impossible worlds is not at all clear, but then again, neither is it clear how we should think about similarity between possible worlds; see Nolan 1997. An attractive idea is that, for any possible world $w$, any impossible world is further from $w$ than any possible world is from $w$. Then impossible worlds will not interfere with the evaluation at possible worlds of counterfactuals with possible antecedents.)

One can of course resist these moves in various ways. One could flatly deny that epistemic and counterfactual concepts are hyperintensional. Perhaps we really do know all consequences of what we know, and perhaps counter-possible conditionals are always trivially true. Lewis (1996) and Stalnaker (1984) defend this view. Stalnaker in particular is at pains to explain away the appearance of hyperintensionality via a metalinguistic approach. But this is implausible in general: my failure to know how best to proceed in this game of chess is not a failure to grasp what certain words mean. Alternatively, one could supplement a possible worlds-based account of these notions with linguistic structure, rather than impossible worlds. Chalmers (2011), Cresswell (1985), King (2007), Salmon (1986) and Soames (1987) take this option. But such approaches do not suffice for, as Frege-problems show, there exist hyperintensional differences without structural differences Ripley (2012).

One might even abandon worlds-based approaches altogether. This last option is extreme. Hyperintensionality worries aside, the world-based approach is the most comprehensive and systematic account of content on offer. It is one of the key theories in formal semantics and is used widely in computer science, artificial intelligence and game theory. We should not abandon the benefits that the theory brings lightly. As always in science, the rational approach is to begin with our best theory, and modify it so as to avoid phenomena which it does not currently accommodate.

For the remainder of the paper, therefore, I will assume that there is desirable theoretical utility to be gained from working with impossible as well as possible worlds. The key question then is, just what are impossible worlds? For some authors, they are worlds which represent some contradiction '$A \wedge \neg A$' as being true (Lycan 1994; Berto 2010). Typically, on this view, the impossible worlds correspond to the relational models of paraconsistent logic (see, e.g., Priest 1987). For others, impossible worlds are worlds governed by some non-classical (e.g., intuitionistic) logic (Cresswell 1973). A far less constrained notion allows that, if it is impossible that $A$, then there is a world which represents that $A$ (Nolan 1997; Priest 2005).

Going further still, one might allow (as Priest 2005 does) that, for any arbitrary set of sentences in some specified language, there's a world which represents all and only those propositions as being true. In order to provide all the fine-grained contents we require, we should accept this latter option. Accordingly, I will develop a position along these lines in §5.

These remarks concern the *granularity* of worlds, the question of whether they may be arbitrarily fine-grained representations, rather than their ontological status. Some argue that we don't need a specific answer to the ontological question in the case of impossible worlds. Graham Priest claims that 'any of the main theories concerning the nature of possible worlds can be applied equally to impossible worlds' (Priest 1997, 580). One main aim of this paper is to argue that this is not so. I'll argue that none of the main theories of possible worlds found in the literature can be extended naturally to make good sense of impossible worlds. The metaphysics of impossible worlds is a pressing issue. Having argued that current theories of possible worlds do not extend to impossible worlds, my second main aim in the paper is to construct a theory of worlds which allows for impossible as well as possible worlds.

The remainder of the paper is as follows. In §2, I argue that impossible worlds cannot be genuine worlds, of the kind proposed by Lewis (1986), McDaniel (2004) or Yagisawa (1988; 2010). In §3, I argue that impossible worlds cannot be ersatz worlds on the model proposed by Melia (2001) or Sider (2002). I present a solution in §4 and finally, in §5, I show how to construct worlds (possible and impossible) using the resources from §4.

## 2. Genuine Worlds

We might take non-actual worlds to exist in much the same way that our own world—the universe in its entirety—exists. I'll call such worlds *genuine* worlds. Lewis (1973; 1986) is the main proponent of this view in the case of possible worlds; Yagisawa (1988; 2010) is its champion in the case of impossible (as well as possible) worlds. An opposing *actualist* view holds that the non-actual worlds are mere representations, or models, of the ways our universe could or could not have been. (Genuine worlds are also representations of ways things could or could not be, but they are not mere representations. They do not represent in a pictorial or linguistic way, as ersatz words do.) There are many of varieties of ersatz world on offer to the theorist, depending on how she wants to represent those ways.

A third alternative, taking a cue from Bolzano (1834) and Meinong (1904), is to accept that there are worlds other than our own but deny that such worlds exist (Priest 2005). I shan't discuss this Meinongian view here. Even setting aside worries about its coherence, there remains the question of whether those non-existent, non-actual worlds are genuine or ersatz in the above sense. Presumably they are not ersatz set-theoretic constructions, since all such constructions exist. If genuine, and genuine worlds are the best theoretical tool for the job, then I'd prefer to say that such entities exist.

It is best not to think of the difference between genuine and ersatz worlds as being between concrete and abstract entities. We need worlds to represent mathematical impossibilities (e.g. that $0 = 1$), impossible properties (e.g. *being a round square*) and other impossible abstract matters. So it cannot be that worlds vary only on concrete matters (as Lewis's worlds do). Consequently, we do well not to define genuine worlds as maximal fusions of spatiotemporally related concrete entities. What marks a world as genuine or ersatz is how that world represents. A genuine world represents the existence of a flying hippo by having a flying hippo as a part. That hippo is a real flesh-and-blood flying animal, the kind you could bump into if you weren't careful. More generally, such worlds represent *de dicto* that $A$ by being such that $A$. Genuine worlds are committed to the existence of non-actual entities, such as flying hippos. Ersatz worlds, by contrast, represent the existence of a flying hippo by picturing, or linguistically describing, a flying hippo. Ersatz worlds are compatible with the actualist slogan that nothing exists but what actually exists.

Lewisian genuine worlds obey the *exportation principle*: if world $w$ represents something as being an $F$, then something is an $F$. For if a genuine world $w$ represents something as being $F$, then $w$ contains an $F$ as a part. And as $w$ is part of the totality of being, that particular $F$ too is part of the totality of being: so something is an $F$. The exportation principle is problematic for any account of impossible genuine worlds, as Lewis (1986) notes. Exporting merely possible entities (or states of affairs) from genuine possible worlds lumbers us with a large and counterintuitive but still consistent ontology. Exporting impossible entities (or states of affairs) from genuine impossible worlds, by contrast, drags us into contradiction. If there is an impossible genuine world according to which there is a round square, then (given exportation) there is a real entity which is both round and square, and hence round and not round: contradiction! So one cannot accept impossible genuine worlds on the Lewisian model.

Yagisawa (1988) tries to bite the bullet here. He holds that 'there are possible worlds in Lewis' sense and also impossible worlds in an equally realistic sense' (1988, 176). He holds the dialethist position (in his 1988 paper), on which there are true contradictions, so that one can 'tell the truth about an impossible thing' (such as our round square) by contradicting oneself (Yagisawa 1988, 203). But merely allowing for true contradictions is not enough to alleviate the worry. Even dialethists want to maintain that many sentences are not true. They make use of a paraconsistent logic, which does not support *ex falso quodlibet*, $A, \neg A \vdash B$. The very point of rejecting this principle is to allow some contradictions to be true, without triviality.

If we can export contradictions from impossible worlds, however, every sentence will be true (simpliciter). Consider any sentence '$A$' and some impossibility '$Fx$'. Then there is an impossible world such that $A \wedge Fx$. If that world is genuine, we can export a genuine $x$ which is such that $A \wedge Fx$, which (classically or paraconsistently) entails that '$A$' is true (simpliciter). Hence, absurdly, any '$A$' whatsoever is true. This is just the conclusion the dialethist wanted to avoid. So even dialethists cannot accept genuine impossible worlds which support the exportation principle.

Yagisawa, in his recent account of modality (Yagisawa 2010), provides an alternative to the Lewisian conception of genuine worlds. He treats modality much as as four-dimensionalists treat temporal matters, whereby entities exist and have properties at a time $t$ by having *temporal stages* at time $t$ which have those properties (Lewis 1986). For Yagisawa, worlds are like times in this respect: entities exist and have properties at a world $w$ by having *modal stages* at world $w$ which have those properties. Bertie is actually beagle-shaped in virtue of having a beagle-shaped actual-stage; he could have been portly in virtue of having a (merely) possible portly world-stage; and he is necessarily canine because all of his world-stages are canine. All of those world-stages (whether the world in question is actual, merely possible or impossible) are genuine parts of reality: they are what we quantify over when using our most unrestricted quantifier.

This approach does not avoid the exportation worry (Jago 2013a). On Yagisawa's view, Bertie has properties-at-world-$w$ in virtue of having $w$-stages with those properties (just as in the temporal case, Bertie has properties-at-time-$t$ in virtue of having $t$-stages with those properties). His $w$-stage is portly, simpliciter; that stage is intrinsically portly, even though Bertie (the collection of all his stages) is not. So the possibility of Bertie's being a portly beagle entails that there is a portly beagle-stage, out there somewhere in modal space. That stage is intrinsically portly, and not merely portly-at-$w$. But by the same token, the impossibility of Bertie's being a portly-and-not-portly beagle entails that there is an intrinsically portly-and-not-portly beagle-stage, out there in (impossible) modal space. So some existent entity is both portly and not portly (simpliciter), and we have not avoided contradiction.

To avoid the exportation worry, one must deny the move from 'according to $w$, $Ax$' to 'something is such that $Ax$'. One way to do this is to insist that all (or at least most) property possession is world-relative: $x$ may be $F$-at-$w$ but not $F$-at-$w'$ (much as I am happy on Sunday but not on Monday), with nothing being $F$ simpliciter. (In particular, on this view, $x$ is not $F$-at-$w$ in virtue of having a $w$-stage that is intrinsically $F$, as on Yagisawa's view. World-stages can play no role in this relational theory.) Then, if it is impossible that $A$, we may infer that there is a world $w$ which represents that $A$, and hence that there something such that $A$-at-$w$. But we cannot infer from this that there is something such that $A$ (simpliciter). We cannot infer that there exists a round square, but only that something is round-and-square-at-$w$ (for some impossible world $w$). McDaniel (2004) defends a view along these lines.

Transposed to the temporal case, the view is precisely what the three-dimensionalist says: objects do not have temporal parts; each object is wholly present at each time. On the modal analogue, each object is wholly present at each world in which it exists. A feature of this account is that one and the same entity may exist according to many worlds, for that entity may bear the *exists-at* relation to more than one world. Because of this feature, the view is sometimes called *genuine modal realism with overlap* (McDaniel 2004).

This approach is just as unsuited for impossible worlds as Lewis's and Yagisawa's, although for a different reason. Suppose to the contrary that impossible worlds are

of the realist-with-overlap variety, and take the case of Richard Sylvan, the New Zealand logician, born Richard Routley. It is impossible for Routley to be other than Sylvan, and hence impossible for Routley but not Sylvan to be a logician. So there is an impossible world $w$ according to which Routley but not Sylvan is a logician. On the overlap view, that is to say that Routley, but not Sylvan, bears the *being a logician* relation to world $w$. So, on this view, Routley bears a relation to $w$ which Sylvan does not bear to $w$ and hence, by Leibniz's law, Routley and Sylvan are not identical (simpliciter). But this is absurd: Routley is (or was) Sylvan! Consequently, overlapping genuine worlds are not a suitable treatment of impossible worlds.

In summary, neither Lewis's nor Yagisawa's nor McDaniels's worlds are suitable candidates for being impossible worlds. Impossible worlds are not genuine worlds (at least, I know of no other account of genuine worlds but these). We need to look elsewhere to get our fill of impossibility. Impossible worlds must be treated as ersatz worlds.

## 3. Ersatz Worlds

Ersatz worlds have a clear advantage over genuine worlds when it comes to impossibilities, for they do not require anything to be impossible or to possess impossible properties (either simpliciter or at-a-world). They are mere representations, and mere representations of impossibility are commonplace. (Just look at an Escher drawing or read the Bible.) Given that we require very fine-grained impossible worlds to make sense of the content of epistemic and doxastic states, the linguistic approach is the way to go. Impossibilities may be incomplete as well as inconsistent, and linguistic representation accounts for incompleteness far better than pictorial representation does. So I will focus on the linguistic approach to constructing ersatz worlds. (Such proposals were put forward, in different ways and for different purposes, by Carnap (1947); Hintikka (1962; 1969), Jeffrey (1983) and, more recently, by Melia (2001) and Sider (2002).)

On the linguistic approach, ersatz worlds are sets of sentences in some 'world-making' language. This can be any language we like, as long as some important constraints are met. First, we need a method of interpreting sentences of the language, so that we can say what a given sentence represents. Second, the language should be disambiguated and precise, so that it is always a determinate matter what a given sentence represents. Third, the language must be expressible enough to represent all the possible and impossible situations we want to represent, and to represent distinct (possible or impossible) situations as distinct situations.

The first two desiderata can be met in a number of ways. Lewis (1986, 145–6), following Carnap (1947), suggests the *Lagadonian* approach, according to which we take particulars to be the names of our language, and properties and relations to be our predicates, each interpreted to refer to itself. Atomic sentences are sequences of an $n$-place predicate (that is, a property or relation) followed by $n$ terms (particulars). Designated set-theoretic constructions will serve as connectives, quantifiers and variables. We allow quantifier-prefixes to be infinitely long and our conjunction

and disjunction symbols to operate on (possibly infinite) sets of sentences, resulting in infinitely long conjunctions and disjunctions.

In this way, we avoid the cardinality objection from Lewis (1973,90) and Bricker (1987, 340–3). The objection, in short, is that there are more possibilities (at least $\beth_2$) than sets of sentences (at most $\beth_1$), and hence sets of sentences cannot represent all the possibilities without conflation. But the objection assumes a countable worldmaking language, and so it is ineffective against the above proposal.

The third desideratum, that the language represent distinct (possible or impossible) situations without conflation, is much harder to meet. This 'problem of descriptive power' is 'an apparently devastating problem' for linguistic ersatzism (Sider 2002, 281). The problem is to represent possible but non-actual particulars, properties and relations without conflation. Since (by actualist assumptions) such entities do not exist, they cannot be invoked as names for themselves. So, it seems, we cannot represent such mere possibilities by naming them.

We can instead describe them, by saying: there are properties and relations $X_1, \ldots, X_n$ and particulars $x_1, \ldots, x_m$ such that $A$ (Skyrms 1981). Yet this approach does not distinguish between what are intuitively distinct possibilities. Bec could have done all Anna did and vice versa and, intuitively, this remains the case even if Anna and Bec are merely possible. Yet descriptive worldmaking sentences will not distinguish between these role-switched situations. Similarly, pairs of possible properties could switch roles and yet, if those properties are merely possible, we have just the one worldmaking sentence to describe them (Lewis 1986, 162).

The argument just given assumes *haecceitism* (Kaplan 1975) in the case of particulars and *quidditism* (Black 2000) in the case of properties. One may deny both. (Lewis (1986) denies haecceitism and dispositional essentialists such as Bird (2007) and Shoemaker (1980) reject quidditism. Heller (1998), who treats properties counterpart-theoretically, also rejects quidditism.) This denial amounts to the claim that each possible property and particular is identified essentially with its role. But consider Wonda, who is wondering whether electrons play this or that role (perhaps it is possible, for all she knows, that electrons have neutral charge). To account for the content of her epistemic state, we need to represent a situation in which electrons do not play the electron-role. The status of quidditism affects whether this situation is possible or impossible, yet we need to represent it regardless. Similarly, Wonda may wonder which role *unicornhood* (or any other alien property) plays. So, in considering how expressive our worldmaking language needs to be, one's take on quidditism (and haecceitism) is beside the point.

Meeting the third desideratum, by solving the problem of descriptive power, is crucial for any ersatz worldmaking language. Melia (2001) and Sider (2002) offer solutions (for the case of possible worlds only). Their solutions utilise similar strategies and so, for brevity, I will discuss Sider's approach only here. His idea is to replace a plurality of ersatz worlds with a single *ersatz pluriverse sentence* (in the worldmaking language), which has the form of a quantifier-prefix:

> There are worlds $w_1, w_2, \ldots$, non-actual particulars $x_1, x_2, \ldots$ and non-actual properties $X_1, X_2, \ldots$ such that

followed by a conjunction of *world-conjuncts* '... $w_i$ ...', each of the form '$x_1$ exists in $w_i$, has property $X_2$, and ...'.

Each world-conjunct fully describes a possible world. And crucially, the quantifiers occur only in the prefix. Because of this, it is easy to say that there are distinct properties $F$ and $G$, and that everything that is $F$ in $w_1$ is $G$ in $w_2$ and vice versa. Sider has given us a single description of all of modal space, rather than a separate description for each ersatz world. What pluriverse sentences represent is none other than a plurality of genuine possible worlds. (In this respect, Sider's approach bears some similarities to Rosen's modal fictionalism (1990).) If successful, Sider has shown how to reap the benefits of Lewis's pluriverse without accepting Lewis's ontology of non-actual entities.

Can Sider's ersatz pluriverse approach be extended to include impossible as well as possible worlds? It faces a painful dilemma: either the extended pluriverse sentence represents *implicitly* by entailment, or it represents *explicitly* by containing an appropriate sentence as a conjunct. Inconsistent world-conjuncts trivialise the former approach. An inconsistent world-conjunct entails every '*A*', so that no two inconsistent world-conjuncts differ in what they represent. But worse, such a pluriverse sentence $S$ will represent that $A$ for any '*A*' whatsoever and so will grossly misrepresents both the possibilities and the impossibilities. In particular, an extended pluriverse sentence cannot represent (by entailment) a structured space of impossible worlds, and hence cannot be used in a theory of epistemic contents or counter-possible conditionals. So, if we are to extend Sider's approach to represent impossibilities, the extended pluriverse sentence cannot represent by entailment.

(Note that it would not help if the extended pluriverse sentence were to represent by some non-classical notion of entailment. If '*A*' entails '*B*' under that notion of entailment, then a world that represents that $A$ will also represent that $B$. But it might be impossible that $A$ and not $B$, in which case, there should be a world which represents that $A$ but does not represent that $B$. This point is crucial if we are to capture fine-grained epistemic contents and counter-possible conditionals using impossible worlds.)

Suppose instead that the extended pluriverse sentence $S$ represents explicitly. $S$ says that there's a world such that $A$ if '*A*' (expressed in the worldmaking language) is a conjunct of some world-conjunct of $S$. This move avoids the trivialisation worry just discussed. But now, if any '*A*' is to be a necessary truth, it must be a conjunct of every world-conjunct. This raises cardinality worries. If set $X$ exists and $W$ is a world-conjunct, then $X$ necessarily exists and so '$X$ exists' must be a conjunct of $W$. That goes for every set $X$. If $W$ is a set-theoretic construction, as is usually assumed, then it must have itself as a conjunct.

This contradicts the axiom of regularity, which implies that sets are no set is a member of itself. (There do exist non-well-founded set theories (Aczel 1988), but these are non-standard. It is pure hubris to think that the correct theory of content should dictate what set theory should look like!) If world-conjuncts are to represent the mathematical universe explicitly (and if sentences are set-theoretic constructions), then world-conjuncts are 'too big' to be sets. They are proper classes. But by definition, no proper class is a member of another class, and so cannot be

conjoined into a quantifier-prefixed pluriverse sentence. So there can be no extended pluriverse sentence that represents explicitly.

Could we escape the dilemma by taking the possible-world-conjuncts to represent implicitly and the impossible-world-conjuncts to represent explicitly? No, we cannot. Consider a situation just like the actual state of affairs (mathematical universe included) which, in addition, contains a round square. That is impossible, and so there should be an impossible world representing it fully. But the situation contains all sets, so cannot be represented explicitly by a worldmaking sentence qua set-theoretic construction. However we define representation, there can be no extended pluriverse sentence that incorporates impossible-world conjuncts.

I have argued that, contrary to Priest's claim (1997, 580), it is not the case that 'any of the main theories concerning the nature of possible worlds can be applied equally to impossible worlds'. Far from it: no theory of genuine worlds, and none of the current theories of ersatz worlds which overcome the problem of descriptive power, can be extended to impossible worlds. In the next two sections, I develop an ersatz theory of worlds which overcomes this deficiency.

## 4. Naming the Possibilities

So far, I have argued that impossible worlds cannot be genuine worlds and that descriptive ersatz approaches, although suited to building impossible worlds, nevertheless conflate what should be distinct (possible or impossible) situations. The moral I draw is that (i) impossible worlds must be ersatz worlds; but (ii) possibilities must be represented referentially, rather than descriptively. Each merely possible particular, property and relation requires a distinct name.

If we continue strictly with the Lagadonian approach, on which each entity denotes itself, then (ii) amounts to the requirement that non-actual entities genuinely exist. This requires an ontology of possibilia, much like Lewis's. Since impossible worlds cannot be genuine worlds (§2), one option to adopt a Lewisian account of possible worlds and use the possibilia they contain to construct ersatz impossible worlds (in the Lagadonian way). This is the view Berto (2010) calls *hybrid modal realism*. On this option, the genuine worlds provide enough non-actual entities for us to represent by name all the possibilities without conflation: they are 'the basic stuff' of the world-building enterprise (Berto 2010, 481). Impossible words are then constructed in some ersatz way, using those possibilist resources.

The view suffers from a serious lack of economy. Suppose we grant that Sider's ersatz pluriverse approach does a good job of representing Lewis's ersatz pluriverse. If we constrain our interest to analysing intensional (as opposed to hyperintensional) content, then Sider's and Lewis's approaches do an equally good job. It is when we turn to hyperintensional contents that Sider's view suffers (§3). So Lewisian non-actual possibilia are not strictly required in the analysis of possible worlds; it is the analysis of impossible worlds that (on the hybrid view) requires them. But this is incredible! If we can construct possible worlds without using exotic non-actual entities, then we should not require exotic non-actual entities to account for impossible worlds. At the least, these considerations undermine the motivation for accepting

non-actual possibilia in the first place and hence undermine the motivation for hybrid modal realism.

An alternative approach to Berto's hybrid modal realism, which affords much the same flexibility, is to adopt Platonic universals into one's ontology. On this view, universals may exist even if uninstantiated. An 'alien' property, such as having $\frac{1}{4}$ charge, is not ruled out of existence. One can then use a broadly Lagadonian approach to constructing worlds, with each property naming itself. How should one represent non-actual particulars on this approach? One option is to accept *haecceitistic* Platonic universals, such as the property of *being Bob Dylan* and (the uninstantiated) *being Sherlock Holmes*. One can then treat such properties as names in the worldmaking language.

I find haecceitistic properties wonderfully obscure. It is highly improbable that there exists a distinct fundamental universal for each existent particular, let alone the non-existent ones. That view would require that the fundamental truths about reality make reference not only to electrons, quarks and so on, but also to you and I, Bob Dylan and Sherlock Holmes. This is unreasonable. One might instead take haecceitistic properties to be non-fundamental and hence decomposable into more basic universals. But that view suffers from the problems of the property-bundle approach to representing particulars (§3): distinct possibilities will be conflated. The objection shows that genuine haecceities cannot be defined qualitatively and hence cannot be treated as property-bundles.

What if one cares about quidditistic differences but not haecceitistic differences between possibilities? Then one could ditch bizarre haecceities and yet stick with the proposal in terms of uninstantiated Platonic universals. This view overcomes the problem of descriptive power and is plausible, so long as one is happy to accept uninstantiated Platonic universals. My concern here is that it is hard to find independent motivation for accepting uninstantiated Platonic universals into one's view. Uninstantiated universals would not be required at all were we to ignore worries about hyperintensionality. It is incredible that uninstantiated Platonic universals should be needed just to account for hyperintensional content, just as it is incredible that Lewisian worlds should be needed solely for that purpose. Perhaps uninstantiated Platonic universals (or Lewisian worlds) can be kept as a back-up option, but I would strongly prefer to do without them.

We cannot name the possibilities (without conflation) without some serious additions to 'ordinary' ontology. What one would like is a serious independent motivation for any such addition to one's ontology. I've argued elsewhere in favour of one such addition to 'ordinary' ontology: the addition of facts (or states of affairs), including negative facts (Barker and Jago 2012). This move affords us a way to name non-actual properties (as I'll argue below) and it has independent motivation. But negative facts are controversial beasts indeed. One needs good reason for accepting their existence. Here, I can indicate how that argument would go only briefly.

It is often said that postulated physical entities 'should make some sort of causal/nomic contribution to the working of the actual world' (Armstrong 2004, 39), and this this thought is often used int he case against negative facts

(Molnar 2000). But if we grant the premise and can show that they do contribute to the make-up of the world, then we have a reason for accepting negative facts. Negative facts (qua real absences, which I identify with negative existential facts) plausibly play a role in causation, as when Bob's not watering his plants causes them to die. They plausibly play a role in the material constitution of holes, gaps, edges and the like, and consequently in the material constitution of entities such as ring-donuts (for which the hole is an essential part!). It's plausible that we can perceive facts (we see that such-and-such is the case) and, if so, it's plausible that some negative facts are amongst the perceivable facts. On this (causal) view of perception, I see that the room is empty, or that there's a gap at mid-on, when I stand in an appropriate causal relation to the fact *that the room is empty* or *that there's a gap at mid-on*.

I've argued for all of these claims in more detail elsewhere (Barker and Jago 2012). But there's a further, independent motivation for believing in negative facts: the demand for truthmakers for all the truths. It's highly plausible that at least some truths (the 'positive' ones) require truthmakers. Yet as I argue elsewhere (Jago 2012), if we need truthmakers for positive truths, then we need truthmakers for 'negative' truths such as 'there are no Vulcans' too. Truthmaker theorists should be truthmaker *maximalists*, who provide truthmakers for all truths. But most maximalist accounts fail (Jago forthcoming). The best maximalist solution, in my view, is to accept an ontology of positive and negative facts. Each step in this chain of reasoning requires careful, detailed argument, and here isn't the place to repeat that reasoning. The salient point here is that there exists independent and (I claim) plausible motivation for accepting negative facts into one's ontology.

How are we to get from the acceptance of negative facts to a worldmaking language which names non-actual entities (without conflation)? We will require that each fact is associated (perhaps primitively) with some particulars and properties, which exist if that fact does. Specifically, a fact *that* $Fa$ must be associated with the property $\lambda x\, Fx$ and the fact *that* $\neg Fb$ must be associated with the property $\lambda x \neg Fx$. One option here is to take facts to be primitive, fundamental entities (Skyrms 1981) from which properties may be *abstracted* (Armstrong 2004) or otherwise constructed. (Abstraction here is not a mental process: the relation between facts and properties should be a mind-independent phenomenon. Just how abstraction is to work needn't detain us here: see Jago (2011) for one account.)

Alternatively, one could take facts to be comprised (in some non-mereological way) of particulars and properties (Armstrong 1997). A third option is to take facts, particulars and properties to be equally metaphysically basic, with primitive relations of association between (e.g.) $F$, $a$ and the fact *that* $a$ is $F$. All that the present argument requires is that one can infer from the existence of the fact to the existence of the associated property.

Now return to the pressing problem of representing alien properties: how are we to represent the property *having* $\frac{1}{4}$ *charge*? Nothing possesses this property. In particular, for any fundamental entity $a$, $a$ doesn't possess *having* $\frac{1}{4}$ *charge*. So, on the current approach, there exists a fact *that* $a$ lacks $\frac{1}{4}$ *charge* and hence an associated (abstracted) property $\lambda x\ x$ *lacks* $\frac{1}{4}$ *charge*. We use this as a predicate

in our worldmaking language, interpreted in the Lagadonian way so as to mean *lacking* $\frac{1}{4}$ *charge*. Now let's add to our worldmaking language a predicate-negation symbol NEG. We stipulate that any predicate (NEG, (NEG, $F$)) represents whatever $F$ represents and that $F$ represents *lacking* $G$ iff (NEG, $F$) represents *having* $G$. Consequently, (NEG, $\lambda x$ *x lacks* $\frac{1}{4}$ *charge*) represents *having* $\frac{1}{4}$ *charge*: precisely the property we wanted to represent!

We can reason similarly with any fundamental alien property in place of *having* $\frac{1}{4}$ *charge*. For suppose that *Fness* is a possible fundamental property. Then (for some fundamental $x$), I claim that there exists either a fundamental fact *that x is F* or a fundamental fact *that x isn't F*. (This is justified by accepting that the fundamental facts are truthmakers for such claims and that all such truths require a truthmaker.) Whichever of these facts exists, we can use the technique just described to represent *Fness* unambiguously. (For the technical details, see Jago 2011.)

I've indicated some independent reasons for accepting negative facts and given one way in which they can be used to represent alien properties. That's progress, but it doesn't give us all we want. It remains to be shown that, having used negative facts to represent alien properties, we can then represent all target possible and impossible situations without conflation. That is the task of the next section.

## 5. Constructing Worlds

In the previous section, I argued that thinking in terms of positive and negative facts (with properties abstracted from those facts) allows us to name alien properties unambiguously and without conflation. Yet representing all the possible properties is just one of the theoretical problems facing the construction of ersatz worlds. There remains the issue of representing possible particulars; and beyond that, we need to represent impossible situations. Let's consider each problem in turn.

I want to revisit the idea, from §3, of representing non-actual particulars using property-bundles. Since we have named all the possible properties, we can help ourselves to any bundle of possible properties. But, as we saw in §3, using property-bundles qua definite descriptions to represent possibilia forces us to conflate what should be distinct possibilities. These are possibilities which are qualitatively indistinguishable, but with different particulars playing the relevant roles.

I propose that we represent such possibilia using property-bundles, interpreted in line with proper names and demonstratives rather than definite descriptions. To see the picture, let's temporarily adopt a possibilist fiction by pretending that all possible particulars exist. A property-bundle qua worldmaking name $n = \{F_1, \ldots, F_n\}$ is interpreted as denoting the unique $F_1 \wedge \cdots \wedge F_n$. We then interpret worldmaking sentences containing $n$ as saying that that very entity is a certain way. The sentence $Gn$ says only that that entity is $G$: it does not attribute any of the bundle-properties $F_i$ to that entity. (As a consequence, the worldmaking sentence $\neg F_i n$ is consistent.) The properties in the bundle pin down what the name $n$ represents, but not how it represents that entity as being. Semantically, $n$ functions similarly to Kaplan's (1978)) 'DTHAT(the $F_1 \wedge \cdots \wedge F_n$)': it is non-attributive and semantically rigid.

When we drop the possibilist fiction, we should no longer say that such property-bundles refer, for there are no possibilia or fictional entities to which they may refer. (For simplicity, I'll continue to say that a bundle represents Sherlock Holmes. By this, I mean only that that bundle, combined with a predicate $F$, represents *that Sherlock is F*.) Despite not referring, the semantic function of these names-qua-bundles is unchanged: they contribute non-attributively to representations. The properties in the name-qua-bundle fix its semantic value qua subject term, but to use that name (in a worldmaking sentence) is not to predicate those properties.

To see why this approach works, consider how the problem of descriptive power has to be set up (§3). We first describe some possible situation involving non-actual particulars. Then we say: *that* particular could have been such-and-such and *that other* particular could have been so-and-so. We make as to refer back, non-descriptively, to the particular previously described. If that way of setting up the problem makes sense (and in particular, if it is coherent to make as to refer back to something which, by actualist lights, doesn't exist), then the line I'm proposing is a genuine solution. If there's something suspect to the idea of rigidly and non-descriptively representing *that thing* (which, as it happens, doesn't exist), on the other hand, then the description of the problem too is nonsensical. I say there is a genuine problem and that using property-bundles as semantically rigid, non-attributive worldmaking names solves the problem.

This approach is particularly useful for dealing with fictional entities such as Sherlock Holmes. The worldmaking name $h$ representing Holmes is a property-bundle {*is a detective*, *lives at 221b Baker Street*, *is a cocaine addict*, ...}. As Conan Doyle wrote the novels, he associated the English name 'Sherlock Holmes' with $h$, by virtue of ascribing the properties $F \in h$ to Holmes. He did not create an abstract entity (how could he?); neither did he describe someone of whom it is *a priori* that he is a detective living in 221b Baker Street. It is coherent to think that Holmes lived elsewhere, or that he was a chemist, or that he did all Watson was said to have done and vice versa.

We can represent this latter situation consistently only if our worldmaking names for Holmes and Watson are non-attributive. The semantically rigid property-bundle approach deals with these situations well. (It may be that none of these Holmes-involving situations are metaphysically possible (Kripke 1980). I remain neutral on the issue: my concern is to represent a range of situations, including Holmes-involving situations, possible or otherwise.)

We must also address the question of representing impossible situations. Logical impossibility requires logical structure: no logical impossibility is logically primitive. Since our worldmaking language contains logical vocabulary (connectives, variables and quantifiers), the language can represent the structure of any logically impossible situation. If we can represent $a$'s being $F$, for example, then we can represent a situation in which $a$ is both $F$ and not-$F$. Similarly for logically complex metaphysically impossible situations such as Bill's being a married bachelor. Since we can represent each conjunct of this conjunctive situation and conjoin them in the worldmaking language, it is easy to represent Bill's impossible situation. Perhaps there

are logically primitive but metaphysically impossible situations: Holmes-involving situations might be such cases (Kripke 1980). Even if so, no problem. We have the resources to represent such entities *de re*, using semantically rigid, non-attributive property-bundles.

The worldmaking language described so far has a serious shortcoming. As one can think about a given entity under various guises (and we want to reflect this at the level of content), we must allow for there to be distinct contents relating to one and the same entity. One may believe that Hesperus is $F$ but not believe that Phosphorus is $F$, and so we need to distinguish between the contents *that Hesperus is F* and *that Phosphorus is F*. If these contents are distinct, then the worldmaking sentences representing *that Hesperus is F* and *that Phosphorus is F* must also be distinct. Yet at present, we have only one name for Venus in our worldmaking language, namely Venus itself. So we have only a single worldmaking sentence representing (by name) *that Hesperus is F* and *that Phosphorus is F*. (We can also denote Venus by definite description. But definite descriptions are attributive and so, for certain *F*s, sentences describing Venus will always say more than *that Hesperus is F*.)

The most promising response to this worry is to replace in the worldmaking language all Lagadonian names with non-attributive property-bundles, all of which uniquely pick out the particular in question. On this approach, all worldmaking names are property-bundles, irrespective of whether the supposed referent exists. There will in general be multiple such names-qua-bundles for each actual particular. When distinct names qua property-bundles are uniquely satisfied by the same particular, those names co-refer and thus give us distinct ways of representing (by name) one and the same individual.

A worry remains, however. Our worldmaking names corresponding to 'Hesperus' and 'Phosphorus' represent the same particular, Venus, because those names are each property-bundles uniquely satisfied by Venus. But not so for the worldmaking names corresponding to 'Superman', 'Clark Kent', 'Pegasus', 'Darth Vader' and so on. Each such property-bundle does not pick out any actual particular. Yet one might well want to say that, in some sense, the English names 'Superman' and 'Clark Kent' represent the same thing (in a way that 'Superman' and 'Darth Vader' do not). If so, we should say likewise for the corresponding worldmaking names. But the approach I've described so far does not allow us to say this. We can say that, according to some world $w$, Superman is Clark Kent; but we cannot say that the worldmaking names corresponding to 'Superman' and 'Clark Kent' represent the same person, simpliciter.

The best response to the worry is to deny that those worldmaking names co-represent. Co-representation is possible only in the case of names for existent entities, on this view: worldmaking names $B_1$ and $B_2$ co-represent just in case some particular uniquely satisfies both property bundles. As a consequence, distinct worldmaking names for fictional entities will never co-represent. Nevertheless, it remains true that, according to the Superman stories, Superman is Clark Kent. These stories represent that Superman is Clark Kent; another story may (consistently) represent otherwise.

On the view I am suggesting, we can say something true using (English or worldmaking) names such as 'Superman' only by using *that*-clauses (or equivalents

such as 'according to the fiction, . . . '). It is true that Alice hopes that Santa will come, that Bob wishes that Superman exists, and that the stories represent that Superman is Clark Kent. But it is false that 'Superman' and 'Clark Kent' (or the corresponding worldmaking names) represent something, and hence false that they represent the same thing. One might object: surely each Superman story represents the same superhero! This I must deny, for there is no such superhero. Instead, we should say that the word 'Superman' has the same content and meaning in each story. Moreover, we can treat all the Superman stories as a combined fiction, in which 'Superman' has a consistent meaning (and according to which Superman exists and does such-and-such).

If we suppose that this approach (or something like it) works, then we have a good handle on how worldmaking sentences represent all that we want them to represent. Ersatz worlds are then sets of worldmaking sentences. These worlds represent by inclusion (and not by entailment). A world $w$ represents that $A$ iff there is a worldmaking sentence $S \in w$ which represents that $A$. To allow the maximum of flexibility, I will count every non-empty set of worldmaking sentences as a world. Consequently, many worlds are incomplete, both in the sense that (for some '$A$') they represent neither that $A$ nor that $\neg A$, and in the sense that they may represent that $A$, which entails that $B$, and yet fail to represent that $B$. Such a world may represent that all men are mortal and represent that Bob is a man, yet fail to represent that Bob is mortal. Both kinds of incompleteness are required in an adequate account of impossible worlds, suitable for making sense of fine-grained contents.

We can classify worlds via a number of syntactic tests. A world is *prime* iff it contains a disjunct of each disjunction it contains. A world is *maximal* iff it contains either '$A$' or its negation, for each sentence '$A$' (of the worldmaking language). A world is *logically possible* (with respect to some logic $L$) iff it is closed under $L$-consequence but does not contain every sentence of the worldmaking language. Just when a world represents a genuine metaphysical possibility is a thorny issue, and one I do not propose to say anything about here. Ersatzers typically resort to primitive modal facts. That's fine with me. My concern has been to represent all the possible and impossible situations, not to give a reductive account of modal talk.

## References


Aczel, P. 1988. *Non-Well-Founded Sets*. CSLI Lecture Notes Vol. 14. Stanford: CSLI Publications.

Armstrong, D. 1997. *A World of States of Affairs*. Cambridge: Cambridge University Press.

Armstrong, D. 2004. *Truth and Truthmakers*. Cambridge: Cambridge University Press.

Barker, S. and Jago, M. 2012. Being Positive About Negative Facts. *Philosophy and Phenomenological Research*, 85, 117–138.

Berto, F. 2010. Impossible Worlds and Propositions: Against the Parity Thesis. *The Philosophical Quarterly*, 60, 471–486.

Bird, A. 2007. *Nature's Metaphysics: Laws and Properties*. Oxford: Oxford University Press.

Black, R. 2000. Against Quidditism. *Australasian Journal of Philosophy*, 78, 87–104.

Bolzano, B. 1834. *Lehrbuch der Religionswissenschaft: ein Abdruck der Vorlesungshefte eines Ehemaligen Religionslehrers an einer Katholischen Universität*. Sulzbach: J. E. v. Seidelschen.

Bricker, P. 1987. Reducing Possible Worlds to Language. *Philosophical Studies*, 52, 331–355.

Brogaard, B. and Salerno, J. 2008. Remarks on Counterpossibles. *Synthese*, 190, 639–660.

Carnap, R. 1947. *Meaning and Necessity*. Chicago: University of Chicago Press.

Chalmers, D. 2010. The Nature of Epistemic Space. In A. Egan and B. Weatherson (eds.), *Epistemic Modality*, pp. 60–107. Oxford: Oxford University Press.
Chalmers, D. 2011. Propositions and Attitude Ascriptions: a Fregean Account. *Noûs*, 45, 595–639.
Cresswell, M. 1973. *Logics and Languages*. London: Methuen.
Cresswell, M. 1985. *Structured Meanings*. Cambridge, MA: MIT Press.
Heller, M. 1998. Property Counterparts in Ersatz Worlds. *The Journal of Philosophy*, 95, 293–316.
Hintikka, J. 1962. *Knowledge and Belief: an Introduction to the Logic of the Two Notions*. Ithaca: Cornell University Press.
Hintikka, J. 1969. *Models for Modalities: Selected Essays*. Dordrecht: Reidel.
Jago, M. 2009. Logical Information and Epistemic Space. *Synthese*, 167, 327–341.
Jago, M. 2011. Setting the Facts Straight. *Journal of Philosophical Logic*, 40, 33–54.
Jago, M. 2012. The Truthmaker Non-maximalist's Dilemma. *Mind*, 121, 903–918.
Jago, M. 2013a. Against Yagisawa's Modal Realism. *Analysis*, 73, 10–17.
Jago, M. 2013b. The Content of Deduction. *Journal of Philosophical Logic*, 42, 317–334.
Jago, M. forthcoming. The Cost of Truthmaker Maximalism. *Canadian Journal of Philosophy*.
Jeffrey, R. 1983. *The Logic of Decision*. Chicago: University of Chicago Press.
Kaplan, D. 1975. How to Russell a Frege-Church. *The Journal of Philosophy*, 72, 716–729.
Kaplan, D. 1978. Dthat. In P. Cole (ed.), *Syntax and Semantics*, pp. 221–253. New York: Academic Press.
King, J. 2007. *The Nature and Structure of Content*. New York: Oxford University Press.
Kripke, S. 1980. *Naming and Necessity*. Oxford: Blackwell.
Lewis, D. 1973. *Counterfactuals*. Cambridge, MA: Harvard University Press.
Lewis, D. 1986. *On the Plurality of Worlds*. Oxford: Blackwell.
Lewis, D. 1996. Elusive Knowledge. *Australasian Journal of Philosophy*, 74, 549–567.
Lycan, W. 1994. *Modality and Meaning*. Dordrecht: Kluwer.
McDaniel, K. 2004. Modal Realism with Overlap. *Australasian Journal of Philosophy*, 82, 137–152.
Meinong, A. 1904. Über Gegenstandstheorie. In A. Meinong (ed.). *Untersuchungen zur Gegenstadstheorie und Psychologie*, pp. 1–50. Leipzig: Barth.
Melia, J. 2001. Reducing Possibilities to Language. *Analysis*, 61, 19–29.
Molnar, G. 2000. Truthmakers for Negative Truths. *Australasian Journal of Philosophy*, 78, 72–86.
Nolan, D. 1997. Impossible Worlds: A Modest Approach. *Notre Dame Journal of Formal Logic*, 38, 535–572.
Priest, G. 1987. *In Contradiction: A Study of the Transconsistent*. Dordrecht: Martinus Nijhoff.
Priest, G. 1997. Sylvan's Box. *Notre Dame Journal of Formal Logic*, 38, 573–582.
Priest, G. 2005. *Towards Non-Being*. Oxford: Clarendon Press.
Read, S. 1995. *Thinking About Logic*. Oxford: Oxford University Press.
Ripley, D. 2012. Structures and Circumstances: Two Ways to Fine-Grain Propositions. *Synthese*, 189, 97–118.
Rosen, G. 1990. Modal Fictionalism. *Mind*, 99, 327–354.
Routley, R. 1989. Philosophical and Linguistic Inroads: Multiply Intensional Relevant Logics. In J. Norman and R. Routley (eds.), *Directions in Relevant Logic*, pp. 269–304. Dordrecht: Kluwer.
Salmon, N. 1986. *Frege's Puzzle*. Cambridge, MA: MIT Press/Bradford Books.
Shoemaker, S. 1980. Causality and Properties. In P. van Inwagen (ed.), *Time and Cause: Essays Presented to Richard Taylor*, pp. 109–35. Dordrecht: Reidel.
Sider, T. 2002. The Ersatz Pluriverse. *The Journal of Philosophy*, 99, 279–315.
Skyrms, B. 1981. Tractarian Nominalism. *Philosophical Studies*, 40, 199–206.
Soames, S. 1987. Direct Reference, Propositional Attitudes and Semantic Content. *Philosophical Topics*, 15, 47–87.
Stalnaker, R. 1968. A Theory of Conditionals. In N. Rescher (ed.), *Studies in Logical Theory*, pp. 98–112. Oxford: Blackwell.
Stalnaker, R. 1984. *Inquiry*. Cambridge, MA: MIT Press.
Yagisawa, T. 1988. Beyond Possible Worlds. *Philosophical Studies*, 53, 175–204.
Yagisawa, T. 2010. *Worlds and Individuals, Possible and Otherwise*. New York: Oxford University Press.