The Limits of Realism

The Limits of Realism

Tim Button

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

First Edition published in 2013

Impression: 1

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data

Data available

ISBN 978–0–19–967217–2

Printed in Great Britain by
CPI Group (UK) Ltd, Croydon, CR0 4YY

Vous m'expliquez ce monde avec une image. Je reconnais alors que vous en êtes venus à la poésie: je ne connaîtrai jamais.

You explain this world to me with an image. I realise then that you have been reduced to poetry: I shall never know.

Le mythe de Sisyphe

Albert Camus

Contents

Introduction

Metaphysics longs for magic. I take this thought from William James, who claims that metaphysicians operate in that tradition of folklore, where to know something's name is to control it:

> 'God,' 'Matter,' 'Reason,' 'the Absolute,' and 'Energy,' are so many solving names. You can rest when you have them. You are at the end of your metaphysical quest.[1]

These particular solving names have fallen out of fashion. But contemporary metaphysicians have found new solving names: 'Causal', 'Fundamental', 'Grounding', and 'Natural'. Added to this, metaphysicians have solving *gestures* (desk-thumping), solving *typographies* ('Really', '*really*', or 'REALLY'), and solving *locations* (the ontology room). These many solutions are so many incantations.

Or so I am often inclined to think. But inclinations are not arguments.[2] I started writing this book with the aim of turning inclination into argument; of *demonstrating* that metaphysics belongs to magic. I finished writing this book with this aim realized in part, but only in part. Indeed, I have come to believe that partial success mixed with partial failure is inevitable. We can show that there are limits to realism, and limits to 'anti'-realism. But most of all, we can show that there are limits to what we can show at all. The aim of this book is to sketch all of these limits.

That aim is realized by considering the relevance of both *semantics* and *scepticism* to the realism debate. Their relevance might seem surprising, given that there is a very direct approach to the realism debate which mentions neither. On the direct approach, you start by choosing some putative entities, such as kitchen appliances, mountains, electrons, virtues, or numbers. Then you simply ask yourself: *should we be realists about those entities?*

That direct question mentions neither scepticism nor semantics. However, it is simply too direct. Until we know what it means to be a *realist* about something, the question has no clear content. Fortunately, there is a way to give the question *some* content, whilst retaining the idea that the

[1] James (1907: 28).
[2] Even if James (1907: 8–9) is right that temperament is 'the potentest of all our premises'.

realism debate should be approached directly. We treat the question as asking us: *are those entities constitutively independent of us?*[3]

Often, this question is clear enough. We did not build mountains with our bare hands, and we did not build them with our naked minds either. So, mountains are constitutively independent of us. *Mountains are real.* Of course, this is hardly profound, but the direct approach to the realism debate takes this to be a sign of its level-headedness.

There is, however, a rival approach. Rather than starting with the question of whether mountains, virtues, or numbers are independent of us, we might instead start by investigating our use of the *words* 'mountain', 'virtue', and 'number'. Perhaps this will give us some insight into the role of mountains, virtues, or numbers in our lives. The realism debate thus shifts from considering the world directly, to considering the world by examining our words.

Philosophers who want to approach the realism debate directly are rarely impressed by this semantic turn. They complain: *I don't want to talk about how we talk about objects; I want to talk about the objects!* And this complaint is good, as far as it goes. No one should deny the difference between talking about the world and talking about words.

However, this difference does not establish that semantics is irrelevant to the realism debate. For suppose we can mount an argument with the following conclusion: *By your own lights, you are simply unable to talk about any of the objects you wanted to talk about.* Then the complaint in the previous paragraph will ring utterly hollow. In the face of this argument, a philosopher who wants to approach the realism debate by talking about objects will have to explain *how we can* talk about them. They will have to engage with some semantic questions after all.

We are, then, looking for an argument that forces realists to consider semantic questions. A natural place to look is in the discussion of meaning scepticism that took America by storm in the 1960s onwards, culminating in Hilary Putnam's model-theoretic arguments against a certain kind of realism.[4] This is my entry point to the realism debate.

I begin by outlining an earnest position, called *external realism*. This countenances good old-fashioned *Cartesian* scepticism and, for this reason,

[3] This is Devitt's (1984a) recommendation.

[4] Putnam's first full public presentation of the argument was in December 1976 (Putnam 1977), but this was somewhat anticipated during an exchange with Dummett in April 1976 (Putnam 1979c).

Putnam's model-theoretic arguments show that it must countenance incoherent meaning scepticism. (Perhaps I should emphasize right now, given how frequently it is misunderstood, that Putnam does not *embrace* meaning scepticism; instead, he uses it as a *reductio* of opposing positions, such as external realism.) This will prompt a loss of faith in external realism, and the remainder of this book explores how we might reconcile ourselves to this loss. We shall find no particular position that deserves our allegiance, not even *internal realism*, which is defined colourlessly in diametric opposition to external realism. By the end of the book, we shall have little more positive to say than that we must be neither external realists, nor internal realists, but something vaguely in-between.

I have just mentioned Putnam for the first time. During the course of this book, I shall mention him several hundred times more. Putnam has provided some of the most powerful arguments in support of the idea that metaphysics longs for magic, and some of the most interesting criticisms of those arguments. All of these form the central subject matter of this book, and the book is consequently a blend of exegesis, reconstruction, and novel contribution. In the interests of providing an unbroken line of narrative, I have not drawn sharp lines between these components. Nevertheless, I do my best to indicate in the text where I am agreeing with, where I am parting company from, and where I am outright disagreeing with (various time slices of) Putnam. Doubtless I shall prove mistaken on certain specifics, but one thing is absolutely certain: without Putnam, this book could not have existed. I owe him tremendous thanks for the pleasure I have drawn from reflecting on his work, and I hope that the depth of my gratitude comes through in what follows.

There are many more people to thank. For artwork: Helen Macdonald (Figure 2.1) and Lawrence Lek (the cover image). For help with German translations: Maike Albertzart and Christine Tiefensee. For community: the philosophy departments in Cambridge, Harvard, and UT Austin. For comments, corrections, discussions, suggestions, and questions over several years: Sharon Berry, Thomas David Button, Christina Cameron, Adam Caulton, Colin Chamberlain, Elijah Chudnoff, Tim Crane, Matti Eklund, Daniel Elstein, Peter Godfrey-Smith, Peli Grietzer, Alex Grossman, Hallvard Lillehammer, Jon Litland, Christina McLeish, Steven Methven, Adrian Moore, Sam Nicholson, Seb Nye, Alex Oliver, Huw Price, Hilary Putnam, Tim

Storer, Peter Sullivan, Amie Thomasson, Sean Walsh, Lee Walters, Nathan Wildman, and four anonymous readers.

A handful of people are owed very special thanks. First, Michael Potter, who was my PhD supervisor when I began what would become this book. Second, Brian King, who first made me realize that I love philosophy and who keeps reminding me that I do. Third, Adam Stewart-Wallace and Rob Trueman, who have both influenced me more than I can express in mere footnotes. And finally—and always—Ben, who may not care about cherries and cats or brains in vats, but who cares deeply about me.

Part A
External Realism

1

The picture of external realism

The model-theoretic arguments should erode one's faith in external realism. That is the central claim of Part A of this book. The aim of the book as a whole is to learn how to live with this loss of faith, and without much by way of a substitute. But to understand all of this, we must first understand the faith held by external realists, and how it came under attack.

External realism goes by many aliases: *metaphysical* realism; *capital-'R'* Realism; *robust* realism; *desk-thumping* realism; *genuine* realism; the choice of adjective is yours. The purpose of that adjective is to indicate just how *serious* the position is about its realism. I have chosen to stick with 'external' realism, because that adjective is most evocative of the philosophical picture which this position employs. That picture is of reasoning from a 'God's Eye point of view'.[1]

Pictures are often difficult to attack, and philosophical pictures are no exception. Fortunately, Hilary Putnam presents us with three principles that flow naturally from the external realist's picture, and shows us how to attack those. These principles collectively enshrine what I shall call the *Credo* of external realism. The precise interpretation of the Credo is not itself a straightforward task, but the Credo is the only place to start.

1.1 Credo in Independence

Generically, realists believe that *we* do not determine what there is; rather, the *world* does that. Of course, this bold statement needs some caveats. No one can deny that we have built cities, sculpted statues, started wars, and invented Esperanto. *These* things would not exist without us. But *most* of

[1] Putnam (1980a: 100; 1981c: 49; see also 1982a: 38; 1983a: x, xviii).

the universe did, does, and will manage perfectly well without us. In this spirit, then, the external realist advances the first of three principles:

> **The Independence Principle.** The world is (largely) made up of objects that are mind-, language-, and theory-independent.[2]

Belief in Independence is undoubtedly a necessary component of realism. It would be controversial to claim that it is *sufficient* for realism,[3] but I shall postpone discussion of this point.

1.2 Credo in Correspondence

If the objects of the world are (largely) independent of our minds, languages, and theories, we need to know how we are able to think, speak, and theorize about them. An early realist answer to this question was the *copy theory* of truth. This held that true representations must copy reality. But this claim is implausible, as William James noted.[4] There are many different conventions for true representation, beyond mere copying. It is not clear why we should care about copying. And in any case, the world is just too rich to make copying it a plausible enterprise.

The *correspondence theory* endeavoured to accommodate these objections. For the correspondence theorist, true theories do not aim to copy the world, but aim only at some kind of *structural* similarity. In the simplest case, names pick out individual objects and predicate letters pick out properties and relations. (The more nominalistically inclined may prefer to read this, and similar expressions throughout this chapter, as saying that predicate letters lasso certain objects together.) Thus for the correspondence theorist, an atomic sentence of the form '$Rt_1, \ldots, t_n$' is true exactly when the relation named by 'R' holds between the objects named by 't_1', ..., 't_n'. This view is enshrined in the second principle of external realism:

> **The Correspondence Principle.** 'Truth involves some sort of correspondence relation between words or thought-signs and external things and sets of things.'[5]

[2] Putnam (1977: 484–5; 1980a: 100; 1981c: 49; 1982a: 30; 1989: 352).
[3] Devitt (1983: 292–3; 1984a: 3–4, 34–40; 1991; 2010: 52–3) makes this controversial claim.
[4] James (1904: 467–9).
[5] Putnam (1980a: 100; 1981c: 49; see also 1982a: 30; 1989: 352; 2000: 126).

Certain versions of the correspondence theory go further, and insist that each individual sentence must correspond with an individual object of a special sort: a *fact*, or *state of affairs*. The Correspondence Principle takes no stance on the existence or otherwise of such objects. It enshrines only the modest view that names pick out objects and predicates pick out properties and relations.

Certain other versions of the correspondence theory make the extremely strong claim that 'there is just One True Theory of the fixed mind-independent Reality'.[6] That Theory is penned in *Ontologese*, a language which is perfectly suited to limning reality's distinguished structure. Putnam often attributes this doctrine to external realism, and there is probably more explicit support for it now than in any recent time. Nevertheless, at the outset my external realist will treat the existence of Ontologese as a pious legend, rather than as Credo-enshrined orthodoxy. She subscribes only to the much more modest Correspondence Principle.

But perhaps even the correspondence theory of truth is too ambitious, since it insists that there must be a structural similarity between the subject/predicate distinction of some language and the object/property distinction in the world. *Truthmaker theory* abandons this structural requirement. It requires only that true sentences (typically) have some worldly (mind-independent) truthmaker.

In these three successive incarnations—copy, correspondence, and truthmaker—the quintessentially realist theory of truth has steadily sacrificed detail. For the sake of generality, it might be best to pin only a truthmaker theory on the external realist, since she could always add greater detail by postulating that the truthmakers have the structure required for full-fledged correspondence. As it happens, though, the choice between truthmaker and correspondence theory will make almost no difference to our subsequent considerations. (If anything, it makes external realism marginally more difficult to attack, if it adheres to a correspondence theory rather than a mere truthmaker theory.) So it will do no harm to assume that the external realist accepts the Correspondence Principle.

[6] Putnam (1989: 352; see also 1979a: 288; 1980a: 100; 1981c: 49; 1982a: 30).

1.3 Credo in Cartesianism

For the external realist, then, truth consists in correspondence. Falsity will consist in a certain failure of correspondence. A worry now arises. Let us imagine that humans have arrived at some scientific theory which is marvellous in every imaginable respect that we can investigate. It predicts all our observations perfectly; it retrodicts equally well; it explains without fault; it is simple to work with; it is harmonious and beautiful; perhaps to learn about and employ the theory gives the sensation of seeing directly into the mind of God. Such a theory is canonically *ideal*.[7] It gives every indication of being true. But if the world really is as mind-, language-, and theory-independent as the external realist believes, then might not appearances be radically deceptive? For example: perhaps the theory tries to talk about hadrons, but there are no hadrons to correspond with hadron-talk. Or perhaps, although the theory says otherwise, the world really came into existence exactly five minutes ago, so that the theory gets everything wrong about the past.[8] Perhaps we are really all just brains in vats,[9] or being deceived by René Descartes's *malin genie*. Such thoughts lead to the third principle of external realism:

> **The Cartesianism Principle.** Even an ideal theory might be radically false.[10]

This captures a kind of anxiety about Cartesian-style sceptical scenarios (hence the name). This anxiety is undoubtedly associated with a certain breed of realism.[11] One way to put the idea is as follows: only God, whose knowledge is not constrained by *mere* science, could say for sure whether any given theory is true or false. For this reason, we might say that the external realist's 'favorite point of view is a God's Eye point of view'.[12]

[7] This follows Putnam's (1977: 485) explication.

[8] A worry mentioned by Russell (1921: 159–60).

[9] Putnam (1977: 485, 487; 1981c: 5–17; 1986b: 110–13; 1989: 352) takes this worry as central to external realism.

[10] Putnam (1977: 485; 1980b: 473; 1989: 352; 2000: 127–31).

[11] Since Dummett (1963: 153), several authors have drawn this link between realism and the threat of Cartesian scepticism. Van Inwagen (1988: 95) explicitly holds that 'truth is radically nonepistemic'. This might seem to express the Cartesianism Principle, but Van Inwagen (1988: 104–7) claims only that truth is a mind-independent relationship between *propositions* (which are 'necessarily existent, abstract' mind-independent objects) and the (largely mind-independent) world; this is just a little icing on the Independence Principle.

[12] Putnam (1980a: 100; 1981c: 49; see also 1982a: 38; 1983a: x, xviii).

We therefore have three Principles to which the external realist adheres: Independence, Correspondence, and Cartesianism. These together form the Credo of external realism. It is worth noting that many philosophers who self-define as 'realists' will reject elements of the Credo. The most salient example is Putnam himself, who called himself an internal *realist* whilst rejecting *external* realism. Accordingly, it is not worth objecting that the Credo fails to capture *realism*,[13] for external realism is just one denomination within the broad church of realism. Moreover, it is not an irrelevant sect. The three Principles of the Credo have, individually and together, a distinguished line of descent.[14]

1.4 Modelling the Credo

To set up Putnam's attack on external realism, our next task is to offer a model-theoretic treatment of the external realist's Credo. (For those totally unfamiliar with model theory, I offer a beginner's primer in Appendix I.)

Suppose the external realist has presented her favourite theory, which she hopes corresponds to the mind-independent world. We shall assume that this theory is expressed in a formal language containing individual constant symbols ('c_1', 'c_2', ...), predicates ('R_1', 'R_2', ...), and function symbols ('f_1', 'f_2', ...). We are to imagine the external realist's mind-independent world as the intended model, $\mathcal{W}$, of the external realist's theory. $\mathcal{W}$'s domain, W, is to be thought of as the objects which make up the world. Each constant, 'c', is mapped to the object $c^{\mathcal{W}}$ in the domain of the intended model. Each predicate, 'R', is mapped to a set $R^{\mathcal{W}}$ of objects (or pairs, triples, etc. of objects) drawn from W; these are to be thought of as the extensions of the predicates. A similar treatment is offered for function symbols.

This model-theoretic treatment sits extremely well with the external realist's Correspondence Principle. The idea of mapping the words of the formal language onto the objects of the world gives us an excellent means for thinking about reference. When the external realist says that the name 'c' refers to some object, a, we can parse this as stating that on our

[13] Though Hansen (1987: 95–7) and Van Inwagen (1988: 107–8) make this objection. Putnam (2012b: 62) now acknowledges that it was a mistake to call the position 'metaphysical realism' (rather than, for example, 'external realism'), since one can be a realist and want to do some metaphysics without being (what I have called) an external realist.

[14] As Putnam (1982a: 30) notes.

interpretation, 'c' denotes $c^{\mathcal{W}} = a$. Derivatively, this gives us an excellent way to understand what it would mean for an atomic sentence to be *true*. Where 'R' is an n-place predicate, and 't_1' through to 't_n' are terms of the language, we have the schema:

$$\text{`}Rt_1 \ldots t_n\text{' is true iff } \left\langle t_1^{\mathcal{W}}, \ldots, t_n^{\mathcal{W}} \right\rangle \in R^{\mathcal{W}}$$

Similar clauses are used to explain the use of function symbols, and correspondence for non-atomic sentences is explained recursively.

The model-theoretic treatment also sits well with the Independence Principle. On the model-theoretic treatment, the external realist conceives of the world as a particular model. Which model the world happens to be does not depend upon our minds, languages, or theories. There are plenty of models and, a priori, any of them *might* be the actual world. This captures some of the idea of Independence. In turn, this gives us a way to understand the Cartesianism Principle: since it is a mind-independent matter which model represents the world, there is always the possibility of stating a bunch of falsehoods about the world.

The model-theoretic approach therefore validates all three Principles of the external realist's Credo. Unsurprisingly, then, this way of modelling external realism has become firmly entrenched among contemporary metaphysicians. To take one example: in passing, Cian Dorr sketches a three-step recipe for doing *fundamental metaphysics*.[15] First, we describe the '*fundamental ontology*'. Next, we specify the '*fundamental ideology*'. Finally, we lay down 'some *laws*'. This is to appropriate model theory for metaphysical ends: Dorr has called domains 'the fundamental ontology'; he has called the (formal) language 'the fundamental ideology'; and he has called axioms 'laws'.

Dorr describes this three-step recipe as a 'standard approach'.[16] I would go further, and describe it as an *orthodoxy* for post-Quinean metaphysicians.[17] Willard van Orman Quine gave us the division between *ontology* and *ideology* upon which Dorr relies.[18] Moreover, Quine told us that a theory's ontological commitment is determined by its regimentation in first-order logic, since 'to be is, purely and simply, to be the value of a variable.'[19] Of course, Quine was no friend to metaphysics, let alone to 'fundamental'

[15] Dorr (2011: 139). [16] Dorr (2011: 139).
[17] Putnam (2004: 15–16, 78–81) plausibly traces the resurgence of metaphysics back to Quine.
[18] Quine (1951b). [19] Quine (1948: 32; see also 1951a: 67; 1957: 17).

metaphysics (whatever that might be). But metaphysicians working after Quine have typically embraced the idea that a theory's quantifiers gauge its 'ontological commitment', and have then quibbled (in a non-Quinean fashion) on the exchange rate between ontology and ideology.[20] The three-step recipe is *the* method of contemporary analytic metaphysics, and it enshrines a model-theoretic approach.

Despite the centrality of these ideas to contemporary metaphysics, an external realist might object to the use of model theory. After all, it is unclear why someone who wants to be a realist (of any stripe) about concrete entities, like cats, cherries, and electrons, should be forced to believe in *models*, which are abstract mathematical entities.

An external realist with such qualms should feel free to think of model theory as nothing more than a convenient tool for discussing correspondence between words and world. This will allow her to *use* model theory to model (in the informal sense) her Credo, without committing her to the *existence* of the models themselves.[21] If she digs in her heels and absolutely refuses to employ model theory in any capacity, then she will have to explain what her correspondence relation amounts to with both hands tied behind her back. Bear in mind that model theory was designed, among other things, precisely in order to discuss different ways to make theories true.

This, however, gives rise to a more serious concern. Model theory was developed by mathematicians for mathematical ends. So, whenever we deploy a result from model theory in order to demonstrate something about external realism, we should ask whether that result is simply an artefact of a theory that is appropriate for mathematics but too permissive for metaphysics. This is not a question that can be settled once and for all in advance; we shall simply have to proceed with caution.

In the meantime, model theory gives us an excellent way to think about the philosophical claims of external realism. We are not asking the external realist to *believe* in the existence of abstract models, but only to employ the model theory. And we are allowing the external realist the freedom to argue, at a later point, that model theory is too permissive. In short, there can be no general objection to a model-theoretic approach to the external realist's Credo.

[20] D. Lewis (1986: 4) is probably the most famous exponent.
[21] Putnam (1994a: 263) offers a similar thought; thanks also to Sharon Berry for suggesting this.

2

The model-theoretic arguments

In Chapter 1, I explained the external realist's Credo: Independence, Correspondence, and Cartesianism. I then explained why model theory offers an attractive way to understand these Principles. With the model theory in place, I shall now explain the model-theoretic results that Putnam brings to bear against external realism.

Putnam's model-theoretic arguments fall broadly into two camps: *indeterminacy* arguments and *infallibilism* arguments. Together these show that, for any ideal theory, there are guaranteed to be many different ways to make it true. This strikes at the very heart of the external realist's Credo.

2.1 Indeterminacy arguments

The indeterminacy arguments aim to show the following: *If there is any way to make a theory true, then there are many ways to do so.* This provides the external realist with an embarrassment of riches, since any one of many candidate correspondence relations would function perfectly well as *the* correspondence relation that she mentions in her Correspondence Principle. The most obvious threat is that it will be utterly *indeterminate* which of these relations *is* truth itself.[1]

The *permutation argument* is the easiest indeterminacy argument to explain.[2] Imagine that we were to lay out all the objects in the world, together with various labels (names) for them, and with other labels (predicates) for

[1] As Putnam (1981c: 33) notes, these relate to Quine's (1960: §§7–16; 1968; 1970; 1987) arguments for the indeterminacy of translation.

[2] See Putnam (1981c: 33–5, 217–18). Permutation arguments have a lengthy history, and are to be found in Frege (1893: §10), Carnap (1928a: §§153–5), and Newman (1928: 145–6). Closer to our present concerns, they occur in Jeffrey (1964: 82–4), Winnie (1967: 226), Field (1975: 376–7), Wallace (1979: 307), and Davidson (1979: 229–30).

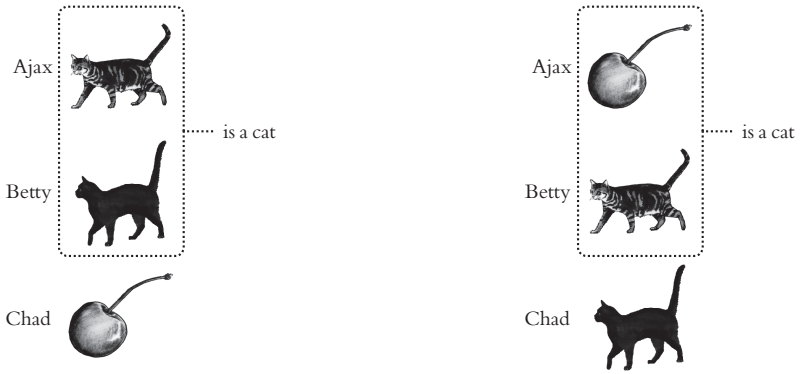

Figure 2.1 Two models of the theory whose sentences are just 'Ajax is a cat', 'Betty is a cat', and 'Chad is not a cat'. We begin with an assignment of names to objects and predicates to collections of objects (left). This model is transformed into the permuted model (right) by shuffling around all of the objects without altering the labels.

collections of them. Suppose we now shuffle the objects around. So long as we do not disturb the labels, exactly the same sentences will come out as true after the shuffling as were true before the shuffling. A very simple instance of this idea is illustrated in Figure 2.1,[3] but the idea is wholly general. If $\mathcal{W}$ models the external realist's favourite theory, then a *permutation* over the domain of $\mathcal{W}$ will systematically shuffle the objects around, yielding a distinct but *isomorphic* structure, $\mathcal{P}$. Since $\mathcal{W}$ and $\mathcal{P}$ are isomorphic, they make exactly the same sentences true or false. Thus we obtain:

The Permutation Theorem. Any theory with a non-trivial model has many distinct isomorphic models with the same domain.

(I offer a formal proof in Appendix I, and a definition of a 'non-trivial model'.) Sticking with our models $\mathcal{W}$ and $\mathcal{P}$, we can regard either of the following schemes as presenting us with the reference relation:

$$\text{`}t\text{'} \text{ refers to } t^{\mathcal{W}}$$
$$\text{`}t\text{'} \text{ refers to } t^{\mathcal{P}}$$

[3] Inspired by Putnam's (1977: 484, 490) diagrams, and by his use of cats and cherries (1981c: 33–5) in discussing permutations. Thanks to Helen Macdonald for these wonderful drawings.

Equally, we can treat either of these two schemes as presenting us with the correspondence relation (for atomic sentences):

$$\text{`}Rt_1 \ldots t_n\text{' is true iff } \left\langle t_1^{\mathcal{W}}, \ldots, t_n^{\mathcal{W}} \right\rangle \in R^{\mathcal{W}}$$
$$\text{`}Rt_1 \ldots t_n\text{' is true iff } \left\langle t_1^{\mathcal{P}}, \ldots, t_n^{\mathcal{P}} \right\rangle \in R^{\mathcal{P}}$$

In short, settling the truth values of every sentence in some language is insufficient to pin down the reference and correspondence relation that the external realist mentions in her Correspondence Principle.

A second indeterminacy argument is only marginally less intuitive. If the external realist's theory has a model, $\mathcal{W}$, then the theory must at least be consistent. But if it is consistent, it also has a *numerical* model. More precisely:[4]

> **The Completeness Theorem.** Any consistent, countable set of sentences has a model whose domain only contains natural numbers.

The proof of this result is slightly more involved than that of the Permutation Theorem, and certainly harder to illustrate. But the overarching upshot is that the supposedly 'intended' model, $\mathcal{W}$, and the numerical model, $\mathcal{N}$, will likely be distinct. This will occur when $\mathcal{W}$ is uncountable, or if some of the objects in $\mathcal{W}$ are not numbers. Nevertheless, both $\mathcal{W}$ and $\mathcal{N}$ make exactly the same sentences true and false. So, as in the case of the Permutation Theorem, we have free choice concerning which model to think of as providing us with *the* correspondence relation.[5]

Further indeterminacy arguments can be offered by employing more sophisticated theorems from model theory (often some Löwenheim–Skolem result). If my focus were the philosophy of mathematics, I would need to discuss such theorems in detail. However, such a discussion would add very little to this book. (I elaborate on this remark in Appendix I.) Indeed, it could be rather harmful, since it might suggest that the model-theoretic arguments depend upon some sophisticated mathematics. Emphatically: *they do not*. They depend only on simple, elementary considerations. So the only indeterminacy arguments that I shall consider in this book are those from the Permutation and Completeness Theorems. (Indeed, in this regard I am largely following Putnam: in all his discussion

[4] I shall not distinguish between first-order and higher-order theories until Chapter 4.

[5] This obviously relates to Skolem's (1922) scepticism that there are any 'genuinely' uncountable sets. The Completeness Theorem also led Quine (1964) to realize that any theory could be given a model in the natural numbers, leading him to investigate (and reject) a kind of modern-day Pythagoreanism.

of the model-theoretic arguments, Putnam only once essentially invokes a result other than the Permutation and Completeness Theorems, and only there to raise problems for a certain kind of mathematical platonism.)[6]

In short, the indeterminacy arguments show that if there is any way to make a theory true, then there are guaranteed to be many ways to do so. This poses a challenging question to the external realist: *Which relation is the correspondence relation?*

2.2 Infallibilism arguments

There are many things that the external realist can say in response to this challenge. I shall consider some of these in Chapter 3. Before that, I shall outline the second kind of model-theoretic argument. Arguments of this kind aim to show the following: *If a theory is ideal, then there must be some way to make the theory true.* The immediate threat is that ideal theories will turn out to be infallible, thereby undermining the external realist's Cartesianism Principle, which requires that an ideal theory might be false. I shall therefore call such arguments *infallibilism* arguments.

Putnam's most famous direct attack on the Cartesianism Principle is his celebrated brain-in-vat argument. Though this is extremely important, it is not model-theoretic in nature, so I shall reserve discussion of it for Part C of this book.

The model-theoretic infallibilism argument runs as follows. If a theory is ideal, then it is presumably at the very least consistent. After all, if it is inconsistent, then it (classically) entails everything, and so it can hardly count as ideal. But if it is consistent, then, by the Completeness Theorem mentioned earlier, it has a model containing only natural numbers. If we like, we can substitute this for a model containing only concrete objects, by appealing to an idea like that used in the Permutation Theorem: we simply swap distinct natural numbers for distinct concrete objects. (I offer a formal explanation of this in Appendix I.) But the general problem is simply that, if the theory is ideal, then there is guaranteed to be some way to make the theory true. So, if the external realist thinks that an ideal theory might

[6] Putnam (1980b: 467–9) invokes the Downward Löwenheim–Skolem Theorem in discussing the Axiom of Constructibility in set theory; I discuss this in Button (2011). Putnam (1977: 485) invokes a General version of the Completeness Theorem which connects with the Upward Löwenheim–Skolem Theorem, but the vanilla Completeness Theorem will do the job (I explain this in Appendix I).

really be false, then she faces a second challenging question: *What prevents the model given by the Completeness Theorem from making the theory true?*[7]

2.3 Correspondence versus more generic truthmaking

Putnam's model-theoretic arguments apply pressure to the external realist's conception of truth (and falsity) in terms of correspondence between words and world. However, as noted in §1.2, truthmaker theory has recently gained support among realists. It is worth briefly explaining why abandoning correspondence in favour of a more generic truthmaker theory will not affect the model-theoretic arguments.

Recall from §1.2 that, in so far as truthmaker theory differs from correspondence theory, it is less prescriptive about the nature of the truthmakers and truthbearers. For a correspondence theorist, there must be some structural relationship between a sentence and the way the world is. By contrast, truthmaker theorists (who are not also correspondence theorists) are committed only to the view that a sentence is made true by some truthmaker. The sheer *lack* of structure demanded by such truthmaker theorists means that it is easier to run model-theoretic arguments against them than against correspondence theorists.

To offer a straightforward indeterminacy argument, we can simply permute over the truthmakers, so that each sentence is mapped to an 'unintended' truthmaker. The intuitive idea is depicted in Figure 2.2, and a general permutation result for truthmaker theorists is trivial to obtain.[8]

A completeness theorem is equally easy to obtain. Choose a single object, Top, which is to be the truthmaker for every true sentence and for no false sentence. (If the truthmaker theorist also believes in falsemakers, then choose a single object, Bottom, to be the falsemaker for every false sentence and for no true sentence.) If it is *possible* to make all of the sentences of a theory true, then the theory can obviously be given a 'model' using just Top (and Bottom). We can now use this completeness result to offer indeterminacy and infallibilism arguments, exactly as before.

[7] Putnam (1977: 485–6; 1980b: 472–4; 1989: 353); Taylor (2006: 58–9) also puts the argument this way. The history of this argument can be traced back to Newman (1928: 144–5) and Winnie (1967: 227); I return to Winnie's argument in §5.3.

[8] Let ι be the function that maps each true sentence to its 'intended' truthmaker. Let π be any bijection over the range of ι. Then $\iota \circ \pi$ is a new truthmaking function.

Figure 2.2 Two permuted models for truthmaker theorists. We begin with an initial assignment of true sentences to truthmakers (left). This is transformed into the permuted assignment (right) by a permutation which shuffles around the truthmakers without altering the sentences.

The truthmaker theorist may raise a *modal* objection against the models presented, on the grounds that a truthmaker should in some sense 'necessitate' its truth.[9] However, the correspondence theorist can raise the same objection. Moreover, since the correspondence theorist has access to the full structure of correspondence theory, her objection will be trickier to deal with. I shall return to all this in §3.4.

For now, the general point is straightforward. Truthmaker theorists (who are not also correspondence theorists) adopt a more permissive attitude towards truth than correspondence theorists do. It is therefore always strictly easier to raise problems for them, than for correspondence theorists. Accordingly, I shall not consider (generic) truthmaker theory in the remainder of this book, and shall instead stick with correspondence theory.

The moral of the chapter is easy to summarize. The external realist subscribes to a Credo that is susceptible to model-theoretic treatment. But model theory can then be used to show that every consistent theory (and so every ideal theory) can be made true in many different ways. The immediate challenge for the external realist is to explain what singles out (at most) one of these as capturing *the* correspondence relation. In other words, she must say what makes an interpretation *intended*.

[9] Thanks to Tim Williamson for suggesting this.

3

Attempts to constrain reference

In Chapter 2, we saw that the external realist cannot pin down a single correspondence relation just by producing her favourite theory. On these grounds, one might try to conclude that *nothing* can fix correspondence. But that would be too quick, as several commentators have noted in this connection and, indeed, as an earlier time slice of Putnam once argued against Quine.[1]

In this chapter, I shall sketch four attempts to constrain reference and correspondence. Each constraint presents a version of the worry, mentioned at the end of §1.4, that model theory is too mathematically permissive for metaphysical purposes. I should emphasize that I am merely sketching these constraints. The purpose of this chapter is not to determine what does fix reference and correspondence, but to pave the way for the discussion of Putnam's *just-more-theory manœuvre*.

3.1 The Causal Constraint

There are causal connections between language users and worldly objects. Cats sometimes cause us to exclaim 'cat!' More generally, our causal interaction with cats seems relevant to the question of how the word 'cat' came to refer to cats and not cherries (for example). Such thoughts might lead one to advance:

> **The Causal Constraint.** An intended interpretation must respect all appropriate causal connections between words and objects.

The models generated by model-theoretic arguments are built without any mention of causation. For example, when we advance a permuted interpretation according to which 'cat' refers to cats$^{\mathcal{P}}$, the latter might be a

[1] Putnam (1974: 30–1).

disparate jumble of molecules scattered throughout space and time.[2] Since there is (presumably) no appropriate causal connection between tokens of 'cat' and these cats$^{\mathcal{P}}$, the permuted interpretation would violate the Causal Constraint. The general hope for the Causal Constraint is therefore that it will seriously reduce the number of acceptable interpretations.

It is worth being aware of two intrinsic limitations on the Causal Constraint. First: if there are any entities which have no causal role, then the Causal Constraint will not help the external realist to refer to them. For example, if the external realist believes in an acausal realm of mathematical objects, then we can run a permutation argument which permutes only among them. (Where π is our permutation, we would stipulate that $\pi(x) = x$ iff x is causal, but allow $\pi(x)$ to vary otherwise.)[3] Second: even where the entities in question are causal, there are many well-known problems with any *purely* Causal Constraint on reference. In particular, it is notoriously difficult to specify what counts as an appropriate causally determined connection.[4]

These problems noted, I intend to set them to one side. If it is legitimate to impose the Causal Constraint on interpretation, then it will substantially draw the sting of the model-theoretic arguments.

3.2 The Eliteness Constraint

A related interpretative constraint arises from the idea that model theory delivers lots of *pseudo-properties* and *pseudo-relations* for predicate letters to pick out, but that *genuine* properties and relations are much sparser than model theory might lead us to believe. To spell out this idea, the external realist might invoke David Lewis:

> Sharing of [genuine properties] makes for qualitative similarity, they carve at the joints, they are intrinsic, they are highly specific, the sets of their instances are *ipso facto* not entirely miscellaneous, there are only just enough of them to characterise things completely and without redundancy.[5]

[2] Nothing hangs on my use of 'refers' rather than 'applies'; see §12.5.

[3] Taking a hint from Putnam (1981c: 218, Second Comment).

[4] Putnam (1980a: 101; 1981c: 51; 1984a: 85–90; 1987b: 37–40; 1992a: 50–6) moves from focusing on the interest-relativity of appropriateness, to the interest-relativity of causation itself.

[5] D. Lewis (1986: 60).

The relevance of this is similar to the Causal Constraint: cats$^{\mathcal{P}}$ might be 'entirely miscellaneous', so that *being a cat*$^{\mathcal{P}}$ fails to qualify as a *genuine* property. If the external realist can insist that the interpretation of each predicate letter must be a genuine property or relation, rather than a mere pseudo-property, then she can ignore the permuted interpretation where 'cat' refers to cats$^{\mathcal{P}}$.[6]

External realists who are tempted by this thought need not draw a sharp line between genuine properties and pseudo-properties. Indeed, Lewis himself suggests that we should think of properties as being *more* or *less* elite, where more elite properties and relations carve Nature *closer* to its joints.[7] This suggests:

> **The Eliteness Constraint.** An intended interpretation of each predicate is (typically) an *elite* (or, quite elite) property or relation.

In a slogan: more elite properties and relations are more *referentially magnetic* than their gerrymandered colleagues.

The main problem with the Eliteness Constraint is that it sounds so strongly like an article of faith. Lewis tells us that 'there are only just enough [elite properties] to characterise things completely and without redundancy'. It is unclear what warrants Lewis's confidence here, since surely there could be disjoint sets of properties, both of which are individually sufficient to characterize the world completely, but such that taken together there are redundancies. Of equal importance, it is something of a mystery what determines (and how we are to determine) just how elite any given property is. In short, nothing guarantees that there is one and only one maximally eligible interpretation of our very best theory.[8]

As with the discussion of the Causal Constraint, I raise such worries only in order to set them aside. As things stand, it seems that Putnam's model-theoretic arguments require a thorough assessment of the merits and defects of the idea of the Eliteness Constraint.

3.3 The Fullness Constraint

The next interpretative constraint I have to offer is rather more specific and technical (I elaborate on the technicalities in Appendix I). Suppose that the

[6] Merrill (1980: 80) makes this point, though he was no external realist.
[7] Lewis (1984: 227–8; 1986: 61).
[8] For problems in this area, see Williams (2007; forthcoming).

external realist has advanced a *second-order* theory. In Chapter 2, I used the Completeness Theorem to offer both an indeterminacy and an infallibilism argument. Certainly this result holds for second-order logic (just as it holds for first-order logic).[9] But the (Henkin) models of second-order theories yielded by the Completeness Theorem need not be *full* models. In particular, when the domain is infinite, the second-order variables do not range over all of the uncountably many combinatorial aggregates of objects in the domain, but only over a countable handful of those aggregates. This suggests:[10]

> **The Fullness Constraint.** An intended interpretation of a second-order theory must be a *full* interpretation.

This Constraint would rule out any model-theoretic arguments which invoke the Completeness Theorem.[11] The obvious limitation of this Constraint is that it would not block the permutation argument, since nothing in the proof of the Permutation Theorem depends upon the use of first-order logic.[12]

3.4 The Modality Constraint

The final constraint that I wish to consider is modal. This is the most complicated constraint to explain, since it attacks the use of model theory from a rather different angle.

The models produced by the Permutation and Completeness Theorems assign a truth *value* to every sentence in the external realist's language. However, the external realist may demand more; she may demand an assignment of truth *conditions*. That is, she may insist that we have some grasp, not merely on the truth values that various sentences actually have, but on the truth values our sentences *would* have if things were different.

Putnam claims that this demand can be accommodated using model theory. A sentence's truth conditions comprise an exhaustive list of the

[9] Hence Putnam (1980b: 481; 1994c: 459*n*33; 1999: 16*n*33) is brisk on the distinction between first- and second-order theories.

[10] Shapiro (1991: 203–18) advocates this.

[11] Or any Löwenheim–Skolem theorem; see Appendix I.

[12] As observed by Hale and Wright (1997: 451).

situations in which that sentence would be true.[13] We can think of 'situations' here as *possible worlds*. And just as we have thought of the actual world as a particular model, so we can think of each possible world as a further model. Putnam now suggests that we can raise problems for the external realist by simply applying the model-theoretic arguments to *each* world.[14] For example, to run a permutation argument, we start with some given possible worlds. For each possible world, $\mathcal{W}_\gamma$, we then define a permutation, π_γ, to obtain a permuted world isomorphic with $\mathcal{W}_\gamma$. Each sentence will receive the same truth value in the permuted world as it received in the original world. And so, Putnam claims, the resulting system of permuted possible worlds will preserve the truth conditions of every sentence.

This thought, though, is a little bit too quick. Vann McGee has suggested:[15]

> **The Modality Constraint.** An interpretation must supply truth conditions rather than mere truth values, and must also respect the fact that certain designators are *rigid*. Such designators denote the same object in every world where that object exists, and are empty otherwise.

To illustrate the intended effect of this constraint, suppose the name 'Ajax' is to be a rigid designator. So on the standard interpretation, 'Ajax' denotes Ajax in every world where he exists. Given the way in which permuted interpretations are generated, the Modality Constraint would demand that π_γ(Ajax) is the same object in every world. This rules out many permuted interpretations immediately, since if we simply offer arbitrary permutations for each world, 'Ajax' will name different things in different worlds.

The obvious response would be to offer the same permutation for *every* world. But this will not be possible if some objects (including Ajax himself) only exist contingently, since different worlds will then have different domains. Moreover, the way we permuted the interpretation of 'Ajax' (in a world) was to take Ajax as the argument to a function defined over the

[13] There is a potential ambiguity here. A *truth clause* for a sentence φ is just a statement of the form: φ is true iff ____. As Wallace (1979: 316–18) and Taylor (2006: 56) note, the model-theoretic arguments supply each sentence with a truth *clause*. What concerns us here are truth *conditions*, as I have defined them in the main text.

[14] Putnam (1981c: 25–7, 32–5, 217–18); Hale and Wright (1997: 451–2) make the same suggestion.

[15] McGee (2005: 405–8).