# CONVENTIONALISM AND STRUCTURALISM IN MADHYAMAKA AND SCIENCE

Justin P. Holder

Dissertation submitted for the degree of Doctor of Philosophy

Lady Margaret Hall
The University of Oxford

2023

CONTENTS

ABSTRACT

This dissertation offers a novel reading of the 2nd century Indian Buddhist philosopher Nāgārjuna and explains how, having adopted his philosophical perspective, one may successfully navigate the contemporary scientific realism debate and explain how theories connect with reality. The view I ascribe to Nāgārjuna is that sentences proclaiming the existence or nonexistence of anything are conventions: statements which lack truth-values and should be accepted or rejected based on convenience. I call this ontological conventionalism. I argue that his methodology can be understood using certain Kantian ideas, and that that his form of metaphysics is compatible with even a very strict naturalistic stance. I also show that the French scientist and philosopher Henri Poincaré espoused ontological conventionalism. I argue that Poincaré's phenomenal relationism can serve as an antirealist account of science that is compatible with Nāgārjuna's philosophy. I further argue that an ontological conventionalist can make careful use of epistemic structural realism to explain the efficacy of theories (whether scientific, Buddhist, or otherwise). I do so by developing an original formulation of structural realism which revitalises (and goes beyond) Bertrand Russell's original claim that all we know about even the observable world is its structure – an idea thought to have been defeated by M.H.A. Newman in 1928. It's this form of 'structural realism' that I rely on to explain how an ontological conventionalist can account for the success of science without denying that scientific theories give us knowledge about the world beyond what is observable. I thus present both an antirealist and a realist route through which the success of science can be accounted for in ways consistent with Nāgārjuna's philosophy.

ACKNOWLEDGEMENTS

This project was made possible through the supervision of Jan Westerhoff and Adam Caulton. I thank them both for believing in it and trusting me to take it where I felt it needed to go. Their guidance has been invaluable. I need to thank Adam especially for the unreasonable number of hours he spent walking me through the philosophy of science literature when it was still quite new to me.

One of my greatest challenges was developing my proficiency in Sanskrit. I am indebted to Bjarne Wernicke-Olesen, Diwakar Acharya, and Shree Nahata for teaching me virtually everything I know about the language (which is barely a fraction of everything they know). Thanks also to Lea Cantor for helping me with my French.

Thanks to James Read for his helpful comments on much of the philosophy of science work, and to him and Patrick Duerr for inviting me to speak about Poincaré at the Oxford Philosophy of Physics seminar. And thanks to all those who attended and gave comments. Thanks also to James Ladyman, who took the time to read a few key chapters and gave very helpful feedback that is likely to inform where I go next with my research.

I thank my parents, Joanna and Michael, for their unwavering belief in and support for me. And I thank my incredible wife, Kiri, for her steadfast partnership. I can't imagine where I would be today without her loving support.

I was first taught the philosophy of science by the late Edwin Brandon, who played an irreplaceable role in my introduction to philosophy through my undergraduate degree at The University of the West Indies, Cave Hill. Ed tragically passed away during the year that this dissertation was submitted. I owe him a lot, and I wish he could have read it.

# INTRODUCTION

This dissertation can be described as belonging to the fields of South Asian or Buddhist philosophy and the philosophy of science. That is accurate. To understand the impetus behind it, though, one should understand that I am interested foremost in metaphysics, and this is where that interest led. That is to say, I am writing about the Buddhist philosopher Nāgārjuna only because I think his metaphysical perspective is correct, and I engage with science to the extent that I do only because I think that metaphysics should. The objectives of this project are, accordingly, to explain what Nāgārjuna is right about, and to account for the implications of his views on the interpretation of scientific theories.

I will present and explain my interpretation of Nāgārjuna's philosophy but my goal, here, is not to defend his conclusions – not directly, at least. The main goal is to answer a question. It is an urgent question because, if it lacks a satisfactory answer, then I think that Nāgārjuna's philosophy should be rejected. The question is: how can one explain the success of science in a way consistent with Nāgārjuna's philosophy?

Let us step back and establish some context. Nāgārjuna was a Buddhist monk who lived somewhere in India somewhere in the vicinity of the 2nd century CE. There is not much more we can say for sure about the place(s) and time of his life (see Mabett 1998 and Walser 2005). There is a little more to say about how his work relates to developments in Buddhism at the time (for further reading on this see Westerhoff 2018, Chapter 2). From about the 1st century BCE, texts referred to as the Perfection of Wisdom (*Prajñāpāramitā*) began to appear, which laid some of the early doctrinal basis for Mahāyāna Buddhism, the most commonly practiced 'branch' of Buddhism in the world today. The most well-known and widely celebrated of these texts that the non-expert reader may well have heard of is the Heart Sūtra

(*Prajñāpāramitāhṛdayasūtra* lit. 'The Heart of the Perfection of Wisdom Sūtra'). Nāgārjuna's philosophical works can be interpreted as rational defences of (his interpretation of) key doctrinal innovations that the Perfection of Wisdom literature introduced to Buddhism – chiefly the doctrine of emptiness, which directly repudiated the canonical account of reality's composition created by earlier Buddhist philosophers.

The body of literature expressing the first systematic account of reality in Buddhism is called Abhidharma, and the authors and proponents of this literature are called Ābhidharmikas. The doctrine of emptiness in the Perfection of Wisdom literature seems to involve the negation of every element of the Ābhidharmikas' story about what reality is like, including every essential component of Buddha's teachings. This was a deeply radical move within a religious tradition. How is it that many self-identified Buddhists were drawn to the doctrine that nothing of which the Buddha spoke was ultimately real? I won't be attempting to answer that historical question here, but it is the intriguing context surrounding the difficult task of interpreting Nāgārjuna's masterpiece, *Mūlamadhyamakakārikā* (MMK), in which the Abhidharmikas play the role of opponents in a philosophical dialectic which attempts to demonstrate the emptiness of all things. My own interpretation of what precisely that meant for him will be the topic of Chapter One.

Nāgārjuna was later regarded as the founder of a 'school' of philosophy called Madhyamaka, which means "middle-most" in Sanskrit. That is, Nāgārjuna's view is, by definition, the Madhyamaka view and many other philosophers who have been proponents of this view are said to belong to the 'Madhyamaka school' and are called Mādhyamikas. I will use the term "Madhyamaka" interchangeably with "Nāgārjuna's philosophy" in this project. I tend to favour "Nāgārjuna's philosophy" to make it clear that I am concerned with the philosophical thought of a specific person – Nāgārjuna, the author of *Mūlamadhyamakakārikā* (MMK) – not the collective thought of every person who has been identified as a canonical

Mādhyamika. But “Madhyamaka” is sometimes more convenient for legibility and aesthetic considerations (such as in the title of this work).

I have so far only given a slight indication of what Nāgārjuna is arguing: i.e., in some sense, he does not think things are ultimately real. There is no consensus on what precisely this amounts to. David Ruegg summarized the situation in 1981 by saying that Nāgārjuna’s philosophy “has been variously described as nihilism, monism, irrationalism, misology, agnosticism, scepticism, criticism, dialectic, mysticism, acosmism, absolutism, relativism, nominalism, and linguistic analysis with therapeutic value” (Ruegg 1981, 2). The situation isn’t quite as desperate as this implies; most scholars only seriously entertain a narrower spectrum of interpretations. But the fact that all of these have been suggested (and more since) indicates the difficulty of understanding and categorising Nāgārjuna’s thought despite the fact that its importance is undisputed. My interpretation will be explained in Chapter One, but these notes should be enough to help the reader understand that Madhyamaka will not lend itself to the view that scientific theories are true in any straightforward way. So, we return to the question: if Nāgārjuna was right, how is the success of science to be explained?

This leads us to the scientific realism/antirealism debate. Scientific realism, generally, is the view that at least very successful scientific theories are approximately true. Scientific antirealism is the negation of this. As Hilary Putnam put it, “The positive argument for realism is that it is the only philosophy that doesn't make the success of science a miracle” (Putnam 1979, 73). This idea has come to be known as the ‘no miracles argument’ and I do not think that it is an exaggeration to say that contemporary scientific realism rests on it entirely. Most contemporary scientific antirealists are motivated by what is known as the ‘pessimistic (meta)induction’ (see Laudan 1981), which has it that we can expect, based on the fact that previous successful scientific theories have been shown false, that current ones will be as well. Antirealists argue that successful theories have an incredible ability to accurately represent and predict observations and deny that this fact needs to be explained by appeal to the unobservable systems referred to by scientific theories. It's not obvious that the institutions and methodologies which produce successful theories are more likely to produce ones that are true than they are to produce ones that are merely good at capturing regularities in observations.

We have two questions to answer: (a) what is a Madhyamaka antirealist explanation for the success of science (if there are any)? (b) what is a Madhyamaka realist explanation for the success of science (if there are any)? We need to ask (a) because Nāgārjuna seems to seek to undermine our commitment to the reality of the observable world. If we are not allowed truths about the observable world, then how can we express the regularities that science supposedly tells us about? As for (b), is there any hope that a Mādhyamika can consistently hold that a scientific theory gives us any ultimately true information about objective reality which goes beyond regularities in experience? The question is interesting in itself, but it also puts immense pressure on the Mādhyamika because, if the answer is that there is no way to do this, then the no miracles argument is an argument against Madhyamaka, and this is a difficult philosophical enemy to have. Being incompatible with the NMA would be sufficient for many reasonable people to reject Madhyamaka.

To keep in clear sight just how incredible the success of science is, consider Einstein's prediction in 1915, in accordance with his theory of General Relativity, that if one were to visually locate a star when it appears near to the sun, its cosmological location would appear to be different than normal (to a precisely specified degree) due to the mass of the sun causing a deflection in the path of the light from the star. Stars normally cannot be seen during the day, so this had to be tested during a solar eclipse. The experiment was carried out by Arthur Eddington during an eclipse in 1919 and Einstein's prediction was confirmed. This is incredible, and any philosopher should have the humility to admit that any philosophical view which cannot satisfactorily account for the possibility and actual occurrence of these events should be rejected. Of course, what accounts count as 'satisfactory' will differ widely between philosophers. Some think that an antirealist account is sufficient – others see believing antirealist accounts as akin to believing in miracles. I won't be solving that debate, but I want to show that none of these considerations should lead to the demise of Nāgārjuna's philosophy. I want to show that both scientific antirealists and realists can account for the success of science in ways that are consistent with Madhyamaka.

**Chapter summaries**

In Chapter One, I explain what Nāgārjuna was trying to accomplish with his MMK. I will identify the text's overarching argument in broad terms: what he wants his readers to believe and why he claims they should believe it. I contend that Nāgārjuna is defending what I call ontological conventionalism, which holds that claims about the existence and nonexistence of worldly things cannot be true or false. They are, by necessity, adopted or rejected by free choice, based on convenience. His argument for this is that such claims could only be ultimately true or false if they pertained to things with determinate natures – but nothing has a determinate nature. I distinguish this interpretation of Nāgārjuna from others in the literature.

In Chapter Two, I provide a philosophical account of Nāgārjuna's methodology in MMK. I argue that Kant's notion of fundamental categories of the understanding can be used to explain how Nāgārjuna's arguments can plausibly establish what he says they establish. Specifically, I think that Nāgārjuna shows that judgments involving the use of the categories of causation and inherence always fail to be coherent when applied to things which are taken to have intrinsic nature (*svabhāva*).

In Chapter Three, I investigate whether Nāgārjuna's philosophy can be considered a legitimate intellectual project from a firmly naturalistic stance. I begin with a sympathetic discussion of Ladyman and Ross's (2007) criticism of non-naturalistic metaphysics. Proposing a useful redefinition of the word "metaphysics", I show that Ladyman and Ross's arguments apply only to a subset of its referents and argue that Nāgārjuna's philosophy is a member of a different subset which could be described as 'reflective' in the sense that Bitbol (2010) uses to describe the type of metaphysics recommended by Kant. There are no apparent reasons to object to this style of metaphysics from a naturalistic stance.

In Chapter Four, I present an exegesis of Henri Poincaré's metaphysical position in three key essays within his book, *The Value of Science*. In doing so, I argue for three theses: First, that Poincaré's metaphysical position in these sources is incompatible with his metaphysical position in his earlier book, *Science and Hypothesis*. Second, that the phenomenological relationism defended by Poincaré in these sources is not a form of structural realism but a structuralist form of empiricism, and (by design) has no greater metaphysical commitments than constructive empiricism. Third, that Poincaré holds in these sources that the existence of the external world is merely a convention. An important figure in the history of science, Poincaré serves as an example of how an ontological conventionalist can give a sophisticated antirealist account of the success of science.

Chapter Five will begin the move towards a realist account of the success of science. Bertrand Russell defended the notion that our knowledge of objective reality is limited to its structure as a response to antirealist concerns about objects of perception. Newman's (1928) objection was meant to demonstrate that any claims about the world's structure alone were trivially true, given that the world contains a sufficient number of things. Though there have since been rejoinders to Newman's objection in the context of contemporary scientific structural realism, these all make at least one concession: that our knowledge of objective reality cannot be *purely* structural as Russell originally claimed – it must extend also to observable facts. I argue that this consensus is mistaken. One can be a successful structural realist even as a brain in a vat. Structural realism is therefore not only a middle-ground answer to scientific antirealism: it is an effective middle-ground answer to even the most radical forms of scepticism.

In Chapter Six, I address a metaphysical view which Nāgārjuna would presumably want to reject but could not have anticipated – ontic structural realism. I argue that the positive metaphysical content of ontic structural realism – the claim that structure is ontologically fundamental – is indefensible from a naturalistic perspective. My strategy is to show that the evidence and arguments advanced by ontic structural realists to motivate their positive thesis underdetermine the choice between it and another, contradictory thesis. I argue that there is no apparent way to break this underdetermination without adopting an anti-naturalistic approach to the human mind. The negative metaphysical content of ontic structural realism, however – the elimination of fundamental individuals with intrinsic nature – is untouched by this criticism and may be defended independently.

In Chapter Seven, the conclusion, I argue that the structural realist proposal I developed in Chapter Five can be adapted to work for a Mādhyamika. That is to say, it can allow a Mādhyamika to make ultimately true claims about objective reality such that the success of science is explicable in a way that can satisfy the realist, without running afoul of Nāgārjuna's philosophy.

CHAPTER ONE

## Nāgārjuna as an Ontological Conventionalist

I contend that Nāgārjuna, in his *Mūlamadhyamakakārikā*, is defending what I call ontological conventionalism, which holds that claims about the existence and nonexistence of worldly things cannot be true or false. Such claims, according to him, are to be adopted or rejected by free choice, based on convenience. This makes him the earliest example of a philosophical conventionalist in the sense associated with Poincaré and Carnap – a new perspective through which Nāgārjuna can be situated in the history of philosophy. I outline Nāgārjuna's motivation and argument for his conventionalism, and I distinguish this interpretation of his work from others in the literature.

### 1.1 Introduction

It is standard practice to translate "*saṃvṛtisatya*", which represents one of the 'two truths' in the Buddhist tradition, as "conventional truth". I have, however, found no literature investigating what connection, if any, exists between this concept and the notion of a 'convention' as used by philosophers in Europe and North America since the $19^{th}$ century. In this article, I show that the influential $2^{nd}$ century Indian Buddhist philosopher Nāgārjuna can be read as a conventionalist in precisely this latter sense. Specifically, I will argue that Nāgārjuna holds that all sentences which declare the existence or non-existence of any worldly thing are conventions in the same sense that, e.g., Poincaré held that axioms of geometry are conventions: these sentences lack truth-values and are to be accepted or rejected by free choice, guided only by convenience. For Nāgārjuna, *saṃvṛtisatya* is language full of the merely conventional existence and non-existence of worldly things, which can never describe the true nature of reality. In what follows, I will outline Nāgārjuna's argument and motivation for this view, go into more depth about the nature of conventions, and then explain how this interpretation of Nāgārjuna should be distinguished from others in the literature.

### 1.2 Nāgārjuna's philosophical project

Whatever Nāgārjuna's philosophical point is, he encapsulates it by saying that all things are 'empty' (*śunya*). It is often noted that Nāgārjuna does not rely on any single 'master argument' for universal emptiness (Westerhoff 2009, 92; Siderits and Katsura 2013, 7). While this is undeniable to a certain extent – the various chapters of his masterpiece, *Mūlamadhyamakakārikā* (MMK), are not presented as steps in a single argument – it is easy to overstate. MMK is a systematic work which seeks to convince its reader of *something* by some means (despite Nāgārjuna's occasional boast of positionlessness). As such, it must be possible to summarise both that something and those means. This is precisely what I mean by MMK's (one) overarching argument.

As I read it, MMK's argument can be summarized like this:

**P1**: Unacceptable consequences follow from any sentence analysed in terms of intrinsic nature.

**P2**: Only a sentence analysed in terms of intrinsic nature could establish the existence or nonexistence of a worldly thing.

**P3**: A sentence with unacceptable consequences cannot establish the existence or nonexistence of a worldly thing.

**C**: The existence or nonexistence of any worldly thing cannot be established.

The qualification "worldly" is meant to focus Nāgārjuna's argument where it was quite clearly intended: real objects in the world, which would today be considered within the domain of physics or metaphysics. His arguments don't prevent us from saying, truthfully, that some fictional object exists within some fictional setting, nor establishing as true some sentence which quantifies over purely formal objects within some well-defined formal system. I also exclude experiences from the category of "worldly thing" only and precisely to the extent that an experience's existence is entailed by its being actually had. Nāgārjuna is, of course, not trying to convince us that we do not have experiences; arguing that his view entails this is tantamount to attempting a refutation. I will discuss this in greater detail below. Throughout the discussion, I will not use the qualifier "worldly" every time it is intended. It should be assumed to be present when it obviously should be and absent when it obviously should be. In cases where that is not obvious, it will be my responsibility to specify what I mean.

We speak of 'sentences' here rather than 'propositions' to preserve the notion that they can be analysed in different ways to produce different meanings. Some clarificatory remarks are in order: what does it mean to analyse something in terms of intrinsic nature, and what is the connection between intrinsic nature and existence?

### 1.2.1Analysis in terms of intrinsic nature

Analysing a sentence in terms of intrinsic nature means interpreting any worldly thing referred to in the sentence as possessed of its own intrinsic nature (*svabhāva*) or as ontologically dependent on things which are. This will usually include every noun in the sentence but might also include other words which might be substantivized, such as how the verb "goes" (*gacchati*) has the same reference as "the act of going" (*gamana*) in MMK 2.

But what is it to think that these things have intrinsic nature? We know for certain that the concept of intrinsic nature is of paramount importance to Nāgārjuna based on his reliance on it in the first two verses of his climactic argument in MMK 24.16-19:

> If you look upon the existence of things as coming from intrinsic nature, you thus see things as being without cause and conditions. (24.16)
>
> Effect and cause, and indeed agent, instrument, and act, arising and ceasing, and fruit, you deny. (24.17)[1]

Here, Nāgārjuna identifies his opponents' philosophical flaw as their view that things in the world exist through intrinsic nature. This is something Nāgārjuna hopes to have shown in the previous chapters of the text. Nāgārjuna does not take the time to explain precisely what his interpretation of intrinsic nature is, likely because it was a familiar concept to his opponents, the Ābhidharmikas, who employed it in their own philosophy.

The Ābhidharmikas were reductionists who argued that composite things could not be said to truly exist because they had no existence independent of the existence of their constituent parts. Based on this reasoning, the only things which could be said to ultimately exist were things whose natures could not be reduced to the natures of their parts – i.e., things with their own intrinsic nature. These partless simples were called dharmas, and they were identified as the only ultimately real components of reality. In accordance with the Buddha's teachings on impermanence and dependent arising, any given dharma was thought to exist only momentarily and in accordance with its specific conditions.

[1] The translation is my own, and the Sanskrit is from Siderits and Katsura (2013). I will use Siderits and Katsura's translation where I cannot see how it could be improved.

To have intrinsic nature means that a dharma is what it is independent of anything else. Its identity is not borrowed, derived, or constructed from elsewhere. Nāgārjuna completes his famous argument as follows:

> Dependent origination we declare to be emptiness. It [emptiness] is a dependent concept; just that is the middle path. (24.18)
>
> There being no dharma whatsoever that is not dependently originated, it follows that there is also no dharma whatsoever that is nonempty. (24.19)

To analyse a sentence in terms of intrinsic nature, then, means to interpret the worldly things in the sentences as having, or being ontologically dependent on things which have this non-derived identity. Nāgārjuna's chief problem with this is that, according to him, this interpretation renders those worldly things incompatible with dependent arising. However, we don't need to be adherents to this Buddhist doctrine to find Nāgārjuna's arguments relevant. As we can see from 24.17 above, Nāgārjuna considers such fundamental concepts as cause and effect and arising and ceasing to be encompassed by the notion of dependent arising. His claim is that the very notion of a conditioned event or entity crumbles if we interpret the world through the lens of intrinsic nature. This is the largest factor in why sentences analysed in terms of intrinsic nature result in absurdities, as claimed in P1. Since the concept of intrinsic nature does not get along well with concepts as fundamental as cause and effect and so on, sentences analysed in terms of intrinsic nature reliably fail to be coherent under rational scrutiny. We will develop this point thoroughly in the next chapter. Now that I have outlined the meaning of "analysis in terms of intrinsic nature", I will explain the connection Nāgārjuna sees between this and existence.

### 1.2.2 Existence and metaphysical determinacy

It is tempting to think that establishing that all worldly things lack intrinsic nature is the entirety of Nāgārjuna's point, but it is not. That is the thesis which is most difficult for Nāgārjuna to establish and the one to which most of his argumentation is devoted, but it is not his final thesis – not the ultimate point that he wants his readers to absorb and accept. His ultimate point is that the very notion of existence does not pertain to how reality truly is. He wants his readers to accept that it is never ultimately true or false that anything exists or does not exist. How might he get there by arguing against intrinsic nature? The bridge, I would argue, is metaphysical determinacy. As Jay Garfield puts it, "To say that something exists . . . is to presuppose that it has an identity. If we cannot say what it is that exists, an existence claim is empty" (2015, 36). In other words, in order for there to be a fact of the matter of whether some *x* exists, there must be a fact of the matter of what parts of the world are *x* and what parts are not-*x*. That, at least, is what Nāgārjuna needs us to believe. This ought to be argued because some authors have argued that reality itself is in some respects indeterminate, including some Chinese Madhyamaka thinkers (Ho 2019). But in MMK, Nāgārjuna seems to be saying that things *would have to be* determinate in order to exist:

> Further, without intrinsic nature and extrinsic nature how can there be an existent?
>
> For an existent is established given the existence of either intrinsic nature or extrinsic nature. (15.4)

He goes on to argue that a thing would need to be determinate in order to truthfully say that it does *not* exist:

> If the existent is unestablished, then the nonexistent too is not established.

> For people proclaim the nonexistent to be the alteration of the existent.
>
> (15.5)

In these verses we can see the connection between intrinsic nature, metaphysical determinacy, and existence. It is only through intrinsic or extrinsic nature that there can be a determinate metaphysical fact of the matter of whether a designated thing is *x* or not-*x*. It is therefore only through intrinsic or extrinsic nature that there can be a fact of the matter of whether *x* exists. This does raise the question, though: if either intrinsic or "extrinsic" nature can work to achieve metaphysical determinacy, why have we been so focused on intrinsic nature exclusively? The answer is that Nāgārjuna thinks extrinsic nature cannot exist without intrinsic nature.

We have explained that a thing's intrinsic nature (*svabhāva*) is what it is independent of everything else. A thing's extrinsic nature (*parabhāva*) is what it is dependent on other things. If, according to the Ābhidharmikas, a composite thing is made of three different dharmas, then the extrinsic nature of that thing would be the intrinsic natures of those three dharmas, collectively. I would argue that this, at least, seems to be the sort of scenario Nāgārjuna has in mind when he composes the following verse:

> Given the nonexistence of intrinsic nature, how will there be extrinsic nature?
>
> For extrinsic nature is said to be the intrinsic nature of another existent.
>
> (15.3)

Nāgārjuna thinks that vanquishing intrinsic nature is sufficient for vanquishing extrinsic nature because he understands extrinsic nature to *be* intrinsic nature in the final reckoning. Perhaps the reasoning is something like this: if an entity has a nature which is not its own, then that nature must belong to one or more other things, since it could not be that a nature comes from nowhere in order to characterise something extrinsically. And if it does not come from nowhere, then it is the intrinsic nature of that from which it comes. This implies that something which is identified only through extrinsic nature has only a borrowed, derived, or abstract identity – nothing that can be taken too seriously as a substantial part of reality. That this is his perspective is confirmed in MMK 22.3, where he rejects the possibility that something with only an extrinsic nature could have become Buddha.

It's clear that Nāgārjuna took this to be something his opponents already accepted and likely therefore saw no need to argue for it. Today, this is hardly adequate support for such an important part of Nāgārjuna's case, but that is not an issue I have space to tackle in this work. For now, we must just note that Nāgārjuna does not focus on intrinsic nature because he is particularly interested in its being *intrinsic*. He is interested in it because he – or, perhaps more importantly, his historical opponents – thinks that it is the only nature that an unequivocally existent thing could have. For Nāgārjuna, this is all about existence. He is trying to get his opponents to abandon the conceit that their theories of what exists and what does not exist actually apply to reality. His point is not restricted to intrinsicality or to a certain kind of existence. If, in Nāgārjuna's philosophical context, there were prominent proponents of the substantial existence of things with extrinsic nature in the total absence of intrinsic nature, we should expect that Nāgārjuna would have spent more time arguing against that, too.[2]

This has been a general introduction to Nāgārjuna's argument strategy in MMK. But, based only on what I have said so far, there are many ways one might interpret his philosophical point. Is it nihilism? Some form of idealism? Is Nāgārjuna opposed to metaphysics? All of these have been suggested, but I will argue that Nāgārjuna is an ontological conventionalist. His position is that the only way to legitimately accept or reject the existence of something is the convenience of doing so – there is no fact of this matter determined by reality.

[2] As evidence, consider MMK 5.8 and 15.10, which will be quoted and discussed later in this chapter.

## 1.3 Nāgārjuna the conventionalist

In step with Buddhist tradition, Nāgārjuna says that there are two truths: the 'ultimate truth' (*paramārthasatya*) and what is literally translated as the 'obscuring truth' (*saṃvṛtisatya*), also referred to as the 'worldly, obscuring truth' (*lokasaṃvṛtisatya*). The latter is usually translated as the 'conventional truth' in contemporary English-language scholarship. There is much discussion of what the properties of the conventional truth are (see The Cowherds, 2011), but these discussions do not compare Nāgārjuna's view with the existing tradition of philosophical conventionalism in European philosophy. This is not intended as a criticism of the existing literature; that the word 'conventional' is often chosen to gloss *saṃvṛti* in this context does not mean that there is necessarily a significant philosophical connection between what Nāgārjuna had in mind and what other philosophers writing about 'conventions' had in mind. But I will be arguing that this is indeed a very fruitful comparison to pursue. First, I will lay out what it means to interpret Nāgārjuna as a conventionalist. Then I will touch on what, in my opinion, makes this reading more compelling than its rivals.

### 1.3.1 What is a convention?

Conventionalism has a significant, if arguably understudied, history in European and American philosophy. Its earliest adoption there seems to be in the reflections of 19th century French scientists, the most influential of whom was Henri Poincaré.[3] Yemima Ben-Menahem (2006) has, at this time and to my knowledge, the only book length study of conventionalism as such in this tradition, starting from Poincaré, who famously argued that axioms of geometry were conventions, and discussing later figures such as Carnap, Quine, and Wittgenstein. I will not dwell on the details of these other philosophers' conventionalism, but I will give a definition

[3] See Schmaus (2020) for more context on the origins of this tradition of conventionalism in France.

of "convention" which firmly matches Poincaré's use of the term. There are discussions to be had about how the use of the term, and the ideas with which it is associated, evolved over time, but the reader should consult Ben-Menahem's book for these. My goal is to explain precisely the sense in which Nāgārjuna was a conventionalist so that he might be included in future discussions about conventionalism, and so that conventionalism may play a larger role in future discussions about Nāgārjuna.

We can define a 'convention' as a proposition which we legitimately accept or reject by free choice, based on convenience, not its truth-value. Conventions are common – we all use them frequently. A good example of one is any definition, e.g., '*this* specific length is a metre'. Neither of these propositions is true or false independent of people's choice to accept it. No one *discovered* how long a metre is by investigating the world or through logic, and it cannot be found that we are mistaken on this issue. As this example illustrates, a genuine convention is not a proposition whose truth-value is ignored for some reason. Rather, conventions do not have truth-values. Convenience and consensus exhaust the considerations which determine their use.

When philosophers argue about conventions, it is to show that some propositions which others believe to be non-conventions are in fact conventions, or vice-versa. We can observe from the above discussion that the notion of truth by convention is anathema to these thinkers. Their philosophical points rely on a strict incompatibility between truth and convention. As Ben-Menahem stresses: "To the conventionalist, the very idea of truth by convention is as incongruous as that as meaningful nonsense" (2006, 1).

Now that we have seen what a convention is, let us take a close look at what it means to interpret Nāgārjuna as a conventionalist.

### 1.3.2 A conventionalist interpretation of obscurational truth

To read Nāgārjuna as a conventionalist, it is not so simple as claiming that the ultimate truth encompasses propositions with truth values while the conventional/obscurational truth is made up only of conventions. To avoid confusion on this point, I will favour the literal "obscurational truth" translation of *saṃvṛtisatya* over the more popular "conventional truth". What makes obscurational truth worth differentiating from ultimate truth is not the presence or even the number of conventions used in it, but rather *which* propositions are regarded as conventions. In the obscurational truth, all existential sentences – i.e., sentences asserting the existence and nonexistence of worldly things – are conventions. Because of this, the obscurational truth is obscurational in the sense that, when someone makes claim from this semantic stance, they have no commitment to the ontology they reference therein. Thus, by its nature, this language obscures, rather than describes, what reality is really like.

We must say more about what existential sentences are. First, we take it as axiomatic that "existence", along with its various translations in other languages, taken literally, has a single possible meaning.[4] It is among the most fundamental concepts we have, so it can hardly be explained: to exist is to be (and to be is to exist, for a thing); existence is the state of existing. An existential sentence is one which claims that something exists or does not exist. We can afford to keep this simple and stipulate that existential sentences are specifically those which contain the statement 'θ exists' or 'θ does not exist' (or their translations) where θ represents any set of worldly things. I say that we can afford to keep this simple because the implications of treating existential sentences so defined as conventions naturally extend everywhere Nāgārjuna might need them to. For example, we don't need to argue that any or every instance

[4] This isn't strictly necessary, but to avoid it we would need to define some technical term, say, existence$^{T}$ which would have one meaning by stipulation, and use this to define existential sentences. This seems unnecessary because I think that most people would agree that "existence" normally only has one literal meaning.

of the sentence 'the house is on the hill' is a convention once we have established that 'the house exists' and 'the hill exists' are both conventions. Having established the existential sentences as conventions, we still have available to us multiple viable interpretations of 'the house is on the hill': we might read it as a convention, too, or we might read it as a phenomenological report or description of a fictional world, in which cases it would have a truth value. The unavailable reading is the one which sees the sentence as implying the literal existence of the house or the hill and as having a truth value. The principle at work is that Nāgārjuna wants us to interpret any piece of language as consistent with the essentially conventional nature of all existential sentences.

The fact (according to Nāgārjuna) that the existence and nonexistence of things cannot be established does not mean that we are supposed to avoid speaking of existent and nonexistent things. On the contrary, this fact should free us to speak about such things at our convenience, unfettered by metaphysical concerns. My phrasing of C above is intended to make clear that Nāgārjuna's conclusion isn't itself an attempt to describe what reality is like. He is rather trying to demonstrate the fatal problems with other philosophers' earnest efforts to do so. The defining characteristic of obscurational truth (for virtually any Buddhist) is that it is unconcerned with whether it represents what reality it truly like. On the other hand, if something is said to be ultimately true or existent (the Sanskrit *paramārthasat* can mean either), then this is explicitly meant to be an accurate description of reality as it truly is. According to Nāgārjuna distinctively, "the ultimate" (here he does not include "truth" or "existence") can only be 'attained' (*āgamya*) through the obscurational truth (MMK 24.10). We will discuss how this might be possible in the final chapter. For now, the important observation is that he does not dismiss the idea of ultimate reality as a mistake only entertained by fools, as he does the idea of ultimately existent and nonexistent things:

> But those of little intellect who take there to be existence and nonexistence with respect to things,
> they do not see the auspicious cessation of what is to be seen. (5.8)

> "It exists" is an eternalist view; "It does not exist" is an annihilationist idea.
> Therefore the wise one should not have recourse to either existence or nonexistence. (15.10)

For Nāgārjuna, even though existence and nonexistence do not pertain to reality, there is a reality that they fail to pertain to. This reality, of course, is not something itself with a nature that can be described by a metaphysical theory and said to exist – referring to it is simply acknowledging that there is a 'way things are'. Not only does Nāgārjuna not deny this, but he thinks that reality as such is something which we should be trying to know better. However, the only way to do so is by relying on language and concepts which will never represent it.

As I mentioned, the obscuring truth does contain non-conventions – i.e., sentences with truth values. It must, or else it wouldn't be useful. It can't be that we can legitimately choose to accept or reject any proposition we feel like all the time. This is, at very least, because we actually do have experiences. Nāgārjuna does not dispute that (how could he?) so, completely independent of questions of existence, there are facts of the matter of what we experience and what we don't. Language is definitely capable of reporting some of these facts – we wouldn't have much success communicating with each other otherwise. So, we know that, at the very least, reports of experiences can be true or false. What Nāgārjuna's philosophy throws into question is the ontology which those reports may reference. For example, it may be true or false that someone has the experience which could be reported as "I see a ball" even if "there exists at least one ball" taken literally does not have a truth-value. Talk about balls in this context is to be understood as a convenient way to structure and convey experiential content rather than a commitment to the idea that the existence of balls can be established.

Here, it could be objected that we have allowed for the existence of experiences, which means that the existence of at least one sort of thing (i.e., experiences) can be established. But this is not quite right. Nāgārjuna's analysis, if it works at all, can certainly show that it does not make sense to see experiences as discreet objects possessed of their own nature. Because of that, there is no way to rigorously determine that *this* experience exists now and *that* experience existed then. Nāgārjuna does not (cannot, wouldn't want to) show that we don't have experiences, but he can show that it is folly to attempt to construct an ontology of experiences through metaphysics. When we talk of "having an experience" we are referring loosely to the fact that experience is going on – we do not need to hypostatise over that fact and marry ourselves to some metaphysical picture of what it means for the nature of reality. It is this latter move that Nāgārjuna's analysis is meant to help us resist. Consider the following verse from MMK, which Candrakīrti holds up as the definitive meaning of "emptiness".[5]

> Liberation is attained through the destruction of actions and defilements; actions and defilements arise because of falsifying conceptualizations;
>
> those arise from hypostatization; but hypostatization is extinguished in emptiness. (MMK 18.5)

The goal is to end hypostatization, which, in this context, would mean reifying experiences as worldly things. Any attempt to move towards regarding phenomena as the foundation of reality, or some other kind of metaphysical idealism, will necessarily cross this line. Thus, Nāgārjuna does not regard experience or anything else regarded as mental to be metaphysically privileged in any way; his point is that we should stop imposing such speculative notions onto the way

[5] In his commentary on MMK 24.7 in the *Prasannapadā*.

things are. We don't need to do that to acknowledge that we have experiences. So, while it may sound dubious to argue that there is something wrong with saying that experiences exist, it is not too hard to see that it would be no simple task to make rigorous sense of the existence in the world of some entity called "an experience". It is precisely that kind of project that Nāgārjuna is rejecting. He regards such a goal as unwise in itself for the reasons we have seen: no determinate nature for such a thing will withstand scrutiny, so there will be no rigorous basis on which to affirm "it exists" or the opposite.

The most pressing concern with this complete rejection of worldly ontology is how we could possibly explain the intersubjective facticity of experience if it is true. If balls neither exist nor not-exist, how come many people can agree on the properties of the 'same' ball over time and space? How could science have been as incredibly successful as it has been if existence pertains neither to its theoretical objects nor its empirical subjects and instruments? We won't solve this in this chapter. My solution to this issue is one of the main contributions I intend this dissertation as a whole to make to the study of Nāgārjuna, and as such will only be fully articulated in the final chapter, where it can draw support from the chapters before it. For now, we should just note that the obscuring truth is not made entirely of conventions – it is just that every existential sentence within it is a convention.

One critically important property of conventions which will come up frequently in my discussion of Nāgārjuna is that the negation of a convention is necessarily also a convention. We could not have the choice to accept or reject P if we did not have the choice to accept or reject not-P. Because of this, interpreting Nāgārjuna as a conventionalist makes clear and easy sense of his otherwise somewhat mystifying rejection of both existence and nonexistence. It makes no sense to say that '*x* exists' is merely a convention while '*x* does not exist' is true. If we can legitimately choose whether to accept one, then, necessarily, we can legitimately choose whether to accept the other. So, when Nāgārjuna says that a wise person commits neither to the truth of ‘x exists’ nor its negation, we should not see this as an esoteric spiritual claim, nor should we move to dress it in qualifications. This is simply a straightforward consequence of taking existential sentences to be conventions.

In the next section, I will consider the role of conventions in the influential “semantic interpretation” of Nāgārjuna’s position. That will allow us to differentiate clearly between it and the ontological conventionalist reading I am suggesting here.

### 1.4 Conventions and the semantic interpretation

The semantic interpretation is an understanding of Nāgārjuna’s conclusion originating (at least in these terms) with Mark Siderits. The defining idea here is that Nāgārjuna’s project is a perspective on truth rather than on metaphysics, the summary of his conclusion being, “the ultimate truth is that there is no ultimate truth”. Siderits (2003) contrasts this with metaphysical interpretations of Nāgārjuna, which are identified as those which take him as teaching nihilism and those which take him as teaching that the true nature of reality transcends conceptualisation. I admit to some uncertainty on where my own ontological conventionalist reading might fall within this framing. I do think that, from a Nāgārjunian perspective, the nature of reality transcends conceptualization, and that this can be defended with textual evidence.[6] But I don’t think that that is Nāgārjuna’s point, *per se*. It is a corollary, not the main thesis. Nāgārjuna’s goal, on my interpretation, is to expose philosophers’ attempts at identifying the ultimate constituents of reality as futile. He is trying to correct unwise attitudes and behaviour, not teach us something about the world. I am therefore unsure whether my interpretation counts as a metaphysical one by Siderits’ definition, and it is certainly not encapsulated by “the ultimate truth is that there is no ultimate truth” or any other overtly

[6] See MMK 18.9, for example.

semantic slogan. I consider Nāgārjuna to be doing what I call 'reflective meta-physics' in Chapter Three. For now, let us see what the main differences are between my interpretation and the two major semantic interpretations in the literature, with special attention paid to the role of conventions. A thorough discussion will not be possible here, but I will briefly characterise these interpretations, contrast them with my own, and suggest some advantages mine has over them.

First, we have Siderits' own version of the semantic interpretation. Siderits claims that Nāgārjuna's doctrine is "equivalent to the rejection of what Putnam (1981, 49) calls metaphysical realism. This is the view that there is one true theory about the nature of reality; with truth understood as correspondence, and reality understood as consisting of a fixed number of objects with natures that are independent of the concepts we happen to employ" (2003, 12). Obscurational truth,[7] for him, is any discourse which does not involve metaphysical realism in this specific sense. That means it does include discourse which embraces the existence of objects dependent on our concepts and interests. There is no significant philosophical conventionalism here. The rejection of metaphysical realism is not supposed to leave us using truth-value-less sentences. Rather, we are encouraged to adopt a different metaphysics – one based on concept- and motivation-dependence – which does result in truth-valued existential sentences.

Why might we prefer ontological conventionalism to Siderits' non-conventionalist semantic interpretation? One concern is that Siderits' interpretation seems a bit tame. He identifies Nāgārjuna's philosophical project with Hilary Putnam's rejection of metaphysical realism, but this is somewhat awkward since Putnam still considered himself a realist, with an

[7] Siderits, like every other author I will discuss, favours "conventional truth" as a translation of *saṃvṛtisatya*.

unhindered ability to learn about what the world is really like. For him,[8] conceptual frameworks are the mandatory avenue through which the human mind interacts with objective reality. He regards metaphysical realism as incoherent because the only way to represent reality is through the use of conceptual schemes; he argues that there is no "God's eye view" from which one can represent things exactly as they are intrinsically. But Putnam doesn't see this as an obscuration of reality; for him, in thinking about reality using a conceptual framework, one is really thinking about what reality is *really* like. As a simple example, Putnam (1987) raises the question of whether mereological composites should be counted as objects. In a universe containing only A and B are there only two objects {A, B} or are there three {A, B, (A+B)}? Putnam regards it as obvious that this is entirely dependent on how we choose to define 'object' and, like any definition, that is just a convention. There is no 'correct' definition of object and therefore no independent fact of the matter of the number of objects in this universe. This is Putnam's intention behind the incoherence of metaphysical realism, but we can see that this only minimally disrupts the idea that we can understand the nature of reality. For Putnam, both the theory that there are two objects and the theory that there are three objects are true in this universe relative to each theory's respective definition of an object. The nature of reality is accessible in Putnam's view – it is just such that accessing it involves making a conventional choice between theoretical axioms (at least). If this is what Nāgārjuna's philosophy amounts to, its reputation for radicalness and supposed soteriological potential seem unearned.

Of course, Siderits or others might pursue a different account of the implications of rejecting metaphysical realism. Many philosophers did, after all, take serious issue with Putnam's claim to being a realist (see, e.g., Anderson 1992). But then we have returned to metaphysics, with Nāgārjuna's view being a specific metaphysical account of what follows

[8] At least at the time when he wrote the pieces in which he defends this view. Later in his life, Putnam gravitated towards direct realism.

from the rejection of metaphysical realism; this requires argument which may be difficult to ground in primary sources.

Another weakness of Siderits' semantic interpretation is the extent to which it must qualify Nāgārjuna's originally unqualified claims about existence and nonexistence. Nāgārjuna says that existence and nonexistence do not pertain to things – he doesn't exclude things dependent on concepts and interests, so it is difficult to see why existential sentences pertaining to those things could be literally true or false, which should be possible on this semantic interpretation since their existence is consistent with the rejection of metaphysical realism.

My ontological conventionalist interpretation does not have any of these difficulties. Since, according to this reading, *all* existential claims are undermined, there is no metaphysical qualification available that can compromise the impact of Nāgārjuna's bold, categorical claims.

The other major form of the semantic interpretation is the dialetheist one associated with Yasuo Deguchi, Jay Garfield, and Graham Priest (Deguchi et al. 2008). Dialetheism is the view that there are true contradictions. Accordingly, this interpretation embraces the apparently paradoxical nature of the slogan 'the ultimate truth is that there is no ultimate truth'. Here, each instance of "the ultimate truth" has the same meaning and is taken to have the same extension as "the literal truth". On this reading, every proposition other than the one expressed by 'there is no ultimate truth' is taken to lack a truth value (and be, thus, a convention), including Nāgārjuna's teachings about things lacking intrinsic nature, and so on. Here is Garfield:

> Nāgārjuna makes it plain that even the statement that phenomena are empty of intrinsic nature is itself a merely conventional truth, which, in virtue of the necessary involvement of language with conceptualization, cannot capture the nonconceptualizable nature of reality. *Nothing can be literally true*, including this statement. (Garfield 2015, 68, emphasis added)

This is near-universal conventionalism. That is certainly not too tame a thesis for Nāgārjuna, but is it the best way of understanding him, and can it be defended anyway?

There is no consensus on the plausibility of dialetheism or its propriety as a tool for interpreting Nāgārjuna.[9] But in keeping with the theme of this chapter, I will focus my brief critical comments on the plausibility of (near) universal conventionalism. I have already explained why this is untenable. At the very least, it is implausible to claim that someone's report of an experience they claim to have had cannot be true or false because of metaphysical considerations to do with intrinsic nature, causation, and so on. We can ascribe purely phenomenological – not metaphysical – meanings to such reports, in which case there is no reasonable basis on which to deny them truth values.

Here, a defender of the dialetheist semantic interpretation may grant that Garfield went too far in saying that *nothing* can be literally true. They might offer the correction: "no *metaphysical* statement can be literally true", with the semantic interpretation's slogan, "the ultimate truth is that there is no ultimate truth" being clarified to mean, 'the metaphysical truth is that there is no metaphysical truth'. But this is problematic for the dialetheist. For, if we read "the ultimate truth" in their slogan as having a different extension than "the literal truth", it seems neither advantageous nor necessary to represent Nāgārjuna's view as paradoxical. Why not go with the non-paradoxical summary, 'there are no metaphysical truths'? This can be literally true without any contradictions arising. I see no motivation to insist that 'there are no metaphysical truths' must itself be a metaphysical truth, especially since Nāgārjuna was at pains to insist that he did not have any thesis in the same sense that his opponents did. This issue will arise for the dialetheist even if they use a qualification other than 'metaphysical', once it results in an extension different to 'literal'. But they cannot use any qualification with the same

[9] See Siderits 2013 for some critical remarks.

extension as 'literal' because phenomenological reports, at least, do have literal truth values. There doesn't seem to be a winning formula for the dialetheist when we acknowledge this.

This begs the question, what about the semantic interpretation I just came up with: 'there are no metaphysical truths'? What, if anything, makes this rendering inferior to my "the existence or nonexistence of anything cannot be established"? It comes down to precision. As we have seen in some of the verses quoted above, Nāgārjuna explicitly targets existence and nonexistence as the problematic views to be abandoned. This is quite enough to substantiate the radical and transformational nature of his thought, and by focusing on existence specifically rather than "metaphysics" – a word with no Sanskrit correlate – as Nāgārjuna's target, we can avoid getting bogged down in any debates over which statements count as metaphysical and which do not. It is also not important whether we consider what Nāgārjuna himself is doing "metaphysics".

In conclusion, 'the semantic interpretation' of Nāgārjuna in recent literature resolves into two different interpretations, of which one is not conventionalism at all and the other involves conventionalism that is too extreme. By arguing only that existential sentences are conventions, my ontological conventionalist interpretation of Nāgārjuna finds a happy middle ground.

### 1.5 Two kinds of existence?

We saw in the discussion of Siderits' semantic interpretation that Nāgārjuna's point became a rejection not of all truth claims but only of truth interpreted according to a philosophical perspective named metaphysical realism. This strategy of associating Nāgārjuna's object of negation strictly with a certain metaphysical theory is not unique to the semantic interpretation; it is a prominent way to qualify his claims in order to avoid reading him as a nihilist. When the focus is on metaphysics rather than semantics, this qualification usually manifests as a rejection specifically of the existence of anything with intrinsic nature. Sometimes this is economically referred to as ‘intrinsic existence’. On this view, we are allowed to assert the ‘mere’ existence of things in accordance with responsible epistemic practices. Responsible epistemic practices do not reveal anything with ‘intrinsic existence’ – ‘intrinsic existence’ is seen as a deep-rooted, persistent error. This is obviously a different way of articulating Nāgārjuna’s ideas than my conventionalist reading, in which there is only one form of existence. However, it is not obvious how stark this difference is; is it a substantial philosophical disagreement or a mere semantic difference? Could it be, for example, that ‘mere existence’ just refers to existence as a convention? To understand this ‘two kinds of existence’ reading so that it might be contrasted with my conventionalist one, we will focus on the explanation offered by Tsongkhapa, an incredibly influential Tibetan monk who lived at the turn of the 15th century. His views on this matter are in turn interpretations of Candrakīrti’s – one of Nāgārjuna’s Indian commentators.

Tsongkhapa emphasises the importance of clearly differentiating between what is to be negated from what is not to be negated in Madhyamaka. The danger of failing to observe this distinction, he explains, is that “then existing things would exist in their own right, and things that do not exist in their own right would then become nonexistent” (2021, 273). We see here that, according to Tsongkhapa, existing things as such are not to be negated. In fact, one of the desirable philosophical outcomes of the Mādhyamika’s negation is to prevent certain existent things from being regarded as nonexistent. It’s important to note that Tsongkhapa also targets existence *and* nonexistence as targets of negation. But, unlike in my ontological conventionalist interpretation wherein existence and nonexistence simpliciter are denied to reality except as conventions, Tsongkhapa sees Nāgārjuna as specifically negating the ‘intrinsic existence’ of anything, as well the utter nonexistence of things which exist without intrinsic nature. In other words, the object of negation is the stance which holds that that existence by virtue of intrinsic nature is the only way for anything to exist. According to this undesirable stance, things only exist insofar as they exist 'intrinsically' and insofar as there is no intrinsic nature, there is only void. Tsongkhapa writes:

> Therefore, being free from all extremes of existence – through understanding that nothing exists by virtue of its own essence – and being free of all the extremes of nonexistence in that one can posit everything from cause and effect to the very facts of the world as possessing no intrinsic existence appears to be a distinctive interpretation of the noble Nāgārjuna's thought by the master Buddhapālita and the revered Candrakīrti. So distinguishing two types of existence and two types of nonexistence is crucially important. (2021, 274)

We may wonder whether the permissible kinds of existence and nonexistence here are to be understood as stipulated by convention. This is difficult to say. Certainly, Buddhapālita, Candrakīrti, and Tsongkhapa did not explain their interpretation of Nāgārjuna's obscurational truth to be defined by free choice. But, even though they did not reason in these terms explicitly, we might wonder whether they would reject such a proposal if it was presented to them.

The crux of the matter, framed within Tsongkhapa's concepts, can be represented by the distinction between two views. One view is that 'intrinsic existence' is the only sort of existence that could be established or negated by a truth-valued existential sentence. That would imply that the things Tsongkhapa says we are allowed to posit as existent despite our rejection of 'intrinsic existence' (i.e., "everything from cause and effect to the very facts of the world as possessing no intrinsic essence") are existent only by convention. This reading is materially the same as my ontological conventionalism – it is only that the strong connection between intrinsic nature and existence in Madhyamaka is made explicit by qualifying non-conventional existence as ‘intrinsic’. In this case, the Prāsaṅgika[10] doctrine of adopting the world’s ‘conventions’[11] would amount to accepting commonly accepted beliefs about existents as conventions in the technical sense that I have explained.

The other view is that ‘intrinsic existence’ and mere existence are both established by sentences with genuine truth-values, but each has its respective type of truth condition. The following passage by Tsongkhapa, in which he is explaining one of Candrakīrti’s statements, would seem to fit neatly with this interpretation: “each phenomenon has two natures, and the ultimate is the one that is found by the cognitive process that apprehends reality, and the conventional is the one that is found by the cognitive process that perceives that which is unreal” (Tsongkhapa 2006, 483). The idea here would be that people indeed have cognitive processes which perceive what is unreal and that there is a fact of the matter of what they perceive with those processes. What they perceive is by definition ‘conventionally’ existent, so there is a fact of the matter of what is ‘conventionally’ existent. This means that what is ‘conventionally’ existent is not established by convention.

I don’t know of any material by Tsongkhapa or Candrakīrti that would definitively rule one of these interpretations out. The point of this discussion, though, was not to determine what they thought, but to explore the possibility of interpreting Nāgārjuna as teaching of two distinct types of existence claim instead of the nature of existence claims in general (as per my proposal). If one favours the second view explained just above, then one’s rejection of conventionalism will depend on seeing existence (of any sort – even ‘false’ or ‘deceptive’) as something which is delivered to us in perception via the cognitive processes that deliver to us

[10] Prāsaṅgika is the name used in Tibet to distinguish the interpretation of Madhyamaka associated with Buddhapālita and Candrakīrti from that of other Indian commentators, who were identified as of the ‘Svātantrika’ school. Tsongkhapa is a key figure in developing and reinforcing the Prāsaṅgika- Svātantrika distinction.

[11] In this context I would take worldly conventions as having the meaning of the Sanskrit *vyavahāra* (common practice, accepted custom, etc.). It does not refer to conventions as I have explained them.

other perceptual information. This is as opposed to the view that existence is a metaphysical postulate that we impose onto experience. I have already explained that a sentence reporting an experience cannot be a convention, so if existence is folded inextricably into perception, ontological conventionalism will not work. There is much one can say about this matter of whether existence is built into perception. I will not pursue the matter here. But I would suggest that regarding existence as inextricable from perception would be in serious tension with one of Nāgārjuna's more evocative statements in his Vigrahavyāvartanī:

> If I perceived anything by means of perception, I would affirm or deny. But because that does not exist, there is no criticism applicable to me.[12]

Nāgārjuna is not here disclosing his physical disabilities. His point is that perception as such does not inherently present objects as existent. He claims to be beyond reproach because he neither holds '*x* exists' nor its negation to be true. He is not in the business of defending such claims.

### 1.6 Spackman's Conceptualist Reading

Of existing interpretations of Nāgārjuna, John Spackman's (2014) "conceptualist" interpretation arguably comes closest to my own. Spackman frames his reading as an effort to forge a position between nihilist readings (those which understand Nāgārjuna as negating the existence of all things) and anti-essentialist readings (those which understand him as negating *svabhāva* but allowing for the existence of things). He proposes that Nāgārjuna means not to make a point about existent things as such, but about the very concept of existence, and formulates his "conceptualist thesis" like this:

[12] VV30 translated by Westerhoff (2010).

> It is a core part of the meaning of the concept of existence (*bhāva*) that whatever exists either exists independently (is *svabhāva*) or is dependent for its existence on the independent existence of something else (is *parabhāva*). (Spackman 2014, 152)

According to Spackman, when we realize emptiness we realize that our understanding of what it means for something to exist can never fully be realized. To say that something exists 'conventionally', for Spackman, is to recognize that it meets some conditions we use to apply the concept of "existence," but fails to meet the *svabhāva*/*parabhāva* (intrinsic/extrinsic nature) condition. In such a case we may recognize that we don't have a complete grasp on what the concept means when applied. With this interpretation of Nāgārjuna, he is showing us that we should abandon the idea that our concept of existence can ever be satisfactorily applied so that we don't cling to theories about what exists or does not exist.

It should be easy to see that Spackman's interpretation has a strong affinity to mine. I also consider Nāgārjuna's philosophy to be an attempt to problematize the existence or nonexistence of things on conceptual grounds – not an attempt to come to some information about the nature of the world itself. Spackman's paper is well-argued and insightful, but I think that my interpretation improves on it in some ways. I will first explain what makes my proposal different and then argue for why these differences should be seen as improvements.

Spackman's version of obscurational truth rests on what he calls the "pragmatic thesis":

> *The Pragmatic Thesis*. It is possible truly to apply a concept without endorsing all features of its meaning – without treating all features of its meaning as truth conditions for applications of the concept – as long as one endorses enough of its meaning that the concept retains its identity. (2014, 154)

Spackman is suggesting here that one can make use of the concept of existence to make true claims despite the fact that the full concept cannot apply to anything. It is in this way that one can make true existential claims in the context of obscurational truth. Because Spackman's interpretation sees existential sentences as having truth values, it is clearly not conventionalism. This is the first difference. The second difference is that Spackman's account hinges entirely on the properties of "the concept of existence" while mine does not. I do rely on there being a concept of existence with some stable meaning – this is what gives us a definition of 'existential sentence'. However, my reading does not require that we believe anything in particular about the precise properties of this concept. It requires us to think that it can only be rigorously applied to something which is identifiable, and that worldly things are only identifiable by their intrinsic or extrinsic nature.

The fundamental reason my interpretation is an improvement is that it achieves a comparable goal with less extravagant claims. Needing to make true claims about the core features of the meaning of 'the concept of existence' is a tough place for a philosophical view to put itself. Consider Spackman's Pragmatic Thesis, which has it that we can truly say that things exist even when 'core features' of the concept of existence fail to apply. Why is it not sufficient to say, as I did before, that something exists when it is, and to be is to exist? In all the history of the world when people have said, truthfully (per the Pragmatic Thesis), that things exist, and have understood each other just fine, must we believe that none knew quite what they *meant*? If this is at all plausible, it does not seem like something a philosopher could establish convincingly. If it can be established at all, it would be by some future cognitive science with a rigorous and explanatorily successful theory of concepts. But we don't need to make such claims to find Nāgārjuna convincing. His philosophical argumentation is focused on establishing that things cannot have natures (whether *svabhāva* or *parabhāva*). That is the hard sell. If he can convince us of that, we can get to the issue of existence via the idea that things need natures to be identifiable and only existential sentences which refer to identifiable things can have determinate truth values. We don't need a counterpart to the Pragmatic Thesis to account for obscurational truth because we say that existential sentences are never true – they are conventions.

The strategy of focusing on the truth conditions of a certain kind of sentence is one that is properly suited to philosophers, and it is exactly in line with Nāgārjuna's general stance. He does not want to teach us about what is in the world, not even the concept of existence. He wants to encourage people to stop committing themselves to unwise judgments.

### 1.7 Conclusion

Conventions are declarative sentences which lack truth-values. They are to be accepted or rejected by free choice based on convenience. I have explained my reading of Nāgārjuna as an ontological conventionalist. What he is trying to get readers of his MMK to believe is that claims about the existence and nonexistence of things all lack truth-values; they cannot accurately describe reality. Such claims should only be entertained for the convenience they offer in thinking about and navigating the world.

For Nāgārjuna, this was a corrective project. Fellow Buddhists, enthusiastic in their efforts to understand reality in accordance with Buddha's teachings, attempted to articulate theories which identified the ultimate existents of reality. In attempting to demonstrate that this intellectual activity was misguided, Nāgārjuna advanced a series of stunningly creative rational arguments. It should be obvious that this unique contribution to philosophy is of value not just to his original interlocutors, but to anyone with an interest in metaphysics.

I have sketched the complicated relationship between my ontological conventionalist reading of Nāgārjuna and existing interpretations of his philosophy. We have seen that some alternatives are not conventionalists at all in the relevant sense, some result in a conventionalism that is too extreme, and, for some, the extent to which they would embrace conventionalism is unclear. This paper is the first to interpret Nāgārjuna through the lens of conventionalism in the sense in which that term has been used in European and American philosophy since the $19^{th}$ century, especially in the philosophy of science.

CHAPTER TWO

## Reading Mūlamadhyamakakārikā with Kantian Ideas

I provide a philosophical account of Nāgārjuna's methodology in MMK. I argue that Kant's notion of fundamental categories of the understanding can be used to explain how Nāgārjuna's arguments can plausibly establish what he says they establish. Specifically, I think that Nāgārjuna shows that judgments involving the use of the categories of causation and inherence always fail to be coherent when applied to things which are taken to have intrinsic nature (*svabhāva*).

### 2.1 Introduction

Having explained, in the last chapter, what conclusion Nāgārjuna wants us to accept, I will take up in this chapter the task of explaining how he seeks to convince us of it. That is not to say that I will present his arguments in all their nuance and comment on them – that is well beyond the scope of this project. My goal is to broadly explain his methodology in such a way that one might evaluate whether it is realistic that he could establish what he seeks to establish. To this end, I will show that the Kantian notion of fundamental categories of understanding is a valuable conceptual resource in accounting for how Nāgārjuna's arguments function and how his conclusion might be satisfactorily established.

In §2.2, I will briefly examine some of the history of reading Nāgārjuna along with Kant. §2.3 will advance my proposal for how certain Kantian ideas should be used to interpret Nāgārjuna's methodology. And in §2.4, I will endeavour to clearly distinguish Nāgārjuna's ontological conventionalism from Kant's transcendental idealism.

A note on language: In the previous chapter, I wrote of *svabhāva* in translation as "intrinsic nature", primarily to make the reading experience easier for readers unfamiliar with

Sanskrit. In this chapter, I have opted to leave the term untranslated in order to distinguish it clearly from similar Kantian ideas as well as for economy and precision of expression.

### 2.2 Earlier Kantian readings

Most early European Buddhologists, such as Eugene Burnouf, Max Walleser, and Louis de la Vallee Poussin presented Madhyamaka as thoroughgoing nihilism.[13] The first major Western effort to resist this reading came from Fyodor Stcherbatsky, who was not shy about interpreting Indian philosophical thought through the lens of European idealism. Madhyamaka, which he discusses primarily in his *The Conception of Buddhist Nirvāna* (1968), was no exception. However, it is misleading to call his interpretation of Madhyamaka a *Kantian* one as, e.g., Tuck (1990, 53) does. According to the index, Kant's name is only mentioned a few times in the book, and each of these instances is a footnote. In those footnotes, Stcherbatsky never identifies Kant's views with those of Nāgārjuna or Candrakīrti. In the most substantial of them (1968, 235, 247), he only speaks to the possibility of comparison with Kant on various points. In his brief section on "European parallels" with Madhyamaka (1968, 84-88), Kant is not mentioned once.

The European parallels which Stcherbatsky thinks are most striking are those with F. H. Bradley and Hegel. Of Bradley he writes,

> Very remarkable are then the coincidences between Nāgārjuna's negativism and the condemnation by Mr. Bradley of almost every conception of the every day world, things and qualities, relations, space and time, change, causation, motion, the self. From the Indian standpoint Bradley can be characterized as a genuine Mādhyamika. (1968, 86)

[13] See the discussion and more examples in Westerhoff (2014, 352-3). Wood (1994, 1) offers a similar list.

And regarding Hegel, Stcherbatsky points to his doctrine in *Phenomenologie des Geistes* stating that the properties of objects are exhausted in their relational qualities except for something that can only be described as "thisness". Stcherbatsky identifies that "thisness" with the Sanskrit *tathātā* in Mahāyāna thought and takes relativity in Hegel's sense as the "exact meaning of the term *śūnyatā*", no doubt with MMK 24.18 in mind. Understanding both Hegel and Nāgārjuna to be attempting to approach the true substance of reality through the negation of the merely relative, he declares, "Both philosophers assure us that negativity (*śūnyatā*) is the soul of the universe".

One sense in which Stcherbatsky's reading can be considered "Kantian" is in the fact that he employs Kantian language liberally in his translations and commentary. By "Kantian language" I mean terms used in technical senses associated primarily with Kant's work and innovations, such as "synthetic", "sensibility", "intuition", and of course the distinction between "phenomenon" and "thing-in-itself". He certainly relies on such concepts to formulate the Madhyamaka position and arguments in a way that he and his peers can appreciate, but he doesn't seem to think that Kant and Nāgārjuna were saying very similar things, or that Kant's developed opinions were particularly relevant to understanding Nāgārjuna.

Although Stcherbatsky's reading wasn't truly Kantian, he formulated Madhyamaka in a way which positioned it squarely within Kant's philosophical arena. It was T.R.V. Murti who built the first truly Kantian reading of Madhyamaka on that foundation. Obviously, I cannot offer a comprehensive review of his work on this topic here. But, in this context, the most pressing issue is what the deficiencies in Murti's project are that the present project will avoid. Therefore, I will focus on addressing that concern.

The core problem with Murti's reading is his identification of Nāgārjuna's ultimate/obscurational (*saṃvṛti/paramārtha*) distinction with Kant's phenomenon/noumenon distinction. According to Murti:

> The purpose of the dialectic is to demonstrate the subjectivity of the categories of thought, namely, that they are of empirical validity and can be significantly used within phenomena only; the Noumenon (tattva) is transcendent to thought. (Murti 1980, 294)

The "dialectic" here refers both to Kant's Transcendental Dialectic in the *Critique* and to MMK. According to Murti, they share the same goal and arrive at functionally identical pictures about the intellect and reality. Murti himself recognizes some limitations of this reading at times:

> There is, however, an important difference, almost a vital one, between the Mādhyamika dialectic and the Kantian Critique. For Kant the Critique has no value *per se*; it is intended to make the path safe for faith; and it is Practical Reason that secures for Kant his noumenal values. The Critique of Pure Reason is a brilliant luxury, and is but externally related to the practical discipline which guarantees Freedom, Immortality, and God. In the Mādhyamika, the dialectic itself is true metaphysic or philosophy *par excellence*. (1980, 213)

> Not that the Mādhyamika takes the Absolute and the world of phenomena as two different sets of entities; but the Absolute is nowhere explicitly shown to be *in* things constituting their very soul. The relation between the two is not made abundantly clear. This may be said to constitute a drawback in the Mādhyamika conception of the Absolute. (1980, 237. Emphasis in original.)

But he doesn't take these difficulties seriously enough.

If Kant sees himself both as giving metaphysical claims a firm ground as a "science" and as paving the way for God, the immortal soul, and personal freedom, then how could his system be suitable for the elimination of views – Nāgārjuna's goal according to Murti himself (1980, 212)? Kant was very consciously trying to create a system wherein certain views could be *justified*: metaphysical views according to pure reason and religious views according to practical reason. Murti stakes his case on the point that Kant rejected speculative metaphysics which claimed knowledge of things-in-themselves, identifying this as Nāgārjuna's concern as well (1980, 295). But this is inadequate for interpreting Nāgārjuna, who is commenting on the possibility of a certain mode of being – *svabhāva* – and the consequences of the negation thereof. That is much more substantial than the mere recognition of epistemological limitations, and it directly problematizes the metaphysical and religious views that Kant takes his system to enable. There is a stark incompatibility here.

Murti has been criticised on this issue before. Tuck takes him to task for claiming, on one hand, that Nāgārjuna offers nothing but criticism of other philosophical positions and, on the other, that he is primarily concerned with revealing a pure, transcendent Absolute (Tuck 1990, 48-52). According to Murti, the former is the methodological route to the latter; the Absolute can only be approached through utterly non-conceptual experience, which he glosses as Pure Intuition (in the fully Kantian sense) and identifies with *prajñā* (Murti 1980, 209-12, 236, 299). The intractable difficulty with Murti's position is how we could possibly take Nāgārjuna's goal to be teaching us about something named "The Absolute" which by definition is beyond all predication, including – according to Murti (1980, 229) – existence and non-existence. Murti seems to be trying to have his cake (the absolutist view) and eat it too (the dialectic eliminates all views), making the charge of inconsistency unavoidable.

Murti has also been criticized in this context for misunderstanding the Mādhyamika take on the empirical. Jacques May, in a critical review of Murti's book, argues that Kant places a great deal of value on empirical knowledge while the only interest the Mādhyamika has in it is in demonstrating its falsity from the ultimate perspective (May 1959, 107). This is related to Murti's strange claim (throughout his 1980 but see, e.g., 229, 295), in line with Stcherbatsky (1968, 88) before him, that Nāgārjuna restricts the use of reason and logic to the empirical realm and denies that it can be used to come to terms with how reality really is – a very Kantian point that is very nearly a complete inversion of how reason is used in MMK. Murti and Stcherbatsky never explain how Nāgārjuna's use of reason to show that various ordinary phenomena make no sense when rationally investigated is supposed to be an *affirmation* of reason's usefulness in that sphere exclusively. To the contrary, and in accordance with May's objection, Nāgārjuna seems to be appealing to reason's authority to prove, *from an ultimate perspective*, that empirical reality fails to hold up. In other words, rational determinacy is the standard for ultimate reality in the Madhyamaka tradition; Nāgārjuna doesn't claim things are empty by denying that reason could grasp what is real; he thinks they are empty because reason *would* be able to grasp them *if* they were real and it *doesn't*. To exclude some class of being or beings from the efficacy of reason is what Kant wants since it makes room for his God, soul, and freedom; but this undermines Nāgārjuna's methodology for precisely the same reason.

When Nāgārjuna states or implies that the way things are (*tattva*) is unarisen, unceasing, neither existent nor non-existent, indeterminable by concepts, he is not talking about some mysterious, mystical, absolute, ground of reality; he is talking about sprouts, pots, and human beings. For what Nāgārjuna is doing is using the categories of thought and authority of logic to *do* something – prove the senselessness of *svabhāva*; he is not trying to show that that process (using categories, reasoning) doesn't grasp the ultimate nature of reality. He is attempting to show that according to the correct use of those very categories, things given in our empirical experience cannot be determined. That realization doesn't point to a noumenal reality beyond the categories – it demonstrates that phenomenal things are no more metaphysically graspable than noumenal things are supposed to be.

Although I've focused on the deficiencies of Murti's reading, as have others,[14] I do not think that it is a bad effort. I don't hold it in lower regard than other readings which disagree with my own. In fact, I think Murti did an admirable job in rooting his reading within the

[14] E.g. Garfield (1995, 305-6, n.119) and Bitbol (2003). Though, in criticizing Murti, Bitbol relies for the most part on May and Garfield. Bitbol's paper is itself is an attempt to connect Kant and Madhyamaka. He argues that aspects of both can be useful and complimentary in interpreting quantum mechanics. This work is exciting and important, but it is not and does not claim to be a Kantian reading of Madhyamaka and need not be discussed in the present context.

Buddhist and wider Indian tradition, and his attempts to avoid the problems I drew attention to here were respectable, though ultimately unsuccessful, in my opinion. Furthermore, some of his comments are truly insightful, such as this: "The dialectic is not an avenue for the acquisition of information, but a catharsis; it is primarily a path of purification of the intellect" (1980, 212). This is a point which I think arises out of my interpretation, too, as we will see. From Murti's point of view, it seems to arise from his claim that Madhyamaka is a critical philosophy in the Kantian sense: it is not concerned with forming metaphysical doctrines but with interrogating the possibilities and limits of the philosophical process itself. May criticizes him strongly on this point as well, but I would want to defend Murti here. The basis of May's objection is his insistence that Madhyamaka isn't limited to the criticism of knowledge alone but treats the epistemic and ontological as deeply interconnected (May 1959, 105-8). However, even if we accept his point, and even if we resist the problematic ideas behind Murti's claims, we can still find that the primary goal of Nāgārjuna's work is to get us to recognize the problematic nature of most metaphysical doctrines *as* metaphysical doctrines, rather than to get us to embrace some belief about how things in the world are. My reading takes that path, and I think it is the more promising option for understanding how Madhyamaka is unique, defensible, and potentially soteriologically efficacious.

In the next section, I will advance my positive account of how Kantian ideas should be used in interpreting Nāgārjuna's philosophy.

### 2.3 The Kantian connection

For the reader's convenience, this, again, is how I summarize the argument that MMK is making:

**P1**: Unacceptable consequences follow from any sentence analysed in terms of intrinsic nature.

**P2**: Only a sentence analysed in terms of intrinsic nature could establish the existence or nonexistence of anything.

**P3**: A sentence with unacceptable consequences cannot establish the existence or nonexistence of anything.

**C**: The existence or nonexistence of anything cannot be established.

We saw some textual evidence for P2 and P3 in the previous chapter, but they are assumed rather than argued for in MMK. The only premise Nāgārjuna is committed to rigorously arguing for is P1, and it is more than enough for his plate. Undermining *every* sentence involving *svabhāva* is a daunting task to undertake.

With P1, I am not claiming that Nāgārjuna attempts to demonstrate that every sentence analysed in terms of *svabhāva* (let's say S-sentence) is *self*-contradictory or absurd in and of itself. That would indeed be too much. As the word "yields" indicates, he wants to show that any S-sentence unavoidably leads us to absurdities and/or contradictions in further thought.

Nāgārjuna goes about this by showing that certain kinds of S-sentences yield contradiction and absurdity. The idea is that the types of judgments he focuses on are so fundamental or general that taking anything to have *svabhāva* is implicated in their unacceptable consequences. A reasonable person should have doubts about such an extreme claim. It would be key for Nāgārjuna's project to realize a systematic connection between the judgments he chooses to focus on and the totality of possible judgments. This is where Kant comes into the picture, for Kant argues that there are a set of concepts that serve as the basis for all judgment, all thought: his categories, or pure concepts of the understanding. According to this theory, every judgment we make must be ultimately rooted in the application of some set of these concepts. This suggests that it should be possible to undermine all S-sentences by demonstrating a fundamental incompatibility between the categories and what all S-sentences have in common: svabhāva.

The goal of the present essay is to show that this is a viable reading of the MMK's philosophical strategy. This is not to claim, of course, that Nāgārjuna shared Kant's thoughts on this matter, but that the arguments he provides are methodologically appropriate, meaning that a Kantian commentary could lend the text systematic support and elaboration. It should be remembered that Kant offered the Critique not as a metaphysical view but as a criticism of how metaphysical views can and cannot be properly defended and understood. In this capacity, a Kantian reading of the text is not so much comparative as critical. To make our point, we must look at the nature of the categories, the form of Nāgārjuna's arguments, and what connection between them we can read into the text.

The following is how Kant presents what he believes is a full list of the categories.[15]

**1.**
**Of Quantity**
Unity
Plurality
Totality

**2.**
**Of Quality**
Reality
Negation
Limitation

**3.**
**Of Relation**
Of Inherence and Subsistence
(*substantia et accidens*)
Of Causality and Dependence
(cause and effect)
Of Community (reciprocity
between agent and patient)

**4.**

[15] A80/B106. All quotations from the Critique are taken from the 2003 translation by Guyer and Wood.

**Of Modality**
Possibility – Impossibility
Existence – Non-existence
Necessity – Contingency

This table is notoriously controversial. Luckily, this piece is a commentary on MMK, not the *Critique*, so we only need to engage with it insofar as it is relevant to the arguments presented in the former. I would contend that the arguments of MMK are overwhelmingly concerned with two of the concepts on this table: causality and dependence, and inherence and subsistence (hereafter just "causality" and "inherence"). The Kantian reading of this text need not be committed to the full accuracy or completeness of Kant's table. It must be committed to the statuses of causality and inherence as categories,[16] it must show that Nāgārjuna's arguments undermine their use in S-sentences, and it must show how this may undermine confidence in S-sentences generally.

First, let us look at Nāgārjuna's treatment of causation. MMK 1.1 presents the thesis of the chapter:

> Not from itself, not from another, not from both, nor without a cause:
>
> Never in any way is there any existing thing that has arisen.

The presumption here is that the judgments being considered are S-sentences. This thesis represents the absurd consequence that Nāgārjuna argues follows from an analysis of causal S-sentences in general. Let us look at how he argues his case.

One of his strategies focuses on the relation between conditions and the conditioned. He argues that, since the intrinsic nature of an existent cannot be found in its conditions, it is

[16] This includes reading, with Kant, Hume's rejection of real causal relations as a rejection of real causation, not a valid alternate understanding of it (Cf. Siderits 2013).

impossible to establish why and how that nature should spring from these conditions and not non-conditions (MMK 1.2, 1.11-12). This resembles Hume's point about it being impossible to establish causal relations analytically, but Nāgārjuna's argument has a crucially different focus. Nāgārjuna isn't interested in denying causes and conditions – he uses the premise that all things have them in his arguments (E.g., MMK 4.2, 24.19). Rather, he is interrogating how *svabhāva* could find purchase among causal processes and relata. His strategy is therefore consistent with a Kantian understanding of causation, for even if we accept that causation is impossible to verify or negate, how we ought to make sense of *svabhāva* in causal contexts remains a legitimate question. The idea in this argument is that it is difficult to see how something specifically defined by intrinsic-ness and independence can have some special connection with its conditions that it doesn't have with non-conditions, despite being present in neither.

His other arguments in this chapter revolve primarily around what Siderits and Katsura (2013, 8) call the "Three Times" argument form: it is impossible to identify a time at which conditions produce their effects. Keeping in mind an ultimately real effect with an intrinsic nature, it must either exist or not exist – there is no in-between. It doesn't make sense to say the conditions function before the effect exists – what are they conditions *of* at that time? It also doesn't make sense to say the conditions function when the effect exists, because then what would their functioning accomplish? And there is no third time at which the effect is both existent and non-existent, as that would be contradictory (MMK 1.4-7; Siderits and Katsura 2013, 22). Since there is no time at which an entity can be causally produced, we can be assured that it doesn't happen at all. And, of course, entities do not arise without conditions; this would mean, absurdly, that anything could come into existence under any circumstances.

It is tempting to think that each chapter of MMK attempts to establish the emptiness of some specific thing, but this is easily shown to be inaccurate. In the chapter just discussed the argument is specific not in *what* it shows to be empty but in *how* it targets things: through the concept of causation. The strategy of aiming at generality through specific concepts is typical of MMK and reveals itself even when it appears, superficially, that Nāgārjuna has a specific target in mind. A good example is MMK 4, which superficially seems to be making an argument about *rūpa* (form, matter), but is actually repeating the point of MMK 1: that causal relations imply emptiness. He does this by focusing on the question of whether *rūpa* is identical or distinct from its cause – a question which, if *rūpa* had *svabhāva*, should have a perfectly determinate answer – and finds neither option satisfactory. Nāgārjuna reveals in 4.7 that *rūpa* is a stand-in for not just any *skandha* but any existent whatsoever.

Other chapters, including MMK 8, 12, and 22, join MMK 1 and 2 in arguing that things with *svabhāva* could not participate in causality. These threads culminate in the climax of the text, where Nāgārjuna advances this argument:

> If you look upon the existence of things as coming from intrinsic nature, you
> thus see things as being without cause and conditions. (24.16)
>
> Effect and cause, and indeed agent, instrument, and act, arising and ceasing,
> and fruit, you deny. (24.17)
>
> Dependent origination we declare to be emptiness.
> It [emptiness] is a dependent concept; just that is the middle path. (24.18)
>
> There being no dharma whatsoever that is not dependently originated,
> it follows that there is also no dharma whatsoever that is non-empty. (24.19)

Because of his extensive campaign against causal S-sentences, Nāgārjuna declares that anyone who sees things as having *svabhāva* is necessarily denying cause and effect. He stresses the disastrous consequences this would have for a Buddhist and derives the inseparability of the doctrine dependent origination and emptiness. But more on this later.

We have seen that Nāgārjuna thinks himself to have undermined all causal S-sentences. S-sentences of inherence receive a similarly thorough (if not as climactic) treatment in the text. MMK 2 argues that motion could not be predicated of something in an S-sentence without absurdity. Recent commentators disagree on the exact nature of his argument but agree that it takes a form meant to be highly generalizable (Westerhoff 2009, Chapter 6; Oetke 2011; Arnold 2012). In MMK 3 it is argued that it makes no sense to understand vision as inhering in an ultimately real seer. This is extended to all senses and, we can venture, any case in which we might want to see a capacity as inherent in an object. MMK 9 reinforces this by attempting to show that the bearer of the senses could not exist before, simultaneous with, or after the senses it supposedly bears when analysed in terms of *svabhāva.* In MMK 5, it is argued that it makes no sense how a characteristic inheres in what it characterizes when they are considered non-empty. MMK 6 goes the same route with a state (desire) and its possessor (the one who desires). This cluster of chapters early in the text aggressively attack S-sentences of inherence toward the same end as the chapters which do the same for causal S-sentences.

What can we make of the attack on S-sentences of causation and inherence from a Kantian perspective? Nāgārjuna's arguments, if successful, show that it would be impossible to use these two categories to think about things with *svabhāva*. But the categories, according to Kant, are *constitutive* of our ability to understand, think of, and even experience things (A93/B126). And, as Béatrice Longuenesse emphasizes throughout her work,[17] they are constitutive of our capacity to *judge*. Keeping P1 in mind as something we want to take from MMK, this reading shows that Nāgārjuna's use of causation does not merely address a kind of judgment that we happen to make frequent use of – it addresses our innermost *capacity* to make judgments as such.

[17] For further reading consider Longuenesse 1998, 2005, 2006.

For Kant, all the categories are tied directly to specific forms of judgment applied to objects of experience. The category of causality and dependence is tied to the capacity to make hypothetical judgments; it is the concept which allows us to think of objects such that we may make hypothetical judgments about them. Nāgārjuna's prohibition of causal S-sentences is therefore a prohibition of all hypothetical S-sentences. For Kant, causality is schematized in our experience according to "the real upon which, whenever it is posited, something else always follows" (A144/B183). In other words, for any judgment of the form "if *x* then *y*" where the variables are understood as existents, the category of causality *consists in* that judgment's intelligibility. Since the contents of S-sentences are, by definition, real or potentially real existents, the intelligibility of any hypothetical S-sentence consists in the applicability of causality to its contents.[18] Nāgārjuna's arguments are designed to show that the application of causality to things with *svabhāva* can never hold. According to Kant's system, this implies that no hypothetical S-sentence can hold.

Inherence/subsistence is associated with categorical judgments: the form of judgment where a predicate is asserted of a subject. By undermining the applicability of this category to *svabhāva*, we can interpret Nāgārjuna as ruling out all categorical S-sentences. In combination with the rejection of hypothetical S-sentences, this means that for any S-sentence, we could predicate no properties of its contents, nor conceive of them in relations of dependence. This is not strict proof of P1's truth, but it is very hard to see how any such judgment could be maintained without absurdity following in some form or other.

[18] Some judgments about things can have a hypothetical form without seeming to rely on causation, like the judgment, "If this is an earth atom, this is hard." But note that the content of this judgment reduces to an application of "Earth atoms are hard." Note also that any even slightly more complicated judgment of this form invokes causation: "If this is a ball, this rolls." Cases like this may or may not frustrate Kant's system, but since I will argue that categorical judgments and hypothetical judgments meet identical fates under Nāgārjuna's arguments, this matter can be safely put aside for the purposes of this paper. I am indebted to an anonymous reviewer for bringing such cases to my attention.

It is possible that someone would find no absurdity flowing from some judgment about something free from causality and inherence.[19] This is difficult to address since people's intuitions on what counts as absurd vary widely, especially when it comes to religion and spirituality. But Nāgārjuna is addressing his arguments in MMK to a specific audience: a Buddhist Ābhidharmika audience. He is counting on the fact that they would find it unacceptable for their *dharmas* to be unaffected by the doctrine of dependent origination. Therefore, at the climax of the work he relies on the unargued premise that all *dharmas* are dependently originated (MMK 24.19ab). It is the thrust of MMK 24 as a whole that Buddhist teachings could not apply to non-empty things; Nāgārjuna does not think that his opponents will accept this fate for *dharmas* – his argument would be undermined if they did.

In a less Buddhist, more Kantian vein, Nāgārjuna's arguments show that non-empty things would be immune to the most fundamental elements of our experience, two of the concepts which constitute our ability to make sense of the world. It is Kant's opinion that we could never be justified in claiming knowledge of such a thing, as it would necessarily be ungrounded by experience, hence fabricated by our understanding, hence "a mere thought-entity" (A566/B594). A judgment about such a thing (i.e. an S-sentence) which somehow manages to avoid implicating it in causal or inherence relations would still be burdened with demonstrating that it is not mere fancy. That it is unclear how this burden could be met is partly what motivated Kant's mission to reform metaphysics in the first place.

The takeaway from this discussion is that it would misrepresent the argument of MMK to say that all S-sentences should not be asserted because they are intrinsically contradictory, inconceivable, or anything to that effect. They *can* be said – the difficulty is the cost. Nāgārjuna's *prasaṅga* style of argument relies on drawing consequences from the opponent's

[19] Perhaps a theory of eternally existent tropes would qualify. Siderits (2003b, pp.122-23) and Westerhoff (2009, 38-40, p.204) have discussions pertinent to how this might be resisted.

position which the opponent will not be able to tolerate. In MMK, where the opponents are Buddhists, these consequences are both religiously and intellectually perilous. The case is arguably only airtight where the opponent is committed to the applicability of causality and/or inherence to all things, as Nāgārjuna believes Buddhists are due to the doctrine of dependent arising. But it is hard to see how a profitable discussion on metaphysics could be held with anyone who disavows these concepts for any object of significance.

To summarize, the Kantian reading takes Nāgārjuna's arguments as prohibiting the use of the categories of causality and inherence in thinking about things with *svabhāva* on pain of absurdity and contradiction. Since these categories are essential to our ability to think about and experience things in general, *svabhāva* seems to inevitably invite absurdity.

This reading has the attractive quality of offering a systematic understanding of how Nāgārjuna's methodology manages to establish its most important goal: the *universality* of emptiness. Other authors attempt to account for this by pointing to the general applicability of Nāgārjuna's arguments (Siderits and Katsura 2013, 7) or the fundamental importance of the concepts that he focuses on (Westerhoff 2009, 92), but they do not offer a systematic account of what *makes* an argument general, or what the fundamentality of his topics *consists in*. The Kantian reading can contribute by offering a clear answer: the purity of the concepts being used is the key.

Proving this is the largest part of MMK's job, but in my rendition of Nāgārjuna's project it comprises only the first premise. In the next section I will attempt to show that it's the remainder of Nāgārjuna's argument that differentiates it from Kantian transcendental idealism and offers something unique in world philosophy.

## 2.4 **Emptiness and Transcendental Idealism**

In this section I will discuss the wider relations between Nāgārjuna's emptiness and Kant's transcendental idealism. Of course, Kant's transcendental idealism is subject to a plethora of competing interpretations itself and I will try not to beg the question against any of them. The point will be to indicate how what Nāgārjuna is doing is novel in a Kantian context.

We will start by looking at the status of the concept *svabhāva* in a Kantian context. Usually translated as intrinsic nature, it means the nature that an object has in and of itself, independent of outside influence. This has quite an obvious connection with the Kantian notion of a thing-in-itself or noumenon,[20] which is the notion of a thing independent of human cognition. Kant himself offers a conceptual parallel to *svabhāva* when discussing substances. He says that "[s]ubstances in general must have something **inner**, which is therefore free of all outer relations, consequently also of composition" (A274/B330). In discussing the schematism of the categories, he explains that substance is the "persistence of the real in time" that allows us to understand endurance in amidst change (A144/B1833). According to the quote above, Kant seems to think that this relies on the imputation of an inner something to what is understood as a substance. But, as Rae Langton (1998, 49-50) points out, this inner something could only be determined by intrinsic properties, which lie beyond the bounds of our understanding. At best, the pure understanding can concoct a transcendent object corresponding to pure substance, but of this object "in thinking it as a thing determinable by its distinguishing and inner predicates, we have on our side neither grounds of its possibility (since it is independent of all concepts of experience) nor the least justification, and so it is a mere thought-entity" (A566/B594). It follows that something which actually has *svabhāva* –

[20] Some scholars distinguish between things-in-themselves and noumena (e.g. Oizerman 1981) while others don't (e.g. Langton 1998, 12) or acknowledge that the distinction is there and choose to not bother with it (Van Cleve 1999, 134). I am adopting the third approach; there is no need to distinguish between noumena and things-in-themselves for my argument.

substances in the fullest sense[21] – could only be a thing-in-itself, impossible for us to determine, destined to be forever beyond our cognition.

This is not a commitment to the idea that things-in-themselves and things with *svabhāva* are coextensive, nor even that Kant's concept of the noumenon was a metaphysical one. It is merely the point that *if* there were something with *svabhāva*, it could not, by definition, be given in Kant's concept of the phenomenal and would therefore be noumenal.[22] Since nothing with *svabhāva* could be a phenomenon, judging an intuited object to have *svabhāva* would *always be an error according to Kant's system*. It would never be correct to judge an object we can experience as having *svabhāva*, and the understanding cannot judge anything *but* objects we can experience.

Therefore, insofar as MMK denies that anything we could experience could have *svabhāva*, a Kantian already agrees. Insofar as it denies that anything we *couldn't* experience could have *svabhāva*, a Kantian would say that its methodology is not capable of giving us any knowledge of things we cannot experience. From a Kantian's point of view, then, Nāgārjuna is telling them something they already know, and may or may not be additionally trying to make dogmatic assertions about things which fall outside the purview of the understanding. If we think that all Nāgārjuna is trying to establish is the incorrectness of judging anything as having *svabhāva*, a Kantian would find nothing in his work that goes beyond the transcendental idealism which they already uphold. With this in mind, Murti's identification of the two systems does not seem unreasonable. If Nāgārjuna's view is to be fundamentally different from transcendental idealism, it needs something more than the incoherence of *svabhāva*.

---

[21] Van Cleve (1999, 120-1) argues that it is problematic for the Kantian concept of substance to be applied to phenomena for similar reasons as I have. Ameriks (2000, 269) also points out the link between substantiality and essential nature for Kant, including helpful primary references (see n.79). But see Bird (2006, 580-5) for a critique of Ameriks and contrary reading of his references.

[22] Allison, whose interpretation of the noumenal is distinctively non-metaphysical, acknowledges this point explicitly (2004, 52).

It seems to me that the novelty of the Nāgārjuna's position when approached from a Kantian perspective lies not in P1 but in P2, P3, and C. In other words, it comes not from his rejection of *svabhāva* but what he thinks the existential import of that rejection is. I argued in Chapter One that, without appeal to *svabhāva*, Nāgārjuna rejected the idea that existential sentences can have truth values. Kant certainly did think that they could. Kant was no conventionalist. The motivation for his system was in large part the desire to explain how one could make judgments with determinate truth values, including existential ones. This seems to be the most significant difference in how Mādhyamikas and Kantians approach the world. I will now compare Kant's approach to existential judgment with Nāgārjuna's to clarify the disagreement and its consequences.

Kant famously claimed that "[b]eing is obviously not a real predicate" (A598/B626) because, for him, the conceptual determination of an object must be independent of the question of whether it exists. He argues that this is necessary to preserve the strict connection between an object that is thought as merely possible and the object as it exists:

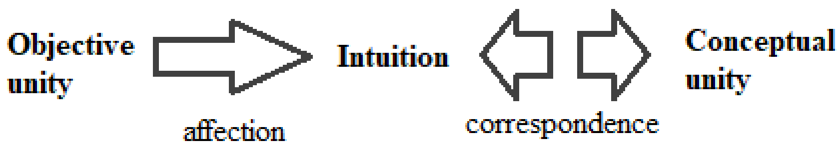
> Thus when I think a thing, through whichever and however many predicates I like (even in its thoroughgoing determination), not the least bit gets added to the thing when I posit in addiction that this thing **is**. For otherwise what would exist would not be the same as what I had thought in my concept, but more than that, and I could not say that the very object of my concept exists. (A600/B628. Emphasis in original.)

Note that Kant says that a positive existential judgment cannot be made if, in the process, even the slightest addition (or, presumably, subtraction) is made in the determination of the existent object that differentiates it from the determination of the conceptual object.[23] We can therefore

[23] In discussing the transcendental ideal, Kant sets impossible standards for perfect determination, taking it to involve use of the "sum total of all possibility" and be therefore just a prescription of pure reason (A573/B601). It's unclear to me how he thinks his frequent appeals to determinacy are compatible with this point. Ultimately, I don't think that there is a solution available to him because I think that Nāgārjuna is right.

reasonably hold him to P3: an existential judgment can only be rigorously established insofar as the object can be rigorously determined. To understand how this works we need to look at Kant's positive view on how existential judgments are made.

For Kant, an object is "that in the concept of which the manifold of a given intuition is **united**."[24] If that phrase seems convoluted it is because Kant is trying to point simultaneously to two separate unities with it. He is saying that the manifold of an intuition is united within a *concept* but also that this concept is *of* something distinct from itself: the object. In a veridical existential judgment, that object affects sensibility, yielding its intuition. We may say that an object exists when it is given to us through an intuition which corresponds to the concept of that object in our mind.[25] Since the object that is represented in intuition is itself meant to be united, we have unities on both "ends" of intuition. The anatomy of a Kantian existential judgment can be visualized like this:




As the quote just above reveals, Kant has a high standard for the strictness of the correspondence between the concept of an object and the intuition through which the object is presented. Such a standard can only be rigorously met if both the intuition and the concept are determinate, so that we can be sure that there is no difference between them.

From Nāgārjuna's perspective, the concept of an object can only be satisfactorily determined through a grasp of its nature. That nature is its unity and we can determine the object's existence only insofar as we have a full grasp of that nature. Nāgārjuna diverges from

[24] B137. Emphasis in original. Compare the definition he gives in his later notes: "What is an object? That whose representation is a sum of several predicates belonging to it." (18:676; 2005, 387).
[25] For a more thorough discussion of this than I have space for here, see Kannisto (2017), especially pp. 20-23.

Kant when he insists that understanding an object's nature necessarily involves understanding whether it is *svabhāva* or *parabhāva*. Therefore, determining its nature necessarily involves determining its *svabhāva* or the *svabhāva* of that upon which its nature depends. It follows that the absence of *svabhāva* among phenomena leaves no basis for the determinacy of object-concepts. There is therefore no basis for rigorously establishing the truth or falsity of an existential sentence. Kant himself lends support to Nāgārjuna's claim (MMK15.5) that judgments of non-existence must go along with those of existence, for Kant holds that "no one can think a negation determinately without grounding it on the opposed affirmation" (A575/B603).

A Kantian response to Nāgārjuna would have to explain how the concept of an object could be metaphysically determinate without involving *svabhāva* and *parabhāva*. Furthermore, it would have to show how an intuition could achieve perfect correspondence with such a determinate concept. For a classic example, if we want to make an existential judgment about a chariot, how do we hold a determinate concept of its nature without having that concept be dependent on the concepts of the natures of its parts, and their parts, down until we find something with a nature that is not dependent at all? How do we verify that an intuition corresponds perfectly with this concept? As shown in the quotation above, Kant himself seems to accept this level of rigor as appropriate. For Nāgārjuna, our inability to attribute *svabhāva* to object-concepts means that there could be no such thing as perfect correspondence between them and any supposedly objective unity given via intuition, which renders them useful merely for use in conventions.

What does this difference mean for the noumenon? How that is answered depends in part on what interpretation of the noumenal we read into Kant, but I think that we can sketch a Nāgārjunian response to each major kind of interpretation that is basically dismissive. Interpretations of transcendental idealism are usually categorized as either "two-object" (alternatively "two-world") or "two-aspect" interpretations. The "two-object" view sees phenomena and noumena as two ontologically distinct kinds of object.[26] The "two-aspect" view sees them as two aspects of the very same thing. Two-aspect interpretations are further subdivided into metaphysical and epistemological variants. Metaphysical variants take the noumenon to be the inaccessible intrinsic nature of an object which is known to us through its non-intrinsic, cognitively dependent nature (the phenomenon). Epistemological variants take the noumenon/phenomenon distinction to not be between ontological kinds but between ways of *considering* things.[27] On this view, an object considered as it is in itself is its consideration as apart from all forms of sensibility. This amounts to the vacuous concept of an object in general, concocted by the pure understanding, as nothing could be known of an object when considered in this light.

I think that Nāgārjuna would be dismissive of the philosophical significance of the distinction under any of these interpretations due to how seriously he takes the idea that no *basis* for the truth or falsity of an existential sentence can be found. We can see how seriously he takes this point in *Vigrahavyāvartanī* 30 and its autocommentary (Westerhoff 2010, 30), where he denies that any object *even appears to him* to be affirmed or denied. Against a "two-object" interpretation, then, he could ask, *by analogy to what* could we posit things-in-themselves? No first kind of object (phenomena) can be established for us to posit a second kind (noumena). Against a metaphysical "two-aspect" interpretation he could ask, what is the noumenon the intrinsic nature *of*? Since no determinate object appears, the first aspect cannot be established – so there is nothing for the second aspect to pertain to. Against epistemological "two-aspect" interpretations he could give a similar response: no object has been found to be

[26] Proponents of this view include Guyer (1987) and Van Cleve (1999).
[27] Proponents of the metaphysical two-aspect view include Langton (1998) and Ameriks (2012); the epistemological view is championed by Allison (2004).

considered in either of the two ways suggested. For Nāgārjuna, an "objective unity" is there neither phenomenally (in space/time) nor otherwise.

These responses are not rigorous analytic take-downs of the phenomenon/noumenon distinction, but that's precisely why they might be effective. Kant designed his system so that reason could never verify or negate the noumenal. A dismissiveness towards the very inclination to apply his distinction via a rejection of there being something for the distinction to pertain to not only circumvents its innate resilience to criticism but is characteristically Nāgārjunian.

It should be noted that interpretations of Nāgārjuna which hold that existential judgments can be defended for 'conventional' objects will have to find a different escape from Kant if they want one.[28]

## 2.5 Conclusion

I have attempted to make two main points in this chapter.

First, I have argued that the argumentation of MMK can be profitably read as seeking to show that the concept of svabhāva cannot be coherently combined with at least two of the Kantian pure concepts of the understanding in judgments: "causality and dependence" and "inherence and subsistence". This reading provides a systematic explanation of how Nāgārjuna's arguments can implicate any and every judgment that involves *svabhāva*, something that is otherwise missing from the literature on this text.

Second, I argued that an adherent to Kantian transcendental idealism would not find Nāgārjuna's arguments against *svabhāva* any threat to their views, as they would readily accept

[28] Spackman's conceptualist interpretation might struggle as well, as if the crux of the matter is that we don't fully understand what it means to say that a phenomenon exists since it lacks svabhāva, a Kantian could say that we do understand existence in that very context when we acknowledge that the object has intrinsic properties which are unknowable by us. One might say that Kant puts the concept of existence back together in a way that is consistent with Nāgārjuna's methodology.

that the understanding could not function with regard to the intrinsic nature of things. In order to establish Nāgārjuna's position as unique and distinct from transcendental idealism, I argued that, rather than emptiness as such, we should appeal to how Nāgārjuna undermines existential sentences about even mundane objects of experience. By taking that move seriously, we can construct for Nāgārjuna a response to transcendental idealism that can be best described as dismissive.

The motivation for this chapter is the desire to show how Nāgārjuna's project could be plausible from a methodological point of view. The chapter itself, of course, is not a defence of Nāgārjuna, but I do not think that he can be properly defended or appreciated unless one has a firm sense of the kind of philosophy he was doing and what that kind of philosophy is capable of establishing. In the next chapter, we will consider the place of this style of philosophy in discussions in which naturalism is the key methodological virtue.

# CHAPTER THREE

## Naturalism and Nāgārjuna

In this chapter, I investigate whether Nāgārjuna's philosophy can be considered a legitimate intellectual project from a firmly naturalistic stance. I begin with a sympathetic discussion of Ladyman and Ross's (2007) criticism of non-naturalistic metaphysics. Proposing a useful redefinition of the word "metaphysics", I show that Ladyman and Ross's arguments apply only to a subset of its referents and argue that Nāgārjuna's philosophy is a member of a different subset which could be described as 'reflective' in the sense that Bitbol (2010) uses to describe the type of metaphysics recommended by Kant. There are no apparent reasons to object to this style of metaphysics from a naturalistic stance.

### 3.1 Introduction

In Chapter One, I explained what Nāgārjuna was arguing. In Chapter Two, I explained how he attempts to argue for it. In this chapter, I will explain what kind of philosophy he was doing. By its subject matter, it seems to be some sort of metaphysics. If it is metaphysics, what type is it, and is it a type of thing worth doing at all? Though these are good questions to ask in any context, I am not pursuing them aimlessly, here. Both because of the nature of this project as partially philosophy of science and due to my own convictions, I find it particularly crucial to figure out how Nāgārjuna's philosophy fares from a naturalistic point of view. I will not attempt a rigorous definition of naturalism; suffice it to say, for now, that it is a stance which places great value on the methods, achievements, and spirit of science, especially concerning science's penchant for advancing knowledge about reality independent of what anyone wants to be true (at least, with more independence than other human institutions).

With that in mind, we will begin by testing Nāgārjuna's philosophy against the standards of naturalism preached by James Ladyman and Don Ross as the primary authors of *Every Thing Must Go* (2007), a book in which they defend a metaphysical view known as ontic structural realism. I will address this view in more detail in Chapter Six. In the first chapter of their book, Ladyman and Ross (hereafter LR) take to task metaphysics which presents itself as giving us knowledge of reality by relying primarily on intuitions and rational thought rather than the up-to-date findings of science. They develop and propose a strictly limited form of 'naturalised' metaphysics and declare that this is the only form in which they are willing to take metaphysics seriously. I am generally sympathetic to their criticisms of traditional metaphysics, but I will show that there is at least one other sense of metaphysics which does not fall within their narrow criteria but against which they have demonstrated no cause to object. It is this style of metaphysics that I think Nāgārjuna is doing.

### 3.2 The Naturalistic Challenge

In what follows I will argue, in harmony with LR, that Madhyamaka is methodologically incapable of reliably informing us of how reality is or is not. It is revealing that, to the best of my knowledge, there is no philosophical jargon which simply means 'metaphysics which attempts to describe the world' as opposed to metaphysics concerned only with conceptual criticism or elaboration. The closest fit must be Strawson's distinction between descriptive and revisionary metaphysics set out at the beginning of his *Individuals* (1959, 9): "Descriptive metaphysics is content to describe the actual structure of our thought about the world, revisionary metaphysics is concerned to produce a better structure." I have not been able to find literature that explicitly declares the equation of revisionary metaphysics with 'metaphysics about the world' and descriptive metaphysics with the opposite; the closest is a

1967 article by R. L. Phillips in which it is mentioned that "most" revisionary metaphysicians are, "of course", trying to speech about the *Ding-an-sich*, the thing-in-itself rather than thought schemes (Phillips 1967, 106). Other than that, philosophers have told me in personal conversation that the descriptive/revisionary distinction is what comes to mind when I say that I am looking for terminology for metaphysics about the world.

It's straightforward enough to see what the connection is between the distinctions: revisionary metaphysicians will usually prefer one way of thinking about the world to another based on what they think the world is actually like. Descriptive metaphysicians are not trying to urge a change, so they do not need such external motivation. But this, it seems to me, is as far as the connection goes. A descriptive metaphysician may easily believe that they describe both thought *and* the world (Strawson has Aristotle as an example from the descriptive camp), and it is conceivable that someone might argue for revisions in the ways we think about the world for reasons unrelated to how they think the world is or is not. In fact, I think that Nāgārjuna is an example of this latter case. He does want us to change how we think about the world, or at least to realize that there is something very wrong with it – a normative exercise unbefitting of Strawson's "descriptive" label. But it would be a mistake to think that Nāgārjuna wants to convince us of how the world is or isn't. He is revisionary but he is not telling us about the world.

Going forward, I will refer to metaphysics about the world as "world metaphysics". This is inclusive of the non-naturalist metaphysics that LR condemn and their own naturalized metaphysics. It is important to use this qualification clearly because it is only with the mistaken and easily adopted assumption that all metaphysics is world metaphysics that Madhyamaka fails the naturalistic challenge. We need to see why this is so, and then we can explore what Madhyamaka is doing differently.

A central problem with non-naturalistic world metaphysics, according to LR, is that it relies on intuitions to achieve its discoveries. This is how they summarize their objections according to their naturalistic point of view:

> Attaching epistemic significance to metaphysical intuitions is anti-naturalist for two reasons. First, it requires ignoring the fact that science, especially physics, has shown us that the universe is very strange to our inherited conception of what it is like. Second, it requires ignoring central implications of evolutionary theory, and of the cognitive and behavioural sciences, concerning the nature of our minds. (Ladyman and Ross 2007, 10)

LR identify non-Euclidean space, a lack of deterministic causation, and non-absolute time as things which philosophers have "confidently" ruled out based on their metaphysical intuitions, but which science proceeded to entertain and verify (2007, 16). This is only small sample; their first chapter is replete with examples of metaphysicians (who overwhelmingly identify themselves as naturalists) cleaving to intuitions which are at odds with contemporary physics. More than once, the common assumption of the world as "mereologically composed of simple atoms" is dismissed as an imaginary image which "has no more in common with reality as physics describes it than does the ancient cosmology of four elements and perfect celestial spheres" (2007, 20). It is well known that quantum mechanics aggressively violates common intuitions. LR take to task metaphysicians who think they can avoid this fact by bracketing QM as a special case while continuing to draw conclusions about the world from more comfortable presumptions. Quoting Crawford Elder's opinion that "With rare and strange exceptions, we suppose that extended objects of any kind cannot simultaneously occupy two discontinuous spatial regions", LR object: "Physics knows nothing of the class of 'extended objects'; and the physical objects that occupy two or more discontinuous spatial regions are basic and ubiquitous" (2007, 23). I agree with LR that one cannot hope to arrive at secure truths about how the world is by ignoring scientists' most profound discoveries about it in favour of more easily digestible intuitions.

Perhaps someone can seek justification for the intuition-world connection within some theory of the human mind itself. But, as LR point out, our best science does not support this strategy either. According to our best theories at present, human beings, like all other living things, came into existence according to reproductive selection pressures over a great deal of time. We also know from cognitive science that our minds function in ways that often do not favour fidelity to the external world, and they may shift quite dramatically according to culture and upbringing. Of course, someone who feels no allegiance to modern naturalistic thinking may find other reasons for believing in the intuition-world connection, as Descartes did by believing a good God would not have given him intuitions which mislead him. Against such a person I have no argument to advance, except to note that I find it unlikely their reasons are on firmer epistemic grounds than the theoretical foundations of physics and biology.

I think that one of the most striking insights into how problematic the methodology of non-naturalistic world metaphysics is comes from a reflection Daniel Dennett has about the philosophy of mind:

> In most sciences, there are few things more prized than a counterintuitive result. It shows something surprising and forces us to reconsider our often tacit assumptions. In philosophy of mind a counterintuitive 'result' (for example, a mind-boggling implication of somebody's 'theory' of perception, memory, consciousness or whatever) is typically taken as tantamount to a refutation. (Dennett 2005, 34)

When they quote this passage, LR (2007, 15) agree with Dennett that this form of conservatism in philosophy is not always a bad idea. However, it is clearly hubristic as a core methodology for making discoveries about the objective nature of reality. This comparison reveals a key virtue underlying naturalism which metaphysics generally cannot claim for itself: humility. A naturalistic attitude implies a willingness to be puzzled by the world. At every step in an ideal naturalistic methodology, reality is given every possible opportunity to violate the assumptions and intuitions of the investigator. Not even the most ingrained common sense is safe from the permanent principle: "We may be wrong."[29] This cannot be duplicated by any metaphysical methodology, because no metaphysical theory is open to empirical testing the way that scientific theories are. Intuition – at best, logical intuition – is the source of rigor in metaphysics as such; reality isn't given the chance to keep a metaphysician humble.

At this point, it will be helpful to see the sense in which LR consider themselves to be doing metaphysics. According to them, the only metaphysics worth doing is a fully naturalised metaphysics which takes as its sole mission the unification and clarification of scientific hypotheses. Their specific guiding principle is the Principle of Naturalistic Closure (PNC):

> Any new metaphysical claim that is to be taken seriously at time *t* should be motivated by, and only by, the service it would perform, if true, in showing how two or more specific scientific hypotheses, at least one of which is drawn from fundamental physics, jointly explain more than the sum of what is explained by the two hypotheses taken separately… (Ladyman and Ross 2007, 37)

The PNC is clarified and defended in detail, but we need not explore it too deeply here. What is important to note is that LR see the proper practice of metaphysics as tied strictly to contemporary science (and physics specifically). Importantly, this means that metaphysics is contingent on the science of its day. When the science changes, the metaphysics will have to change, too. In this way, LR's naturalized metaphysics inherits the naturalistic humility of the science on which it rests.

Let us imagine that Nāgārjuna is, in fact, a world metaphysician. Call this hypothetical character world-Nāgārjuna. His conclusions about things lacking *svabhāva* and not really

[29] We are talking about the ideals of science, here, not actual scientists.

existing (or whatever) amount to discoveries about the true nature of reality (rather than the possibility of justifying a certain kind of judgment, as I've argued). Having found a time machine and some good 21st century academic connections, he gets the opportunity to sit down with James Ladyman to discuss their positions. Ladyman immediately claims that nothing world-Nāgārjuna says can be an acceptable criticism of his position: his *a priori* arguments proceed only on the analysis of domestic concepts like "cause" and intuitions as to what does and doesn't make sense. Ladyman says he will only countenance rival metaphysics which satisfy his naturalistic principles. How can world-Nāgārjuna respond?

In order to maintain that his position is a genuine ontological rival to Ladyman's, world-Nāgārjuna would need to defend his *a priori* methodology as a reliable way of gaining insight into the world. This means defending, in some form, a necessary intuition-world connection, if only to the extent that something which is blatantly absurd or contradictory cannot be the case.

Here is the problem: world-Nāgārjuna does not have a philosophical leg to stand on in defending such a connection. A necessary connection between intuitions and the nature of reality is exactly the kind of thing that Madhyamaka adamantly opposes. To justify it would require the defence of some kind of intuitive grasp on reality, which is what Mādhyamikas are so eager to deconstruct in the first place. Moreover, it is well established that Mādhyamikas cannot rely on belief in epistemic instruments which have as their intrinsic nature the ability to deliver reliable knowledge, (Siderits 2003, Chapter 7; Westerhoff 2009, Chapter 8), so what could world-Nāgārjuna even argue for? LR's strict standards for naturalised metaphysics expressed in their PNC rules out any theories which attempt much more than the unification of scientific hypotheses, which is not what world-Nāgārjuna wants to do. World-Nāgārjuna is therefore stumped: he can't independently defend his methodology against LR, and he can't simply adopt the standards which LR propose.

Perhaps a sympathetic person would embrace this result for world-Nāgārjuna: Nāgārjuna himself says clearly that the ultimate must be approached through the conventional (MMK24.10). In that case, it is unfair to expect Nāgārjuna to be able to defend his approach to the ultimate via absolutely any set of 'conventional' restrictions. On this view, that LR's restrictions don't allow world-Nāgārjuna to develop his philosophy doesn't mean that it fails any more than the fact that the conventions for baking recipes are also too restrictive for finding ultimate truth. But this is a feeble move. LR are not merely presenting another set of conventions that we can take or leave as we please, they are systematically arguing that the kind of methodology which Nāgārjuna relies on is *incapable* of leading us to how reality really is. As such, their position is directly contrary to world-Nāgārjuna's understanding of what it is he is doing. This objection cannot be dodged with relativism and, as we have seen, it cannot be countered in argument, either.

I believe that this very scenario – world-Nāgārjuna in conversation with James Ladyman (or Don Ross, or any of the many philosophers who agree with them on this issue) – demonstrates that Madhyamaka, understood as world metaphysics, simply fails. If it is supposed to be telling us something about how things are outside of our heads then that state of affairs should be certain no matter whom we talk to. The status of the external world presumably doesn't fluctuate based on the philosophical commitments of our conversation partners. If Nāgārjuna's position relies on adopting intellectual practices which he cannot defend and which only a subset of his opponents share, then it cannot be considered knowledge of the world at all.

It is now time to look at what Mādhyamika methods *can* do.

### 3.3. Reflective meta-physics and Nāgārjunian humility

It is difficult to say what exactly metaphysics is supposed to be. Van Inwagen et al. (2023) helpfully differentiate between the "old metaphysics" which was conceived of specifically as the study of being as such, first causes, and unchanging things, and the "new metaphysics" which has a more diverse subject matter (including modality, social categories, freedom of the will, etc.) and for which denying that any of its subject matter exists is not renouncing metaphysis but doing it. I will, to begin with, propose a sense in which the word "metaphysics" can be understood which I think is useful and captures a lot (not necessarily all) of the work that metaphysicians do. This will allow us to define a more limited space for LR's PNC (if one feels compelled to adopt it) while also helping to identify the sort of metaphysics to which it does not (should not) apply.

Consider the word "meta-metaphysics", which is what LR are doing in their chapter that we have looked at, and what I am doing presently. To do meta-metaphysics is to philosophise about metaphysics, in the same way that meta-ethics is philosophy concerning ethics and meta-philosophy is philosophy concerning philosophy. The prefix "meta-" in these cases therefore operates on the name of some philosophical discipline to give us a word which means 'philosophy about or concerning [that philosophical discipline]'. This is not the nature of the "meta" in "metaphysics"; that "meta" is an historical accident coming from the decision of an editor to categorise Aristotle's books concerning being as such, first causes, and changeless things as "*Ta meta ta phusika*" – "the ones after the physical ones" (van Inwagen et al. 2023). My suggestion is to deliberately reinterpret both the "meta" and the "physics" in "metaphysics" as follows:

| | |
|---|---|
| meta-: | "philosophy about or concerning…" |
| physics: | "claims about the world" |

I will signify that I am referring to this conception of metaphysics by using a hyphen: "meta-physics". 'Physics' in this very general sense encompasses any claim about the world, whether it is a child saying that there is a bug on the wall, a scientist offering a hypothesis, or a cleric saying that people have eternal souls. Meta-physics examines the nature, limitations, presumptions, and implications of such claims, to various degrees of generality and specificity. At least some of what is currently called meta-metaphysics will thus be included in meta-physics, but so will plenty of what is currently called metaphysics. To illustrate: LR's attack on the seriousness of the efforts of certain metaphysicians falls squarely into the domain of meta-metaphysics, but it is also clearly meta-physics. On the other hand, when LR tell us that propositions from extremely successful theories tell us about the objective modal structure of reality and never about the properties of fundamental objects with intrinsic nature, this is 'metaphysics' by any definition, not meta-metaphysics, but, since it is a philosophical interpretation of certain claims about the world (the original propositions), it is also meta-physics. We can call this latter example an example of 'world meta-physics': meta-physics which makes claims about the world in its capacity as meta-physics.

It should be clear from the above discussion that LR do not have a problem with meta-physics as such – how could anyone have a problem simply with doing philosophy about claims about the world? In the field of meta-physics, what they are opposed to is non-naturalistic *world* meta-physics, and their PNC is their standard for what qualifies as acceptable world meta-physics. It is obvious that they are not opposed to meta-metaphysics, so any meta-physics which falls within that category is unproblematic on assumption. What I have been building towards is that there is important philosophy which falls within the meta-physics umbrella, and which is neither world meta-physics nor meta-metaphysics. It's into this third category that I think Nāgārjuna's philosophy falls. I will illustrate with a couple of examples.

In a paper aiming for a Kantian interpretation of quantum mechanics, Michel Bitbol (2010) describes the style of metaphysics which Kant argues for as ‘reflective’. That is to say that Kant, in his meta-metaphysical writings, argued that the kind of metaphysics which ought to be done is that which Bitbol describes as reflective. In reflective metaphysics, metaphysical statements are seen as “rules in a grammatical pre-ordering of experience” as part of a discipline which is “nothing else than a reflective analysis of the powers and credence of reason” (Bitbol 2010, 59). Bitbol’s paper does not only label reflective metaphysics; it is itself an effort to do reflective metaphysics in the context of quantum mechanics. I won’t comment on that effort since the details are not relevant to the present discussion, but it should be noted that reflective metaphysics, in Bitbol’s view, is not meant to be limited to Kant’s own philosophy but is supposed to be an active programme of research. Bitbol’s reflective metaphysics is philosophy concerning claims about the world, so it is meta-physics, but it is not itself directed at the world nor the discipline of metaphysics itself. It is therefore neither world meta-physics nor meta-metaphysics, but meta-physics of a different sort.

Judging by his use of the adjective “Kant-like” to qualify reflective metaphysics several times in his paper, Bitbol does not mean to strictly identify his reflective metaphysics with precisely what was laid out by Kant. I will take the liberty of interpreting his “reflective” such that it is possible for there to be reflective metaphysics which is incompatible with Kant’s specific metaphysical ideas once it is in general agreement with the *orientation* that Kant proposed for metaphysics – that orientation being towards the formal, cognitive, or other conditions for a claim’s possible truth, meaningfulness, or other qualities. Meta-physics with such an orientation will thus be classified as reflective meta-physics. This is the category in which I think Nāgārjuna’s philosophy belongs.

We have noted that LR should have no quarrel with meta-physics generally, though they do have a problem with world meta-physics which does not adhere to their PNC. Have they any reason to object to reflective meta-physics? If they were to object, their objection would be that at least some of the subject matter which reflective meta-physics "orients" itself towards is properly investigated by science, not philosophy. For example, cognitive science should investigate the cognitive conditions for whatever qualities of claims we are interested in. I would agree insofar as this applies to claims such as Spackman's (2014) (discussed in Chapter One), which holds that Nāgārjuna is telling us about the properties of 'the concept of existence'. As I stated in Chapter One, this doesn't seem to me the domain of philosophers. But what Nāgārjuna (and Kant, I would say) was doing was not focused or reliant on rigorously establishing empirical facts about concepts, but on exploring the possibility of making rigorously true judgments using various concepts. I don't see how this is something that could be decided by science at all.

To show how important it is to recognize this style of metaphysics, especially in the present context, I will draw attention to a passage where LR failed to recognize it in another philosopher's work, resulting in an unfortunate case of undeserved criticism.

This is a quote from *Every Thing Must Go* (Ladyman and Ross 2007, 17):

> It is rare to find metaphysicians … arguing that if science and metaphysics seem to conflict the latter may trump the former, but here is a breathtaking declaration of philosophical arrogance from Peter Geach: '. . . "at the same time" belongs not to a special science but to logic. Our practical grasp of this logic is not to be called into question on account of recondite physics . . . A physicist who casts doubt upon it is sawing off the branch he sits upon' (1972, 304). Note that the 'recondite' physics we are being advised not to take metaphysically seriously is Special Relativity.

But in the relevant section of his work, Geach is at great pains to avoid being read this way.[30] To think that he was insisting that his metaphysics "trumps" science where they conflict is not a viable reading of him here. So that there can be no mistake about this, I will quote him at length:

> Simultaneity is involved in empirical statements; but it is not an empirical relation like neighbourhood in space. The natural expression for simultaneity is not a relative term like "simultaneous with", but a conjunction like "while" joining clauses; it is an accident of English idiom that "at the same time" seems to refer to a certain *time* that has to be *the same*, and the words for "at the same time" in other languages – Latin *simul*, Greek *ἅμα*, Polish *razem* – have no such suggestion.
>
> These conjunctions joining clauses no more stand for a proper relation than, for example "or" does. If I say I can see with my myopic eyes something over there that is *either* a hawk *or* a hand-saw, I do not claim to observe a hawk in the act of being an alternative to a hand-saw; to try to conceive a relation of alternativeness between such concrete objects would soon land us in paradoxes. Like alternativeness, simultaneity is not a relational concept, but is one of those concepts called transcendental by the medieval, formal in Wittgenstein's *Tractatus*, and topic-neutral by Ryle; the last term is the most informative of the three – it shows us that these concepts are not departmental but crop up in discourse generally.
>
> Because of this topic-neutrality, "at the same time" belongs not to a special science but to logic; its laws are logical laws, like the so-called De Morgan laws for "or". Physicists may have interesting things to tell us about the physical possibilities of synchronizing clocks by the transmission of electromagnetic signals; but this information is wholly irrelevant to the logic of basic simultaneity propositions. Our practical grasp of this logic is not to be called in question on account of recondite physics; for without such a practical

[30] It is on page 312 in my copy of Geach's 1972, not 304. And mine was published by Basil Blackwell in Oxford, not by Oxford University Press, as is listed in LR's bibliography. I have not been able to find an alternate version of this text published by Oxford University Press.

> grasp we could not understand even elementary propositions in physics, so a physicist who casts doubt upon it is sawing off the branch he sits upon. And a theoretical account of this logic must be given not by physicists but by logicians. (Geach 1972, 312. Italics in original)

I am not interested in defending Geach here, but his point is not that his logical intuition about simultaneity trumps contrary evidence from science. Rather, he is arguing that it would be misguided to think that the discovery that it is impossible to say in absolute terms that two events both happen at some time t means that the concept of simultaneity cannot be truthfully employed in judgments. He is arguing that propositions of the form 'P while Q' are grasped by the same type of logical intuition as those of the form 'P or Q', and that these do not represent real relations that can be empirically investigated by science in principle. Geach was trying to address what he saw as an error of equivocation in how people used "simultaneous" in logic and physics. According to Geach, there are two distinct types of judgment at work here which should not be confused. The point of rectifying this error is in helping people properly understand what is involved in the relevant judgments so that they can be properly investigated. This is reflective meta-physics. It is evident that the truth values of relevant propositions hinge on whether Geach is correct about this. I say that to point out that reflective meta-physics is an aspect of the pursuit of truth – the same goal that motivates naturalists in LR's worldview – it is not just idle play of the intellect. Again, one may disagree with Geach's position– it is not my mission to defend his conclusion – but to do so properly one must understand what he is attempting. Lacking this understanding, LR read him as they would someone doing world metaphysics – as if he were deciding between competing claims about the nature of reality.

Nāgārjuna's philosophy, as I interpret it, is reflective metaphysics. Its thesis – that the existence or nonexistence of any worldly thing cannot be established – is not a thesis about the world but about judgments as to the existence of worldly things. Its methodology is neither empirical nor speculative. Its analysis is meant to reveal the frailty of our conceptual scheme, which helps us attain the wisdom necessary to refrain from cleaving to it and projecting it onto reality uncritically.

### 3.4 Conclusion

I suggested earlier that naturalism as an ideal is wed to the humility implicit in the assumption that one can always be wrong about the way the world is. The naturalist always strives to check their every intuition, assumption, and rational conclusion, against empirical facts. I am suggesting now that there is humility implicit in Nāgārjuna's philosophical project as well. Nāgārjunian humility comes from the realization that our basic means of understanding the world are prone to being undermined upon reflection. We come away from it, if we are convinced, not taking our ontological intuitions very seriously. This is compatible with and complimentary to naturalistic humility.

This concludes this chapter, which is the last of three chapter focused primarily on explaining the nature of Nāgārjuna's philosophy. The next three chapters will be focused primarily on topics within the philosophy of science revolving around structural realism. If, according to Nāgārjuna, the existence or nonexistence of any worldly thing cannot be established, then what is science accomplishing, if not greater knowledge of existing worldly things? It is certainly accomplishing *something* spectacularly. The path to solving this problem will begin with an exegesis of Henri Poincaré's views in the next chapter.

CHAPTER FOUR

# Poincaré's Radical Ontology

I present an exegesis of Henri Poincaré's metaphysical position in three key essays within his book, *The Value of Science*. In doing so, I argue for three theses: First, that Poincaré's metaphysical position in these sources is incompatible with his metaphysical position in his earlier book, *Science and Hypothesis*. Second, that the phenomenological relationism defended by Poincaré in these sources is not a form of structural realism but a structuralist form of empiricism, and (by design) has no greater metaphysical commitments than constructive empiricism. Third, that Poincaré holds in these sources that the existence of the external world is merely a convention. An important figure in the history of science, Poincaré serves as an example of how an ontological conventionalist can give a sophisticated antirealist account of the success of science.

## 4.1 Introduction

In attempting to understand how Nāgārjuna's ontological conventionalism accounts for the success of science, it is instructive, in the first instance, to examine a philosopher of science with a similar conclusion who has addressed the issue of the success of science in their own work. Such a figure can act as a bridge between Nāgārjuna's ideas and the scientific realism literature, as well as a point of reference and departure as I develop my original proposal. I have found Henri Poincaré to be such a figure. His conventionalism has been extensively explored as an interpretation of principles in geometry and mechanics, but Poincaré, in certain key sources, extends his conventionalism to external objects as well. This chapter will be a careful reading of the key sources in which Poincaré explains his epistemological and metaphysical commitments. This will provide a working example of how a philosopher might account for the success of science even though existential sentences are merely conventions.

I must clarify the sense in which I am using some key terms in order to prevent some potential misunderstandings. When I refer to "Poincaré's metaphysics", I do not mean some description he has provided of the ultimate nature of reality: he very consciously refuses to provide any such thing. What I mean by "Poincaré's metaphysics" is the collection of his philosophical opinions about reality which are not strictly epistemological. For example, his important claim (explored in section 4) that "the ether is no less real than any external body" (1905/2001, 349) is not a description of reality, nor can it be reduced to a claim about the limits of human knowledge; there, as we will see, Poincaré is explaining what he thinks constitutes the meaningful designation of an external object. That is metaphysics; specifically, it is "reflective meta-physics" as I described it in the previous chapter. I will not continue to use the specialised vocabulary I developed there, however.

Next to be clarified is the term "conventional". This may retread some ground from Chapter One, but it is important to be thorough in this new context. For Poincaré, a convention is something expressed in the form of a proposition which is accepted or rejected by free choice. Though the senses or reason might inform what conventions we adopt, they can never ultimately compel us to accept one; the element of free choice in adopting conventions is ineliminable by definition. We adopt conventions for the sake of the *convenience* they provide: they make it easier for us to think or speak about something productively. When Poincaré claims that "the external world exists" is a convention, he is saying that we choose to accept it rather than its negation because accepting its negation would make it extremely inconvenient (though not impossible) to make systematic sense of our collective experiences.

Let us further develop the notion of a 'convention' as it is used in Poincaré's work. There are two critical questions which I will use to frame this effort. Firstly, do we have absolutely free choice in what conventions we can legitimately accept or is our choice constrained? Secondly, does judging a metaphysical statement to be a convention have only epistemic consequences, or does such a claim have ontological weight?

For the first question, consider this well-known quote of Poincaré's:

> *The geometrical axioms are therefore neither synthetic à priori intuitions nor experimental facts.* They are conventions. Our choice among all possible conventions is *guided* by experimental facts; but it remains free, and is only limited by the necessity of avoiding every contradiction, and thus it is that postulates may remain rigorously true even when the experimental laws which have determined their adoption are only approximate. (1902/2001, 45. Emphasis in original.)

Here, we see that the only strict limitation Poincaré places on our choice in conventions is non-contradiction. This, of course, means that we should not choose conventions which contain logical contradictions. But it also means that we need to avoid contradictions when we attempt to *apply* our conventions to anything. So, when we hope to use conventions, such as the geometrical axioms, in empirical science, our choice is limited to only those conventions which are consistent with the totality of empirical facts. Experimental facts then continue to "guide" us to choose which, from this limited set of conventions, are most convenient for our purposes.

It is important to understand that, when Poincaré writes "rigorously true" in the passage above, he does not mean this literally. Just a few words later he declares that the question "Is Euclidean geometry true?" has "no meaning". He writes that we "might as well ask if the metric system is true, and if the old weights and measures are false." By "rigorously true" he means that postulates are immune to empirical correction because they are *definitions* (in disguise). They are "remain rigorously true" in the same sense that it is always "rigorously true" that a 'gram' is a gram and a 'metre' a metre, no matter what happens in an experiment (Poincaré 1905/2001, 324). Poincaré extends this notion also to the principles of mechanics. For example, according to Poincaré, the laws of motion *define* what force is. No experiment could ever show that force is not the product of mass and acceleration because there is no clear notion of what 'force' means outside of that definition.[31]

This suggests an answer to the second question I posed above: this kind of convention is not the type that we choose to adopt because we do not *know* the ultimate truth of the matter. We choose to adopt these conventions because there *is* no ultimate truth of the matter. The matters they decide are not only stipulative by necessity – they are stipulative by nature. To adopt a convention in this sense is not to decide what is true but to make a determination on a matter to which truth and falsity do not apply. Ben-Menahem regards this as essential to conventionalism in philosophy: "To the conventionalist, the very idea of truth by convention is as incongruous as that of meaningful nonsense" (2006, 1). If it is in this sense of convention, then, that Poincaré claims that the existence of the external world is a convention, he is not claiming that we cannot know whether it exists or not, he is claiming that there is no ultimate truth of the matter of whether it exists or not. I will demonstrate that this is indeed the sense in which he means his claim (*pace* Zahar 1997 and Giedymin 1991). That is why his ontology, at least in the sources that I am investigating, deserves the title of 'radical'.

We must be careful here, however. Even if "the external world exists" is a convention, it does not seem like it could be a definition of anything. Because of that, it does not seem that the analogy between this proposition and the geometrical axioms or principles of mechanics

[31] See Ivanova (2015) for a detailed summary of Poincaré's conventionalism in geometry and the principles of mechanics. Ben-Menahem (2006) has an extended exegesis.

can go the distance. One might also wonder whether the ontological result I described is only attainable if we adopt peculiar rules about what counts as a truth-valued metaphysical claim, hence Zahar (1997, 215) argues that the geometrical conventionalist relies on a "positivist principle" that is "highly questionable"; namely that the empirical equivalence of various possible descriptions implies that there is no objective fact of the matter of which is accurate. It is this sort of rigidity that Putnam (1975) objected to, saying that it reduced conventionalism to dogmatic essentialism. In Section 4, I will argue that Poincaré's radical ontological conventionalism arises neither from the definitions-in-disguise move that we have seen, nor from the mere stipulation of any positivist principle describing what kind of claims count as truth-bearing. It arises rather from his strong neo-Kantian perspective on the nature of metaphysics and reality.[32]

Finally, it is critical for the reader to recognise from the outset that, in claiming that the existence of the external world is a convention, Poincaré is *not* claiming that the external world really does not exist. It should be clear from what I have said already that if some proposition is a convention, then its negation *is also a convention*. We could not have the choice to accept or reject *p* if we did not have the choice to accept or reject not-*p*. For Poincaré, "the external world exists" being a convention does not mean that the negation of that proposition is true; the situation is quite the opposite: its negation *cannot* be ultimately true.[33] This being a conventional matter means that we have the freedom, *independent of questions of truth*, to stipulate whether or not the external world exists. And, contrary to the idea that he is trying to convince us that the external world doesn't exist, Poincaré treats it as utterly obvious that he

[32] It might be argued that he does uphold the positivist principle *because* of his neo-Kantian metaphysics, but in that case he is being neither an essentialist nor obviously dogmatic.

[33] Carnap (1950, p.33) makes a similar point about the Vienna Circle's rejection of "the thesis of the reality of the external world" and other metaphysical theses as pseudo-statements. "It is obvious that the apparent negation of a pseudo-statement must also be a pseudo-statement." This is not to say that conventions and pseudo-statements are the same, but they are at least alike in this respect.

and his readers will accept the convention "the external world exists" and not even consider accepting its negation. Poincaré is not going for some kind of Berkeleyan idealism or nihilism here; he is saying, rather, that these kinds of ontological determinations can only be made as conventions. Truth and falsity don't ultimately apply to them.

I am focusing on three primary sources: the introduction to Poincaré's 1905 book *The Value of Science* and the final two essays in that book, entitled "Is Science Artificial?" and "Science and Reality". These final two essays were developed from a single 1902 paper of Poincaré's (see de Paz 2021, 453). Poincaré's chief goal in this pair of essays is to distance himself from the radical conventionalism that he associates with Edouard LeRoy. This means he has maximum incentive in these essays to explain precisely his commitments pertaining to the nature of reality and the limits of his conventionalism. We should expect that he chose his words very carefully in these sources: he would not have wanted his readers to think that he had ontological commitments which he did not in fact have, but it would have been contrary to his purpose to exaggerate his scepticism. These two essays are therefore arguably where Poincaré pays the greatest attention to the metaphysical implications of his conventionalism.[34] The introduction to *The Value of Science* will also be an important source because Poincaré presents there a summary of his views in helpfully direct language. I will refer to these three sources collectively as 'the VoS sources'. I will only reference Poincaré's other writings for context or support.

This background might seem somewhat bewildering: if I am correct in claiming that Poincaré, in the identified sources, is claiming that the existence of the external world is a convention, how could his goal in these sources be to show that his conventionalism is *less* radical than someone else's? The short answer is that Poincaré thought that science accurately

[34] Ivanova also highlights them as two particularly important essays for this subject (2015, p.114, footnote 1).

represents empirical facts as they are experienced (albeit with conventional labels and categories) while LeRoy thought that empirical scientific facts were conventional creations which did not correspond with experience as it genuinely is, free from conceptual engineering. The specifics of Poincaré's view will be clarified in this paper, but I will not explain LeRoy's view further. For a more detailed comparison of the two thinkers, see de Paz (2021).

In the next section, I will argue that the interpretation of Poincaré as a structural realist which began with John Worrall in his seminal 1989 paper, while appropriate for his position in *Science and Hypothesis*, cannot be maintained for his position in the VoS sources. His metaphysical opinions in these two (collections of) sources are incompatible. In section 3, I will show that Poincaré thought that the only things which are truly objective are the relations between sense impressions. He thought that science, using certain conventions, enabled us to see a profound harmony among these objective phenomenological relations. I will argue that his relationism in the VoS sources is not a form of structural realism, but is a structuralist form of empiricism – i.e., it is meant to establish the uniquely objective status of observable facts, not to establish that we have any knowledge of the unobservable world. In section 4, I will explain how Poincaré extends his conventionalism to all external objects – observable and otherwise. Section 5 will raise a difficulty for Poincaré's view and suggest avenues through which it can be solved.

### 4.2 Poincaré and epistemic structural realism

The first thing to note is that the common labelling of Poincaré as an epistemic structural realist is inappropriate to his position in the VoS sources. Epistemic structural realism (ESR), as defined by Ladyman (1998) to categorise the kind of structural realism advocated by Worrall (1989), holds that science teaches us the true relations among the real objects that ultimately constitute reality. It is true that a direct quote from Poincaré unambiguously describes this idea and was in fact what inspired it. Quoted often in the scientific realism literature, this passage is from his essay "The Theories of Modern Physics" found in the book *Science and Hypothesis,* published in 1902:

> It cannot be said that this is reducing physical theories to simple practical recipes; these equations express relations, and if the equations remain true, it is because the relations preserve their reality. They teach us now, as they did then, that there is such and such a relation between this thing and that; only, the something which we then called motion, we now call electric current. But these are merely names of the images we substituted for the real objects which Nature will hide forever from our eyes. The true relations between these real objects are the only reality we can attain… (1902/2001, 123)

But the idea expressed in this quote is irreconcilable in important respects with the ideas expressed in the VoS sources. This represents a genuine discontinuity in Poincaré's thought.

Herein lies the conflict: the quote from "The Theories" commits Poincaré to the existence of (i) real objects, distinct in nature from the mere "images" conceived of in scientific theories, which are inaccessible to us and stand in distinct relations with one another, and (ii) that certain mathematical relations in science faithfully represent relations that these external objects have with one another. I will show that both of these commitments are irreconcilable with the VoS sources. The case of (i) will take some work to explain and only be settled in Section 4. The case of (ii), however, can be settled from a single quote from a VoS source. The following is taken from the final paragraph of the Introduction – the climax of Poincaré's summary of his metaphysical commitments:

> Does the harmony the human intelligence thinks it discovers in nature exist outside of this intelligence? No, beyond doubt, a reality completely independent of the mind which conceives it, sees or feels it, is an impossibility. A world as exterior as that, even if it existed, would for us be forever inaccessible. (1905/2001, 193)

This "harmony" is identified as "the harmony expressed by mathematical laws". Here, Poincaré is stating strongly and unambiguously that the mathematical relations he treasures so much should not be interpreted as mind-independent. They are not (or, at least, we could never know that they are) the relations which hold between external, hidden objects. As we examine the VoS sources throughout this paper, it will become clear that Poincaré's insistence on this point is not ad hoc. It is a natural component of a metaphysics which takes scepticism as a guiding methodology. There is a genuine contradiction between the two quotes above which could only be avoided by reading one of them non-literally. The upshot is that Poincaré's view in the VoS sources is significantly less realist than his view in "The Theories" and therefore significantly less realist than standard ESR. Poincaré does claim that the "harmony" mentioned above is objective, but he clarifies that "what we call objective reality is, in the last analysis, what is common to many thinking beings, and could be common to all" (1905/2001, 193). Throughout the VoS sources, Poincaré maintains that "objectivity" is supposed to be understood as meaning intersubjective stability and nothing more.

It may be tempting to favour other interpretations of Poincaré to preserve the unity of his corpus, but there is an hermeneutic responsibility to avoid ignoring or dismissing things an author has written in order to preserve an impression of unity across their work. Though it is, of course, fine to suggest non-literal readings of passages here and there to preserve consistency, I submit that the evidence from the VoS sources which I will present in this paper cannot so easily be made consistent with the famous 'structural realist' quote from "The

Theories of Modern Physics". I will not speculate here on why this discrepancy exists. The important thing is to recognise that it does so that the full complexity of Poincaré's philosophical thought is not lost to us.

Milena Ivanova (2015) and Katherine Brading and Elise Crull (2017) have recently drawn from the VoS sources to argue that Poincaré's structuralism should be understood as distinct from ESR. But they do not acknowledge the incompatibility *within* Poincaré's corpus that I have just highlighted. Ivanova argues that Poincaré's "structuralism is internalist and deeply entrenched into his neo-Kantianism and conventionalism" (2015, 121) which makes it a departure from "the view of contemporary structural realists who claim that our best theories can successfully track the structure [of] mind-independent reality" (120). That is correct, but Ivanova's study leaves room for clarification on the metaphysical character of Poincaré's position. She touches on a contrast between realism and mere empirical adequacy (115; 120) but does not ultimately say whether or to what degree Poincaré's position holds that any scientific theories are more than just empirically adequate. In the present study, I will argue that Poincaré (in the VoS sources) does not make any claims about the metaphysical status of scientific theories which go beyond their empirical adequacy. Further, Ivanova claims that Poincaré's conventionalism "is restricted to geometry and some physical constitutive principles only" (p.120) but I will show that (in the VoS sources) he regards the existence of every object in his *ontology* – both observable and unobservable – as conventionally stipulated.

Brading and Crull draw on reconstructed systematic arguments to show that Poincaré's structuralism had different properties to ESR. Of particular interest for us is what they call the argument from objectivity (Brading and Crull 2017, 121):

1. Objective knowledge is necessarily common to all (i.e., intersubjectively stable).

2. That which is common to all is necessarily communicable.

3. That which is communicable is knowledge of relations only.

4. Therefore, objective knowledge just is knowledge of relations.

They reconstruct this argument from the VoS sources, and I agree that this is what is going on in the relevant passages (Poincaré 1905/2001, 345-6). Their point is that, while ESR as introduced by Worrall (1989) is motivated specifically as a means of upholding the no miracles argument in the face of the pessimistic meta-induction, Poincaré's structuralism is derived from his philosophical principles regarding the nature and limits of knowledge. Though Poincaré does discuss the success of science and the issue of theory change, his case for all objective knowledge being relational does not rely on these historical facts. Brading and Crull are right to argue that this tells us that Poincaré's position is different in character to ESR. My mission here is to further interrogate the character of Poincaré's relationism in order to explain the role it plays in his ontology. What is the metaphysical status of these objective relations? What of their relata?

### 4.3 Poincaré's phenomenological relationism

Though they can be interchangeable under some interpretations, I favour the word 'relationism' over 'structuralism' for Poincaré's view.[35] The distinction between relations and structures put to use, notably, by Russell and Carnap is that structures are the abstract patterns formed by relations. A 'pure' structure provides no information whatsoever about the nature of the relations which form it, so multiple otherwise unrelated systems of relations can instantiate the same structure. For example, the relation *is greater than* forms the same structure in the set

[35] Psillos (2014) also opts for this term.

{1,2,3} as does the relation *is bigger than* in the set {The Eiffel Tower, Australia, Jupiter}. Because some mathematical objects and physical systems will share structural properties without instantiating similar intrinsic properties or relations, structure provides fertile ground for the formal analysis of diverse classes of entities. And even if the nature of an entity is unknown or unknowable, its structure might be known. For Russell and Carnap, objectivity was to be found in structure, not in relations. But Poincaré, as we will see, took *relations* of a certain sort as being most truly objective, not structure. He was criticized by Carnap for this (1928/2003, 30), but, as I will argue below, Carnap's objection was not sound. I mention this now only to show that there is some weight to the relation/structure distinction and 'relationism' is the more appropriate label for Poincaré.

It has already been indicated, through Brading and Crull's reconstructed argument, that Poincaré's relationism is motived by his desire to identify what in reality should be taken as objective. Let us take a closer look at how Poincaré understands objectivity and the connection he sees between it and relationality. In "Science and Reality", he writes:

> What guarantees the objectivity of the world in which we live is that this world is common to us with other thinking beings. Through the communications that we have with other men, we receive from them ready-made reasonings; we know that that these reasonings do not come from us and at the same time we recognize in them the work of reasonable beings like ourselves. And as these reasonings appear to fit the world of our sensations, we think we may infer that these reasonable beings have seen the same thing as we; thus it is we know we have not been dreaming.
>
> Such, therefore, is the first condition of objectivity; what is objective must be common to many minds and consequently transmissible from one to the

> other…we are even forced to conclude: no discourse, no objectivity. (1905/2001, 345)

Notice first that in this passage 'objectivity' has not yet come to be equivalent with intersubjectivity for Poincaré. This is him establishing that connection. For the sake of this passage, 'what guarantees the objectivity of the world?' has exactly the meaning of 'what guarantees that the world is not illusory?'. Poincaré's answer is that we can be sure that there are other beings around us having similar sensations and thoughts, so we can conclude that this isn't all make-believe. Poincaré does not venture to explain *why* beings see the same things as one another. He accepts that it is true that they do and that this guarantees that those things we have in common are not made up.

What defines Poincaré's relationism is that he does not use this intersubjectivity argument to infer that we ought to trust the content of our senses. On the contrary, he points out that the qualitative content of our senses is *in*transmissible and we therefore *cannot* trust it. It is this fairly hardcore sceptical intuition that leads Poincaré to look for truth exclusively in information that is shared between persons. That is, his scepticism will not allow him to be confident that anything which cannot be *confirmed* as intersubjectively stable is not illusory. This is why he pivots from sensations to the relations between sensations:

> We have no means of verifying that the sensation I call red is the same as that which my neighbour calls red. …In compensation, what we shall be able to ascertain is that, for him as for me, the cherry and the red poppy produce the *same* sensation, since he gives the same name to the sensations he feels and I do the same.

> Sensations are therefore intransmissible, or rather all that is pure quality in them is intransmissible and forever impenetrable. But it is not the same with relations between these sensations. (1905/2001, 345. Emphasis in original.)

We can confirm that red poppies and cherries stimulate the same sensation in multiple people even though we can't confirm what "red" looks like to each of them. So, the relation of *sameness* of the sensations produced upon seeing various red things is objective for Poincaré, even though the qualitative content of those sensations is not. Note that what is being transmitted are the relations experienced privately and individually by multiple people. We are not talking about relations between the sensations of different people. Poincaré continues:

> From this point of view, all that is objective is devoid of all quality and is only pure relation. Certes, I shall not go so far as to say that objectivity is only pure quantity (this would be to particularize too far the nature of the relations in question), but we understand how someone could have been carried away into saying that the world is only a differential equation.
>
> With due reserve regarding this paradoxical proposition, we must nevertheless admit that nothing is objective which is not transmissible, and consequently that the relations between the sensations can alone have an objective value. (1905/2001, 346)

In this way, Poincaré derives from his "first condition of objectivity" – intersubjectivity – the requirement of relationality. Let us unpack this latest quote. First of all, it would seem that Poincaré knew of a person espousing a view that sounds something like an anticipation of ontic structural realism (OSR) (see Ladyman 1998).[36] True, it seems that Poincaré might be

[36] In the French, it is clearer that Poincaré is here writing about a specific person whose name he apparently doesn't remember. "Je n'irai certes pas jusqu'à dire que l'objectivité ne soit que quantité pure (ce serait trop particulariser la nature des relations en question), mais on comprend que je ne sais plus qui se soit laissé

caricaturing the idea he is recalling but, so far as caricatures go, it is not an inaccurate summary of the form of OSR defended by Tegmark (2008), who defends what he calls the Mathematical Universe Hypothesis: "Our external physical reality is a mathematical structure." Other OSRists are a bit harder to pin down on this matter. Ladyman and Ross say that the question of what makes the world's structure physical and not mathematical is one which they "refuse to answer" (2007, 158). French has a nuanced discussion of the question spanning an entire chapter (2014, Chapter 8). In any case, it is clear that Poincaré wanted to distance himself from any metaphysical position where the nature of the world and the nature of mathematics become so difficult to separate. Judging by his language, he sees this as symptomatic of a lack of restraint: a person who equates the nature of the world with mathematics has, having made a genuinely profound realisation, become too eager and jumped to a poorly motivated declaration. Poincaré is saying here that, while his views do lend themselves to this kind of view, it should be rejected.

Now, the key question we ought to direct at Poincaré here is: if "all quality" is being denied to "pure relation", then what is left of the "nature of the relations in question" beyond the quantitative? Unfortunately, he did not elaborate on this point. However, a satisfying answer is forthcoming when we pay attention to the final line of this latest quote – to Poincaré's claim that "the relations between the sensations can alone have an objective value." This claim is stronger and more important than one might initially recognise from reading the passage in context. Poincaré's position is not merely that the only objective value *in sensations* is the relations between them. His position is that the relations between *sensations specifically* are the only parts of reality that are ultimately objective. This is so important to him that he presents it as his second condition of objectivity:

entraîner à dire que le monde n'est qu'une équation différentielle" (Poincaré 2013, 320). The "je ne sais *plus*" instead of "je ne sais *pas*" suggests that at one time he would have been able to say whom he was talking about.

> An absolutely disordered aggregate could not have objective value since it would be unintelligible, but no more can a well-ordered assemblage have it, *if it does not correspond to sensations really experienced*. … Two conditions are therefore to be fulfilled, and if the first separates reality from the dream,[37] the second distinguishes it from the romance. (1905/2001, 346-7. Emphasis added.)

As a consequence of this, we know that Poincaré does not regard any unobservable aspect of scientific theories (relational or otherwise) as objective. An important passage for helping us understand this point appears towards the end of "Is Science Artificial?":

> Since the enunciation of our laws may vary with the conventions that we adopt, since these conventions may modify even the natural relations of these laws, is there in the manifold of these laws something independent of these conventions and which may, so to speak, play the role of *universal invariant*? (1905/2001, 334)

At the very end of the essay, he answers confidently:

> What now is the nature of this invariant it is easy to understand, and a word will suffice us. The invariant laws are the relations between the crude facts, while the relations between the "scientific facts" remain always dependent on certain conventions. (1905/2001, 336)

As I explained in the introduction, a convention, for Poincaré is something expressed in the form of a proposition which is accepted or rejected by choice. Note well that Poincaré claims that relations between "scientific facts" are *not* intersubjectively invariant, due to their dependence on conventional definitions. What are truly invariant are the relations between

[37] Here, Poincaré clarifies in a footnote: "I here use the word *real* as a synonym of *objective*; I thus conform to common usage; perhaps I am wrong, our dreams are real, but they are not objective." Emphasis in original.

"crude facts". To understand him properly here, we need to understand what the word "fact" meant for him, and then the distinction between crude and scientific facts.

Jerzy Giedymin provides this context for talk about facts in the relevant period:

> The traditional empiricist and Comtian positivist philosophy of science, perpetuated by numerous textbooks on the subject, contrasted "positive science" consisting of facts, with theories and hypotheses. Facts were claimed to be objective, theory-independent and simply discovered, stated and collected by impartial observers. (Giedymin 1982, 199)

In "Is Science Artificial?" Poincaré is positioned as a defender of this orthodoxy (more or less) from attack by Edouard LeRoy, who claimed that scientific facts were conventional through and through. Though Poincaré is forthcoming about being sympathetic to much of what LeRoy says (1905/2001, 318), he is adamant that facts revealed by our senses are not created by us. Poincaré then develops a technical definition of the word 'fact' tailored to its use in science. Instead of the general meaning of 'something that is the case', his sense of 'fact' is limited to what is given in observation.[38] The following passage gives us a lot of information about Poincaré's use of language in this matter:

> When I say: it grows dark, that well expresses the impressions I feel in being present at an eclipse; but even in obscurity a multitude of shades could be imagined, and if, instead of that actually realized, had happened a slightly different shade, yet I should still have enunciated this *other* fact by saying: It grows dark.

[38] My own use of the word 'fact' will not be limited to this technical sense. I will use either sense where it is appropriate and I believe that my intent is clear.

> Second remark: even at the second stage, the enunciation of a fact can only be *true or false*. This is not so of any proposition; if this proposition is the enunciation of a convention, it cannot be said that this enunciation is *true*, in the proper sense of the word, since it could not be true apart from me and is true only because I wish it to be. (1905/2001, 324. Emphasis in original.)

From this we know, firstly, that facts are identified with the content of a person's senses; every miniscule difference in shade experienced is a *distinct* fact that might be enunciated by the proposition "It grows dark". A "proposition", for him, is a linguistic construction that may be used to enunciate a fact or a convention. Propositions which enunciate facts can be true or false; technically, on Poincaré's formulation, when we say that a fact is true, we mean that the proposition which *enunciates* a fact is true, since the fact itself is a sense impression and not linguistic. Propositions which enunciate conventions are neither ultimately true or false (as discussed in Section 1).

Now for the difference between crude and scientific facts. Crude facts are the 'pure', un-enunciated sense impressions that comprise facticity. They are extra-linguistic and theory-independent experiences. Poincaré explains that in order to think and converse about crude facts, we need to introduce conventional systems (i.e., systems at least partly constructed according to free choices) like language and mathematics to represent them. It is those representations that are defined as scientific facts.

> [I]t is necessary to distinguish between the impression of obscurity felt by one witnessing an eclipse, and the affirmation; it grows dark, which this impression extorts from him. In a sense it is the first which is the only true fact in the rough, and the second is already a sort of scientific fact. (1905/2001, p.323)

The distinction between crude and scientific facts, then, is the difference between pure sensation and the enunciation of the content of sensations. On this view, the only difference between, say, a child reporting a fact to their sibling and a scientist reporting a fact to their colleague is that the scientist will use technical conventions tailor-made to be convenient for their purposes. Both reports fall into the broad category of scientific fact rather than crude fact. But note that when a child and scientist both witness some event, the crude facts that they experience will be in strict harmony with one another (for Poincaré, that is what it means for them to have witnessed the 'same' event) even if the conventions they use to describe the event are radically different. It's this idea that there is agreement between crude facts which is preserved through any empirically adequate use of conventions that makes Poincaré's view so much more realist than LeRoy's. While LeRoy thought that scientific facts were virtually independent of crude facts, Poincaré insisted that "the scientific fact is only the crude fact translated into a convenient language" (1905/2001, 327).

So, when Poincaré declares that "[t]he invariant laws are the relations between the crude facts, while the relations between the "scientific facts" remain always dependent on certain conventions", he is saying that the true invariant relations which form the beating heart of his metaphysics and epistemology exist between *sense impressions alone*. They are phenomenological in nature, not mathematical or linguistic. Mathematics and other forms of theorising are just ways of representing them so that we can reason about and discuss them in sensible, productive ways.

Let us recap briefly. According to Poincaré:

- Something is objective insofar as it is intersubjectively invariant.
- Scientific laws are constructed using conventions, which prevents them from being universally invariant.

- Only the relations between sense impressions themselves (crude facts) are truly objective.

We may now return to the question I asked earlier: if Poincaré is saying that objective relations are devoid of quality, how can he also say that they are not only quantitative in nature? We have now seen that Poincaré sees ultimate objectivity in phenomenological relations alone. This helps us understand what he must have meant. It would be odd to say that the relation of *sameness* between multiple sightings of red objects is "pure quantity", even though we acknowledge that the qualitative sensation of redness is utterly absent from that relation. The relation, devoid of quality though it is, is still phenomenological in nature. We are not forced to treat it as an entirely uninterpreted instance of some logico-mathematical notion of sameness. Part of understanding and accepting the existence of the relation is to understand, in at least some way, what it is to experience a sensation. Though 'universal invariant' is a predominantly mathematical concept, Poincaré, in the relevant discussion, identifies an 'invariant' as that by virtue of which different languages are capable of intertranslation (1905/2001, 336). In that passage, he makes clear that the 'universal' in 'universal invariant' means universal among beings who have similar sense faculties. He is saying that, among such a set of beings with common faculties, their scientific facts will always be intertranslatable precisely because the relations between their crude facts will be "invariant laws". To say that objectivity is purely quantitative therefore misses much. The objective relations between sensations are relations between *sensations*, not abstract relations.

Now it is time to address Carnap's objection. Carnap quotes Poincaré as saying that "only the relations between the sensations have an objective value" and proceeds to make these comments on Poincaré's view:

> This obviously is a move in the right direction, but does not go far enough. From the relations, we must go on to the structures of relations if we want to reach totally formalized entities. Relations themselves, in their qualitative peculiarity, are not intersubjectively communicable. It was not until Russell … that the importance of structure for the achievement of objectivity was pointed out. (Carnap 1928/2003, 30)

Carnap flatly denies here that phenomenological relations can be purged of their qualitative content in order to be intersubjectively communicable. For him, objectivity can only exist in structure (according to the definition that I explained at the beginning of this section). But Carnap was wrong about this. Poincaré's argument (which Carnap doesn't engage with) is meant precisely to demonstrate that a person can communicate to me that (according to them) poppies and cherries are the same colour and that I can verify that (according to me) poppies and cherries are the same colour, without either of us needing access to the "qualitative peculiarity" of the other's experience of poppies or cherries. What is needed is that we both know what it means to experience something visually and we can understand the referents of "poppies" and "cherries" using description and ostension. With those tools – and if we are both sighted people able to perceive colours – we can verify that these things are the same colour for each of us even if we have no idea what they *look like* for the other person. For Carnap to prevail here, he would have to convince us that people not only cannot communicate between each other what something looks like, but that they cannot communicate between each other the fact that they are looking at something – i.e., experiencing a sensation. This does not seem plausible. In Poincaré's example, the objective relation is explicitly a relation between certain phenomenological properties, and this is something that people can effectively relay to one another.

Carnap seems to have conflated the idea of "totally formalized" with "intersubjectively communicable". It is true that the same-colour relation is not totally formalized: it requires familiarity with the phenomenological nature of colour. But this does not mean that the relation is incommunicable without access to the "qualitative peculiarity" of a person's experience of colour. Thus, in pursuit of logical formalization, Carnap has glossed over the distinction between the *content* of a sense impression and the *nature* of a sense impression as sensory. Neither of these are formalizable, but only the former is incommunicable.

Poincaré's phenomenological relationism could be described as a relationist or structuralist form of empiricism (cf. Bueno 1999). I am not claiming that Poincaré was arguing from an empiricist stance generally. His stance is neo-Kantian and to give him the title 'empiricist' would obscure more than it would clarify. My claim is rather that the point of Poincaré's *relationism* specifically, in the VoS sources, is to defend the legitimacy and objectivity of observable scientific fact against attack from LeRoy, *not* to identify the knowledge we have of the noumenal or otherwise unobservable world. And, in pursuit of that goal, Poincaré rules out the possibility of objective knowledge of the noumenal or otherwise unobservable world. As we saw above, he goes so far as to claim that correspondence with "sensations really experienced" is a condition of objectivity. I find it instructive to compare this with the well-known slogan of "saving the phenomena" in van Fraassen's (1980) constructive empiricism. While van Fraassen wants to "save the phenomena" by embedding them (that is, embedding their *structure*)[39] within the empirical substructures of theoretical models, Poincaré wants to "save the phenomenal relations" by embedding them within convention-dependent mathematical systems which enable us to see the systematic harmony between them. So, while

[39] "To be matched are two models, a data model and a theoretical model. The matching in question may be as simple as an embedding or partial isomorphism, or it may need to be some measure-theoretic refinement thereof to allow for approximation. But is in any case a mathematical relationship, and therefore purely structural. The claim of adequacy is in the first instance a claim about how two structures are structurally related." (van Fraassen 1997, p.524)

Poincaré and van Fraassen have different philosophical orientations in several ways, they are alike in regarding the mathematical structure of scientific theories as fully in service of our knowledge of the connections between phenomena. This is a noteworthy observation because existing literature on Poincaré would have him as van Fraassen's opponent on this very issue.

That Poincaré's relationism was concerned with phenomena, rather than unobservable or mind-independent reality, is directly relevant to a contemporary debate. John Worrall and Elie Zahar, both inspired by Poincaré, have defended the use of Ramsey-sentences to formulate structural realism (appendix IV to Zahar 2001; Worrall 2007; Zahar 2007). They do so against an objection to realist interpretations of Ramsey-sentences advanced by Demopoulos and Friedman (1985). This objection, which is an adaptation of Newman's (1928) objection to Russell's structuralism, charges that, for any theory T, the truth of T's Ramsey-sentence T* tells us nothing other than that T's purely observational consequences are true (Demopoulos and Friedman 1985, 635). Ketland (2004) further developed and defended the objection, arguing that T* is equivalent to T's empirical adequacy plus a cardinality constraint on its domain. Worrall and Zahar, who consider themselves realists, don't recognise this as an appropriate objection to their position. They write that "on the one hand, the Ramsey-sentence does not in general logically follow from its empirical basis" and "[o]n the other hand, if the so-called empirical generalisations…were to be included in the observational content of a theory, then the Ramsey-sentence might well be one of them; in which case Demopoulos's and Friedman's thesis would collapse into the trivial claim that the Ramsey-sentence follows from itself" (Zahar 2001, 240). They therefore allow that their approach is focused on the structural unification of the empirical world, not any specific claims about what the unobservable world is like. The crucial issue, for them, is not whether a true Ramsey-sentence represents the structure of some specific unobservable system, but whether it tells us something about the world that is not reducible to the atomic sentences of empirical fact which constitute its "empirical basis". Worrall suggests that the critics' discomfort comes from thinking that this structuralism cannot count as "really" realism. He urges that it ought to qualify because it takes the mathematical structure of theories to "globally reflect" the ultimate structure of reality (Worrall 2007, 153-4).

This approach of Worrall's and Zahar's is clearly not far removed from Poincaré's. But the more realist elements to do with "reflecting" mind-independent reality, while consistent with Poincaré's words in "The Theories of Modern Physics", are rejected in the VoS sources. Poincaré, in the VoS sources, thought that science taught us about reality by revealing profound harmony among phenomenological relations – a harmony that is not conventional (though we need conventions to represent it) nor dependent on any *particular* mind or group of minds. But he emphatically did not think that science taught us the ultimate mind-independent structure of 'the world' as such. The objection advanced by Demopoulos, Friedman, and Ketland is of no threat to Poincaré (in the VoS sources) because he is not committed to anything beyond the empirical adequacy of theories. Poincaré's point is that the fact that *there exist* theories which are remarkably (never entirely) empirically adequate, and equally so for everyone (when certain conventions are understood), itself demonstrates science's objective value. That so much of the world we collectively experience can be synthesised into unified mathematical structures is what is important for Poincaré in these sources – he consciously refuses to offer a metaphysical interpretation of those structures.

We have now covered the essentials of Poincaré's relationism and seen that it is only committed to the existence of sensory experiences – our own and others' – and stable relations between them. But what about objects? I have noted some similarity between Poincaré's relationism and constructive empiricism, and van Fraassen regards the unqualified existence of observable objects as philosophically unproblematic. The reader may also have noticed that the example of a phenomenological relation Poincaré uses – red poppies and cherries being the same colour – is naturally stated as a relation between objects, not a relation between subjective experiences. Does Poincaré maintain that the objective relation in question is between poppies and cherries, thereby committing himself to the objective existence of those objects? Or is he only ultimately committed to the relation between the *appearance* of poppies and the appearance of cherries in a person's consciousness? I will show in the next section that the latter is true. According to Poincaré in the VoS sources, the existence (or nonexistence) of any and all external objects is only a convention.

### 4.4 Poincaré's ontological conventionalism

Poincaré's initial and most influential use of conventionalism in the philosophy of science was in arguing for the conventional status of geometrical axioms. His most radical application of conventions, however, is undoubtedly in his ontology. Let's begin this section with Poincaré's understanding of objecthood, in "Science and Reality":

> External objects, for instance, for which the word *object* was invented, are really *objects* and not fleeting and fugitive appearances, because they are not only groups of sensations, but groups cemented by a constant bond. It is this bond, and this bond alone, which is the object in itself, and this bond is a relation. (1905/2001, 347. Emphasis in original.)

The reader will note that the existence of the bond to which Poincaré refers as "the object in itself" is not a convention. According to him, we do not have a choice as to whether such bonds, which are identified as objects, exist. But also note that he is explicitly referring to bonds between *sensations*. For him, the 'real' object – the object that we can say exists without qualification – is a relation between sensations exclusively. His conventionalism will come into play when we ask the separate question of what *metaphysical interpretation* we ought to give to these bonds between sensations. To this new question, his answer is that such an interpretation is only to be given as a conventional stipulation. A few pages after the quote above, Poincaré illustrates his intent by invoking one of science's canonical non-existents – the luminiferous ether.

> It may be said, for instance, that the ether is no less real than any external body; to say this body exists is to say there is between the colour of this body, its taste, its smell, an intimate bond, solid and persistent; to say the ether exists is to say there is a natural kinship between all the optical phenomena, and neither of the two propositions has less value than the other. (1905/2001, 349-50)

According to Poincaré, no object has any more reality than the luminiferous ether. The ether was a convenient way to think about optical relations when these relations were understood as mechanical. It is not that anyone discovered, factually, that there was no such object with the rise of electromagnetic field theory, but rather that this object-idea ceased to be convenient when our story about the objective relations in question changed. We could have continued to refer to the ether as the representation of the bond between optical phenomena if we wanted to (this has been argued by Wayne Myrvold (2020) – which is not to say that he shares Poincaré's metaphysics), and in all this the ether would be no less real than any object we think of as inhabiting the external world. Poincaré continues:

> And the scientific syntheses [like the ether] have in a sense even more reality than those of the ordinary senses [like tables], since they embrace more terms and tend to absorb in them the partial syntheses. (1905/2001, p.350)

Poincaré pushes his point and argues that one could consider unobservable theoretical entities *more* real, in a sense, than ordinary observable objects. This is because the theoretical structure of those objects encompasses a greater array of phenomenological relations than does the structure of ordinary objects. While the idea of a particular table represents the bond between certain colours, shapes, and tactile sensations, the idea of the ether represents a truly vast array of optical phenomena. By "partial syntheses" I think Poincaré means phenomenological "bonds" that are not "solid and persistent" but conditional, such as the various ways in which light is refracted in different conditions. Theoretical objects "absorb" these in the sense of accounting for and explaining their conditionality. This perspective of Poincaré's is, of course, a full reversal of the standard direction of ontological scepticism. Typically in the philosophy of science, especially in contemporary times, the reality of observable objects is taken for granted and it is the theoretical objects whose reality is questioned. Poincaré turns this on its head when combining his thorough conventionalism with his relationism.

Though Poincaré's use of conventions in ontology is only developed in the VoS sources, its radical nature is anticipated in *Science and Hypothesis* despite his apparent belief, there, in real, mind-independent objects. I have in mind his claim that "these two propositions, 'the earth turns rounds,' and 'it is more convenient to suppose that the earth turns round,' have one and the same meaning. There is nothing more in one than in the other" (1902/2001, 91). In other words, the rotation of the Earth is nothing more than a convention. Poincaré defends the idea that the Earth's rotation is not an empirical fact by denying that there is absolute space. Without absolute space, any observation that would seem to confirm that the planet rotates could be interpreted such that it is not rotating with no strict loss of truth. The claim that the Earth's rotation is a convention, while striking, is thus based in scientific theory and is not, on reflection, an especially radical piece of metaphysics. In "Science and Reality", however, Poincaré addresses the reactions of some of his readers to this claim with a much more aggressive metaphysical claim:

> These words [about the Earth's rotation] have given rise to the strangest interpretations. Some have thought they saw in them the rehabilitation of Ptolemy's system, and perhaps the justification of Galileo's condemnation.
>
> Those who had read attentively the whole volume could not, however, delude themselves. This truth, the earth turns round, was put on the same footing as Euclid's postulate, for example. Was that to reject it? But better; in the same language it may very well be said: these two propositions, the external world exists, or, it is more convenient to suppose that it exists, have one and the same meaning. So the hypothesis of the rotation of the earth would have the same degree of certitude as the very existence of external objects. (1905/2001, 350-51)

In an extraordinary move, Poincaré assures his readers that the Earth's rotation, though only a convention like geometrical axioms, is on firm ground because the very existence of the external world is only a convention, too!

Let us make sure we understand Poincaré's claim by looking at each example of a convention he mentions. Euclid's postulates are conventions because they are definitions; one is free to define alternative geometrical systems which are equally capable of representing all spatial facts. As we saw in Section 1, Poincaré regards the question of whether Euclid's postulates are "true" as meaningless. The only reason to use them rather than other axioms is that other axioms would make the process of reasoning about space more difficult. The rotation of the Earth is a convention because, due to the rejection of absolute space in physics, a model wherein the Earth does not rotate might be no less accurate than one in which it does. Again, the reason we don't use such a model is that it would make reasoning about cosmology extraordinarily difficult. We can conclude, then, that when Poincaré claims that the existence of the external world is a convention, he must mean that the proposition "the external world exists" is no truer than its negation. It might be thought that he intended:

(a) There is no fact of the matter of whether it is true.

(b) He makes no judgment as to whether it is true.

(c) It is meaningless.

We may immediately rule out (c) since it couldn't be more convenient to take something to be true if it were meaningless. This is why we must be careful when Poincaré suggests, as we saw in the previous section, that conventions are not true in the "proper" sense. He does not mean that they are not semantically apt for truth. In his discussion of Poincaré, Psillos observes:

> Note that so far there is no tension between calling a principle [a] convention and taking it to be true. … More specifically, though the truth of a principle is not factual, since the principle is not empirical, the principle is nonetheless true in that it is constitutive of a theoretical framework. (Psillos 2014, 110)

When we say that a convention is nether true nor false, what we mean is that it bears no truth value independent of its (or its negation) being stipulated, not that it is meaningless.

What, then, of the choice between (a) and (b)? It is clear enough that (a) is Poincaré's intention in his Euclid and rotation examples. In the former case there is no fact of the matter because the postulates are definitions. In the latter case there is no fact of the matter because absolute space has been judged theoretically incompatible with the facts. But could Poincaré really have believed that there is no fact of the matter of whether the external world exists? Yes – he could have, and I think he did. We saw Poincaré state that "the external world exists" has the same *meaning*[40] as "it is more convenient to suppose that the external world exists". There

[40] Fr.: "ces deux propositions…ont un seul et même sens" (2013, 324).

is no reasonable interpretation of that claim which both takes it seriously and frames it as a merely epistemological point. Fortunately, there is evidence for his philosophical motivation underlying this striking claim in the text.

Based on the evidence in the VoS sources, the reason that he thinks there is no fact of the matter of whether "the external world exists" is not because he somehow considers this a definition, nor because he thinks some scientific finding undermines its ability to bear a truth value. Nor is it simply that he sees this question as external to any contentful linguistic framework, as Carnap (1950) did. Nor is it because he dogmatically asserts a positivist principle which denies such a proposition a truth-value because it is not empirically verifiable. No – in the final analysis, Poincaré's conviction seems to rest on his neo-Kantian stance. This aspect of Poincaré's thought is well-studied as it relates to his opinions on the foundations of mathematics and science. But it must be added that, in the VoS sources at least, Poincaré is fully committed to the Kantian idea[41] that any possible physical or metaphysical judgment is limited absolutely by the innately given concepts of our understanding while reality (not 'external reality' but simply reality, as in the way things are) is not so restricted. That, at least, is the natural explanation of his grounds for the following vital, also from "Science and Reality":

> [It is not only that] science cannot teach us the nature of things[42] but nothing is capable of teaching it to us and if any god knew it, he could not find words to express it. Not only can we not divine the response, but if it were given to us,

[41] I do not claim that Kant had the view which I am attributing to Poincaré. I understand this as Poincaré's own interpretation and use of Kantian ideas.

[42] His use of "things" here should not be conflated with "external objects" ("*objets extérieurs*"), which he discusses just before this passage. Poincaré is using the phrase "the nature of things" ("*la nature des choses*") here to gesture broadly at reality without making any specific metaphysical commitments (including commitments like 'externality' and 'internality'). (Fr: Poincaré 2013, 321).

we could understand nothing of it; I ask myself even whether we really understand the question. (1905/2001, 347-8)

According to Poincaré, it is not simply that we do not know whether or not our understanding can accurately represent the true nature of reality. He positively declares that it cannot and suggests that we might not even have the cognitive ability to frame this question sensibly.[43] In that case, the concepts 'external world' and, indeed, 'existence' fail to meaningfully apply to ultimate reality. Since the external world's existence can't be identified with a crude fact and – by virtue of being a conceptual judgment – can't apply to the ultimate reality, then, according to Poincaré's view, there is no final fact of the matter on whether the proposition "the external world exists" is true or false. Note that the same reason will lead us to deny that "internal world" has an ultimate referent. For Poincaré, the internal/external divide is an imposition of our understanding. This does not mean that sense impressions are conventional after all, it simply means that framing them as 'internal' (or any other metaphysical description) is not to be taken as a true description of their ultimate nature (independent of whether or not we have a viable conventional choice in the matter).

Note that in the passage from the introduction to *The Value of Science* which was quoted in section 2, Poincaré writes, "a reality completely independent of the mind which conceives it, sees or feels it, is an impossibility. A world as exterior as that, even if it existed, would for us be forever inaccessible." His phrasing here suggests that he is making an ontological, not merely epistemological point. When he does make the epistemological claim, it is prefaced by an ontological concession: "*even if* it existed".[44] Poincaré is saying that the external world

[43] It is true that he is contrasting "the true nature of things" ("*la véritable nature des choses*") with "the true relations of things" ("*les véritables rapports des choses*") and saying that science *does* teach us the latter. But he quickly makes clear that "the true relations of things" means just those which are intersubjectively stable as he previously established (2001, 348; 2013, 322). And we have already seen that the only *objective* relations he wants to commit to here are those between sensations. He is therefore referring to relations between sensations again, which he considers real, but not things out in mind-independent reality.
[44] Fr.: "si même il existait" (2013, 183).

*would* be inaccessible even if he is *wrong* and there ultimately existed something matching that description – a situation that he has just described as "an impossibility".

We have established that Poincaré's ontological position, in the VoS sources, is that "the external world exists" and "there exist external objects", etc., are only conventions. In Section 2, I claimed that the *Science and Hypothesis* quote frequently used to identify Poincaré as an epistemic structural realist was incompatible with the VoS sources, partly because it committed Poincaré to the existence of real, inaccessible, external objects. We now see why that is. In the VoS sources, the only objects Poincaré is willing to acknowledge are conventional ones. Their reality is exhausted in their usefulness for systematising objective phenomenological relations.

Importantly, Poincaré does not regard conventions as completely divorced from all objectivity. He acknowledges that which conventions are convenient and which are not is an intersubjectively stable truth:

> It will be said that science is only a classification and that a classification cannot be true, but convenient. But it is true that it is convenient, it is true that it is so not only for me, but for all men; it is true that it will remain convenient for our descendants; it is true finally that this cannot be by chance. (1905/2001, 350)

Science's *objective convenience* is what Poincaré defends with an early no miracles argument. It is convenient for revealing elegant mathematical harmony between countless phenomenological relations, in the form of empirically adequate theories. As should be clear at this point, Poincaré is not going to give a realist explanation of this objective convenience. The external world and everything in it remain conventional for him. But *how convenient these conventions are* is not something that anyone can freely opt in or out of. And if convenience is not itself conventional, and it is intersubjectively communicable and stable, then the convenience of science is determined by the way things actually are – it is not anyone's creation.

We have now surveyed Poincaré's radical ontological conventionalism. In Poincaré's ontology, sense impressions really exist – yours and mine. And the relations between sense impressions are objective: the same for everyone. The external world and everything in it are conventions chosen only to facilitate productive thought and discourse about phenomenological relations. Which conventions are convenient and which are not is an objective truth.

This brings us to an issue which Poincaré did not adequately address, and which might be a problem for his system: the ontological status of other minds.

### 4.5 The other minds dilemma

The first quote from Poincaré we encountered in Section 3 contained his full argument for the existence of other minds. I will reproduce it here:

> What guarantees the objectivity of the world in which we live is that this world is common to us with other thinking beings. Through the communications that we have with other men, we receive from them ready-made reasonings; we know that that these reasonings do not come from us and at the same time we recognize in them the work of reasonable beings like ourselves. And as these reasonings appear to fit the world of our sensations, we think we may infer that these reasonable beings have seen the same thing as we; thus it is we know we have not been dreaming. (1905/2001, 345)

For Poincaré, if there do not exist multiple minds capable of being in accordance with one another, there can be no objectivity. As we can see, he found it very easy to infer that minds other than his exist. But, based on what we have seen so far, is this not *too* easy? Why is the existence of other minds not a convention like the existence of external objects? The existence of another mind obviously cannot be identified with a crude fact (a pure sense impression) and, even if it could, Poincaré only wants us to trust relations between crude facts when they are confirmed by other minds. Every other entity external to our own experiences is supposed to be a mere convention used to refer to regularities in sense impressions. If the argument for the conventional status of external objects can be used with equal force for the conventional status of other minds, Poincaré might face this dilemma, which we'll call 'the other minds dilemma':

1st horn: Other minds only conventionally exist, and objectivity depends on other minds, so objectivity itself is dependent on a convention.

2nd horn: The argument for the conventional status of other minds can be defeated, but so too can the argument for the conventional status of external objects.

This dilemma is a serious threat. If, via the 1st horn, objectivity itself becomes a conventional stipulation in the same sense that we have seen so far – i.e., such that there is no fact of the matter of whether there are objective facts – then Poincaré's position will be cast even further down the antirealism path than LeRoy's. This is an unacceptable result for him. And the 2nd horn opens the door to a realism that contradicts Poincaré's ontological conventionalism. All is not lost, however. I think that there are ways for a defender of the ontological conventionalist Poincaré to effectively respond to this dilemma without abandoning the substance of his position.

We must first take care to understand the reason why Poincaré thinks there is no fact of the matter of whether the external world exists and question whether this reason extends to other minds. The key piece of evidence, which I quoted and explained in the previous section, is the passage wherein Poincaré claims that even a god who knew the true nature of reality would be unable to express it. The question, then, is whether this neo-Kantian meta-metaphysical stance results in the belief that there is no fact of the matter of the existence of other minds. I would answer: not necessarily. If we regard the proposition "other minds exist" as equivalent to or entailed by "external objects exist", then this problematic result would indeed obtain. But the problem Poincaré sees with taking "external objects exist" as a non-conventional proposition is that doing so would require us to think that, in the ultimate reality, there is a real distinction between 'things' that are mind-external and those that are mind-internal, and this is a description which Poincaré thinks just cannot apply to the ultimate reality. But we need not think that "other minds exist" requires us to impose those or analogous distinctions on the nature of reality. I – the author of this paper – know as a matter of fact that there is at least one manifold of experiences including sensations, thought, and memory: the one that I call my own. I know this through direct acquaintance. Let's stipulate, here, that 'mind' refers only to such a manifold and bracket any philosophical questions about the mind's supposed 'possessor'. Since it is factually true (I would insist) that "there is at least one mind", it would seem bizarre, if not incoherent, to insist that there is no fact of the matter of whether there is more than one. To illustrate: I have no recollection of seeing the Eiffel Tower in Paris and am not seeing it as I write this. Insofar as those are acceptable as facts, is it coherent to claim that there is no truth of the matter of whether *there are* recollections of seeing the Eiffel Tower or present experiences of looking at it (and thus minds which are not my own)? Would believing that the nature of reality is transcendent require me to make such a claim? If I could ask Poincaré's god whether such experiences exist, would it be unable to answer? For all three questions, I think the answer is 'no'. This is not a matter which requires knowledge of the ultimate nature of the reality which underlies experience.

Due to this, it would not be inconsistent for Poincaré to say that there is a fact of the matter of whether there are other minds (once the question is interpreted correctly) even though there is no fact of the matter of whether the external world exists. There is still the epistemological question of how we might *know* that other minds exist, since our experiences are arguably consistent with solipsism. Here, the defender of Poincaré has two options: argue, as Poincaré did, that there is evidence against solipsism, or admit that there is no evidence against solipsism but argue that this is not fatal to Poincaré's position. I think that either route has potential.

For the first option, we would ideally take leave from Poincaré's own argument (quoted at the beginning of this section). But it must be given philosophical reinforcement to be convincing. Consider the following two propositions:

(1) I encounter at least some reasonings that seem to be produced by other beings which I know are actually produced by other beings.

(2) I experience at least some sensations whose qualities I know to be objective.

Poincaré justifies his relationism by denying (2). It is precisely because we cannot rely on our own devices to tell us whether a sensation is fabricated or distorted that we should (according to Poincaré) take as objective only what is confirmed by other people (i.e., only relations). But his argument for the existence of other minds seems to rely on the truth of (1), which seems analogous to (2). One way that Poincaré's defender might maintain (1) and deny (2) is by arguing that reasoning is essentially self-conscious. This means defending the view that if someone is producing rational thought then they are necessarily aware that they are producing rational thought, plus the notion that reasoning cannot be evident unless someone has produced it through rational thought. There is extensive applicable literature on self-consciousness both in the 'Western' philosophical tradition (see Smith 2020, especially Section 4.2) and the Indian one (see Kellner 2010), so this is not at all a weird move or one that can be easily dismissed. If we accept such a view, then in cases wherein what Poincaré calls "ready-made reasonings" appear to us, we can know for sure whether or not it was us who produced it and know for sure that it was produced by a mind. This can clearly result in strong evidence that there are other minds.[45]

For the second option, it might be admitted that Poincaré *does* have to take the existence of other minds on faith, and argued that this singular leap of faith, which allows for objectivity and the rest of his system, should not be denied to him because of his otherwise sceptical stance. Poincaré neither declares an absolute Cartesian scepticism for himself, nor does his system logically require it. He can consistently say that it needs to be relaxed specifically for the question of other minds, and it may be argued that this puts his view in the same boat as any other metaphysical system. Attempts to disprove solipsism are rare and rarely considered successful; most philosophers simply regard solipsism as a fringe idea that does not need to be thoroughly addressed. It is not obvious that this latitude cannot be extended to Poincaré, even if we demand that his defenders provide special justification in his case.

I am not claiming, here, that these responses to the other minds dilemma are successful, but I am claiming that they are not obviously unsuccessful. And there are, doubtlessly, other possible responses which I have not thought of. The other minds dilemma therefore should not incite us to jump to the conclusion that the ontological conventionalist position I have attributed to Poincaré is hopeless. Furthermore, the dilemma should not be taken to make the rival 'epistemic structural realist' reading of Poincaré more attractive here, since that version of Poincaré faces comparably dire objections. Chief among them is Newman's objection

[45] The appearance of others' reasoning in dreams or hallucinations will cause difficulty for this position, but there is perhaps an effective avenue of response in the fact that these appearances are paradigmatically not objective whether or not anything is.

(mentioned in Section 3), which some scholars regard as insurmountable for the ESRist (see Ainsworth 2009). Leading efforts to save ESR from Newman's objection (e.g., Melia and Saatsi 2006) do so by modifying the position well beyond what can be represented as Poincaré's own view. And this, of course, is not the only objection to structural realism. So, a structural realist interpretation of the VoS sources is by no means obviously more charitable than my ontological conventionalist one. And mine, I hope to have shown, is far better supported by the text.

### 4.6 Conclusion

This chapter presents three contributions to the study of Poincaré:

- It highlights the incompatibility one of Poincaré's key metaphysical claims in *Science and Hypothesis* with his position in the VoS sources. This discontinuity has not, to my knowledge, been previously acknowledged in the literature.
- It explains Poincaré's relationism in the VoS sources as fundamentally phenomenological and closer in metaphysical import to constructive empiricism than to structural realism.
- It explains Poincaré's radical conventionalism regarding the ontological status of external objects and the external world.

I hope to have made clear that, in the VoS sources, Poincaré was not a realist in the contemporary understanding of the term. Poincaré thought that the existence of theories with any significant degree of empirical adequacy – i.e., theories which accurately and precisely represent regularities which hold between complex and varied phenomena – is a tremendous achievement which demonstrates the objective value of science. He thought that the existence of intersubjectively stable mathematical connections between phenomena throughout the history of science, even across generations and theory change, demonstrated that science was more than just collective imaginings imposed onto the world. What he did *not* do (in the VoS sources) was claim that these mathematical connections represented real relations between real mind-independent objects. While, in “The Theories of Modern Physics” from *Science and Hypothesis*, Poincaré considers the entities of scientific theories to be “images” which stand in for “real objects” which are forever inaccessible, Poincaré in the VoS sources unambiguously refuses to claim that science’s conventional objects represent anything other than assemblages of objective phenomenological relations. For Poincaré in the VoS sources, the ontology of science, and indeed that of the ordinary world of experience, is nothing more than a conventional construction with the sole purpose of facilitating the representation of objective relations between phenomena.

It would be incorrect to say that Poincaré was only interested in the empirical adequacy of scientific theories *instead* of their truth.[46] He chose to begin *The Value of Science* with the sentence, “The search for truth should be the goal of our activities; it is the sole end worthy of them” (1905/2001, 189). However, it was Poincaré’s considered opinion that ontology did not play an objective role in the ultimate truths about reality that humans could access. “The truth we are permitted to glimpse,” he writes, “is not altogether what most men call by that name” (1905/2001, p.191). The truth which science shows us, according to Poincaré, is the mathematical harmony between various phenomenal relations. On his view, any and every metaphysical interpretation of this structure – no matter how modest – should be understood as a convention.

[46] That Poincaré was not primarily interested in truth was alleged by Daston and Galison (2007, p.261).

### 4.7 Poincaré and Nāgārjuna

Is Poincaré's relationism what we have been looking for? Is it the way that a Nāgārjunian should account for the success of science? I have my reservations. Poincaré's view is compatible with Madhyamaka, but it lacks explanatory power. It can serve as a sophisticated antirealist account of scientific knowledge which is consistent with Madhyamaka. As I pointed out when comparing Poincaré with van Fraassen, Poincaré's relationism holds the goal of science to be only the mathematical unification of diverse phenomenal relations. We might say that, on the question of scientific realism and antirealism, it is a basically empiricist perspective which disavows commitment even to the factual existence of the observable world independent of consciousness. As a basically empiricist perspective, it rejects in principle the demand for an explanation of science's success that goes beyond said unification of experience. This is important to develop as an option for Mādhyamikas, and many may opt to go this way, but the question we must ask is whether Nāgārjuna's arguments, if successful, *force* us to adopt this perspective on science. Can Mādhyamikas help themselves to explanations of what makes certain theories more successful than others which appeal to a genuine connection between those theories and reality?

Perhaps that the preceding sentences are advancing in the wrong direction. It might be more prudent to ask whether even Poincaré's story is coherent unless we say at least a little bit more about the role of reality transcendent to experience as the source of objectivity. We have the experiences we do because reality is such that we have them, and the same is true for others, and some phenomenal relations are objective because reality is such that they are, and some conventions are more convenient than others because reality is such that they are. For Poincaré, we cannot explain how reality conditions the world as we experience it because reality is transcendent; we cannot even represent 'it' in our minds because our minds' speculative abilities are limited to the innate concepts of our understanding. Nāgārjuna seems to agree:

> Not ascertained from another, pacified, not fabricated by hypostatisation,
>
> Free from conceptualisation, not having different meanings – this is the nature of reality (*tattva*). (MMK 18.9)

Yet we know something about it: it is the source of all objectivity. It is, by definition, the very way things are, including, but not limited to, experiences and concepts themselves. Experiences are conditioned things, and the nature of their objective conditions transcends them. But may we not make something of the objective conditions simply from the fact that they are conditions? Does their status as conditions not indicate some potential for an explanatory role? These are the thoughts that lead us into the next chapter, where I explain a form of sentence that might lead to a way for even Mādhyamikas to claim some knowledge about the structure of reality.

## CHAPTER FIVE

## Structural Realism for the Sceptics

Bertrand Russell defended the notion that our knowledge of objective reality is limited to its structure as a response to antirealist concerns about objects of perception. Newman's (1928) objection was meant to demonstrate that any claims about the world's structure alone were trivially true, given that the world contains a sufficient number of things. Though there have since been rejoinders to Newman's objection in the context of contemporary scientific structural realism, these all make at least one concession: that our knowledge of objective reality cannot be *purely* structural as Russell originally claimed – it must extend also to observable facts. I argue that this consensus is mistaken. One can be a successful structural realist even as a brain in a vat. Structural realism is therefore not only a middle-ground answer to scientific antirealism: it is an effective middle-ground answer to even the most radical forms of scepticism.

### 5.1 Introduction

Epistemic[47] structural realism in the contemporary philosophy of science literature is a middle-ground position between scientific realism and antirealism. Scientific realists believe that our best scientific theories give us knowledge of the unobservable world, and scientific antirealists deny this. Structural realists hold that our best theories give us knowledge only of the structure of the unobservable world, not its nature. I argue in this chapter that structural realism can be put to further use – in line with its original use suggested then abandoned by

[47] There is also ontic structural realism, but this is a different topic that it won't help to go into. When I refer to "structural realism" in this article, it will be the epistemic sort.

Russell roughly a century ago and not defended since. It can work as a middle-ground position between the view that we have knowledge about reality as it is external to experience and the view that we have no such knowledge. I will show that it is possible to make true, non-trivial claims about the structure of objective reality even if experience is utterly deceptive in regard to its nature.

In the context of the overall project, this chapter serves as a proof of concept for the ability of someone to make true statements about the structure of systems which they cannot understand, but which underlie their experiences. As I indicated at the end of the last chapter, this is the sort of claim that I think a Mādhyamika can use to explain the success of science (and theories in general – including Buddhist theories). It will take more work than what is found here to see how this sort of claim can be used by a Mādhyamika, but this will be addressed in the final chapter. For now, I will deal with this question on its own terms.

During the course of the chapter, I will contrast scientific antirealism with what I call ‘perceptual antirealism’, which should be understood as the view that our senses do not provide accurate information about objective reality. It is stronger than scientific antirealism, which does not challenge the idea that perception is a reliable means of learning about objective reality. Perceptual antirealists would include radical sceptics and transcendental idealists (not limited to Kant’s strict view but anyone who thinks that the nature of objective reality transcends our cognitive faculties). I focus on radical scepticism in the chapter but will have some comments specific to transcendental idealism at the end. I will frequently mention structural realism as “answering” either scientific or perceptual antirealism. By this I mean that it challenges the antirealism in question by proposing knowledge of objective structure as something that can be realistically achieved even if it is conceded to the antirealist that knowledge of the objective nature of reality is unobtainable.

That structural realism can answer perceptual antirealism is a significant finding with regard to any philosophical perspective according to which objects of perception are taken to be distinct from, and therefore possibly deceptive as means of learning about, mind-independent reality. But I think that the case for the significance of this is sufficiently made by showing that structural realism can be effective even in the context of a thought experiment designed to be an epistemic worst-case-scenario: the brain in a vat.

The contents of the chapter are as follows. §2 will explain some of the historical and conceptual foundations of my proposal. §3 will explain how I propose one formulates claims about only the structure of an unobservable system. At this point, the conversation will be limited to scientific structural realism. In §4, I will explain how to formulate my proposal for scientific structural realism in formal language. In §5, I will show how the proposal I will have developed up to that point can be adapted to answer to perceptual antirealism as well. It is here that I will explain how a person who is a brain in a vat could be a successful structural realist. In §6, I will distinguish my proposal from the nearest existing scientific structural realist proposals and comment on whether they, too, could be adapted to answer perceptual antirealism. Here, I will also offer some advantages that my proposal has over theirs even when the discussion is limited to answering scientific antirealism.

### 5.2 Newman's objection and the existing literature

In the early 20th century, Bertrand Russell was convinced that he had solved scepticism of a Kantian sort – at least insofar as it applies to mathematical sciences – through appeal to "structure". He explained that structure is an abstraction from relations: it can be understood as a "map" (Russell's word)

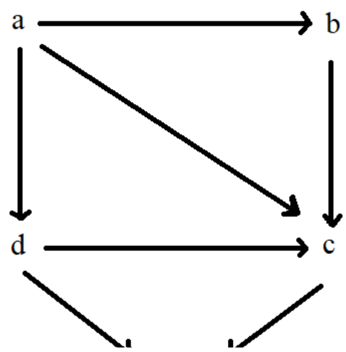


*Figure 1*

which charts the pattern a relation or relations instantiate in some domain. His example is a relation whose extension is the couples $\langle a,b\rangle, \langle a,c\rangle, \langle a,d\rangle, \langle b,c\rangle, \langle c,e\rangle, \langle d,c\rangle, \langle d,e\rangle$. The structure of that relation is pictured in Figure 1. The key feature of structure, here is that it is abstract and multiply realizable. There are many relations whose extensions in some domains will form the same pattern (i.e., structure) pictured in Figure 1.

According to Russell, "a great deal of speculation in traditional philosophy" might have been avoided if philosophers had realised the importance of structure. He claims, for example, that if subjective phenomena "are caused by things in themselves" then, contrary to the notion that the things in themselves would be utterly mysterious, "the objective counterparts [of phenomena] would form a world having the same structure as the phenomenal world, and allowing us to infer from phenomena the truth of all propositions that can be stated in abstract terms and are known to be true of phenomena."[48] According to him, "the only difference must lie in just that essence of individuality which always eludes words and baffles description, but which, for that very reason, is irrelevant to science" (Russell 1919/1993, 60-1). Carnap praised Russell for pointing out "the importance of structure for the achievement of objectivity" and criticised Poincaré for teaching that objectivity was to be found in relations when (according to Carnap) it is *structure* that is the key (Carnap 1928/2003, p.30). Carnap, though, like Poincaré, had a different objective to Russell from a metaphysical point of view. Russell was trying to offer a way to describe the structure of the world as it is independent of the mind. Neither Carnap nor Poincaré were interested in such a project; for them, what objectivity was to be found only in what was intersubjectively stable in experience.[49] Hence, the present paper is in the spirit of Russell, not Carnap.

[48] This isn't quite right. On the hypothesis in question, the phenomenal world would only need to instantiate *substructures* of the structure of the objective world. But Russell's point is otherwise sound, in my opinion (bracketing Newman's objection).

[49] For an analysis of Poincaré's position in the essay that Carnap quoted, see Holder (2023).

Russell utilised this argument in his later work, *The Analysis of Matter* (1927/1992). This prompted a reply from M. H. A. Newman (1928), who argued that if Russell's point amounted only to the claim that we know that *there exist* some stipulated structures in objective reality, then the claim is quite trivial. For, according to Newman, there exist relations among *any* set of objects which instantiate any stipulated structure consistent with the cardinality of the set. Other authors have defended Newman's conclusion with their own arguments (e.g., Demopoulos and Friedman 1985; Ainsworth 2009). It will not help us to dwell on the details of these arguments here. In this paper, let us assume that Newman's objection effectively shows that any system trivially instantiates any stipulated 'pure' logico-mathematical structure which is consistent with how many things are in it (i.e., consistent with the cardinality of the set of its constituents). The structuralist's conundrum is that narrowing down which relations' structure they are interested in seems to require non-structural knowledge of those relations, which defeats the point of appealing to structure in the first place. Russell conceded to Newman at the time (see Demopoulos and Friedman 1985, 631).

In 1985, Demopoulos and Friedman reintroduced Newman's paper to the literature, arguing that it was underappreciated as an objection to more contemporary positions. Demopoulos and Friedman turned their attention to structuralism as expressed through Ramsey sentences[50] and presented a new objection based on Newman's. They argued that the Ramsey sentence of a theory sentence is trivially true once all of the theory's "observational consequences" are true. This objection was further developed by Ketland (2004), who concluded that the truth of theoretical Ramsey sentences can establish no more than the

[50] If we have a sentence from a theory which contains predicates describing the nature of the unobservable world, the corresponding Ramsey sentence takes the predicates in the original sentence which are not 'observational' (there is some disagreement over what precisely this disqualifies) and replaces them with predicate variables which are existentially bound. The effect is that, in a Ramsey sentence, only observational predicates are given interpretations, but unobservables may yet be said to exist and stand in certain relations to each other and to observables.

empirical adequacy of the theory plus a cardinality constraint on its domain. This conclusion has been challenged by many, including Zahar and Worrall (2001), Votsis (2003), Melia and Saatsi (2006), and Smithson (2017). But not everyone finds these rejoinders satisfactory (see Ainsworth 2009).

The move to focusing on Ramsay sentences is concomitant with an important shift in perspective regarding the problem that structuralism was supposed to be solving. Russell was motivated by sceptical concerns about perception; for him, our knowledge of the world was supposed to be entirely structural, even when it comes to those aspects of the world that we would today call 'observable'. By the time that the issue began to be reframed with Ramsey sentences, the concern was no longer objective knowledge of the world as such, but knowledge of the unobservable world. This is the sole concern of contemporary scientific realism, including contemporary scientific structural realism as introduced by Worrall (1989). The observable world tends to be understood as epistemically unproblematic by all participants in the structural realism debate today. This is likely, in part, because scepticism about objects of perception is not currently popular in the philosophy of science. But it is also due to a sense that structuralism which does not make use of knowledge of the observable world will be helpless against Newman's objection (see Zahar and Worrall 2001). This latter point is reasonable on the face of it. It is not apparent how knowledge of the structure of a system could be anything but trivial without appealing to at least some non-structural knowledge about the system. Relying on knowledge of its observable properties is the obvious way to handle this problem. If we have no epistemic access to the nature of external reality ('observable' or otherwise), then to what could we appeal in order to pick out the system we want to target?

My answer is that we can appeal to experience itself. I can acknowledge that I have experiences even if I am sceptical about the efficacy of those experiences as sources of knowledge about the nature of external reality. Here, external means external to experience;

mental things which are not experience also count as 'external reality' for us. I may also suppose that the contents of my experiences are conditioned in part, at least some of the time, by external reality (their being conditioned by external reality does not mean that they tell me anything about the nature of external reality). I won't argue for this – I take it as something that is reasonable to suppose and see no reason why a structural realist would be less likely to suppose it than anyone else. It was, as was shown in the quotes above, an essential part of Russell's original idea. And this is all that is needed to pick out some system I am interested in. If I have some experience and suppose that the content of this experience is what it is because of the existence or activity of some external system, I may start formulating questions and claims about that system even if I have no idea what it is.

This is the foundation of my case: all we need in order to make structural realism work are experiences (which may be universally deceptive as to the nature of what is experienced) and the supposition that external reality causes the content of some experiences to be some way rather than some other way some of the time. Relying on those, we may make non-trivial claims about the pure logico-mathematical structures of those systems which we suspect of causally influencing our (perhaps deceptive) experiences. Using this strategy, one may not only describe the structure of the immediate stimuli of experience, as Russell was attempting, but one may also make claims about the unobservable-in-principle systems described in successful scientific theories, as the contemporary structural realists desire. If this works, we may, excepting the supposition that external reality causally influences our experiences, have the option of holding, without succumbing to triviality objections, that *all* of our knowledge of the external world is structural.

In the next section, we will bracket perceptual antirealism and begin to formulate my proposal first as an answer to scientific antirealism.

## 5.3 Knowing the structure instantiated by unknown relations

We want to say that we know some of the structure of the unobservable world. By this we mean that we know the abstract logico-mathematical pattern instantiated by some unobservable real relations. Newman's objection tells us that it is trivial that there exist relations instantiating any structure amidst some domain once the instantiation of those relations and that structure is consistent with the domain's cardinality. For this our claim to be non-trivial, we must specify which relations in particular instantiate the structure. As structuralists, we cannot do so by specifying the nature of the target relations. But that is not the only option: we can pick out the instantiation of specific relations by certain relata in the domain via the consequences or implications of their instantiation by those relata – this is just a weird thing to do under most circumstances.

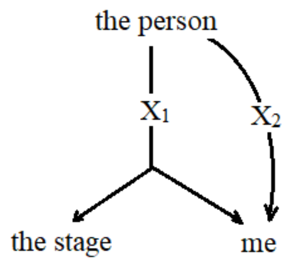



*Figure 2*
*Unknown to my friend are the following interpretations by which this structure represents the cause of my problem:*

*$X_1$: standing between*
*$X_2$: taller than*

For example, say I am prevented from seeing a performance at a concert because a much taller person is standing in front of me. I might later tell a friend, "the instantiation of some relations prevented me from seeing the show," where "the instantiation of some relations" refers specifically to the person being *taller than* me and their *standing between* me and the stage. Let's say that I don't tell my friend that those are the relations to which I am referring, but I present them with the image shown in Figure 2 and say that the culprit relations can be represented as interpretations of $X_1$ and $X_2$. According to Newman's objection, this image alone contains no non-trivial information beyond the labels/descriptions of the objects in the domain and the fact that there are three of them. According to the objection, it is trivial to define relations between these objects which interpret $X_1$ and $X_2$ such that this structure is an accurate representation of the world at the specified time. However, it is not trivial (though it may be easy, with imagination) to define a pair of relations instantiating this structure which could have prevented my seeing a show, and it is certainly not trivial to define *the* specific pair of relations instantiating this structure which *did* prevent me from seeing a show: that is information which must be retrieved empirically. That *those* specific relations instantiate this structure is novel empirical information for my friend, even if my friend never discovers the nature of the relations.

The ability to refer to specific relations only through their connection to identifiable phenomena is the key to the response to Newman's objection that I am proposing. When identifying the structural knowledge imparted to us by a scientific theory, the important claim should not simply be "this is a structure instantiated by a system that exists" (whether that structure is purely abstract, or partially interpreted, or inclusive of modal operators, etc.). The important claim should rather be "this is a structure instantiated by the system *of relations* underlying the attraction and repulsion of magnets", or the dispersion of gases, or whatever. The nature of those relations does not need to be correctly interpreted for that structure to be an incredible discovery. But there is yet much work to do in developing and explaining this claim so that it can be applied with precision and avoid pitfalls.

The biggest potential pitfall to be considered is, of course, that this strategy cannot be specific enough with the relations it picks out when applied to cases that the structural realist cares about. In the concert example, the full interpretation of the objects in the domain was allowed – something that cannot track to cases in which realists want to claim knowledge of only the structure of an unobservable system. In the structure that I show my friend (Figure 2), imagine that we replace the named relata in the structure with letters standing in for unidentified relata, yielding Figure 3. If I were to present my friend with *this* diagram and say that $X_1$ and $X_2$ stand in for relations which

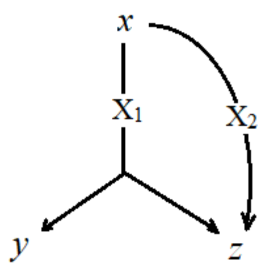


Figure 3

prevented me from seeing the show, we will run into difficulties. For, even though this claim is not strictly trivial, there is an immense number of relations which make it the case that I cannot see the stage in the situation described in the example. For example, if people were transparent, then a tall person standing between the stage and me would not prevent me from seeing the show. Therefore, among the relations which make it the case that I cannot see the show at that time are all of the physical relations which make it the case that people are opaque. If *any* of those relations instantiate the structure in Figure 3 as interpretations of $X_1$ and $X_2$, then my claim that "this is a structure instantiated by the relations which stopped me from seeing the show" will come out true. This is not an acceptable result; my goal in presenting the structure was not to pick out some obscure substructure in optics or something. Let's call this 'the problem of background relations' and ensure that it does not show up when we apply my strategy to scientific theories.

I will make reference to models in the presentation of my proposal. A model is some non-linguistic entity which exemplifies theoretical axioms or equations. Mathematical models offer mathematical descriptions of systems in which theoretical axioms are true and/or which demonstrate solutions to theoretical equations. We may also speak of physical models such as material artefacts, computer programs, or even natural systems, in which a set of axioms or equations are similarly true. Though I rely on models for my account, I would not claim to be using the 'semantic approach' to scientific theories, which identifies theories with classes of models. My perspective on models is similar to that of Morrison and Morgan (1999) in that I see them as instruments which have certain relations to theories and to the world but are distinct from both. I will not dwell on this here because I think it is immaterial to my argument: it is not necessary to have any specific conception of what a model is (beyond the basics I have already mentioned) in order to appreciate their use here.

I will now define some of my own technical terms. First is 'phenomenal fact'. This means something that is the case in the observable world.[51] Note that it does not mean something that is *observed to be* the case. For example, "the sun has risen every morning for the past five thousand years" and "it is impossible for light to pass through lead" might be considered facts about the observable world even though they are not verifiable through observation. A phenomenal fact may consist simply in the existence of some phenomenon (e.g., a reading on an apparatus, a light, sound, or movement), or it may consist in a regularity (e.g., the sun is never visible in Barbados at 1:00am local time; under *x* conditions, the reading on this apparatus is always a value greater than *y*). Models can be used to represent and help explain any sort of phenomenal fact.

Next, I must explain what I mean by what 'underlies' a fact. Take two propositions P and Q; the fact that P underlies the fact that Q iff Q is true, P is true, and Q is true because P is true. Further, take an object or relation *x* and a proposition Q; *x* underlies Q iff the fact that *x* exists or is instantiated underlies the fact that Q in the sense just explained. Clearly, the content of 'underlies' rests heavily on what is meant by 'because'. It is not a logico-mathematical concept: it must be supplied meaning by some theory of causation.[52] But, for the purpose of answering Newman's objection, what theory of causation we use to interpret this does not matter too much. All that matters for our purpose is that facts of the form '*x* causes *y*' are rarely trivial. I will be assuming a basic counterfactual notion of causation (Q would not be true if P were not true) but I will not make any effort to flesh it out for the reason just given.

[51] By 'observable' I mean something like van Fraassen's (1980) use of the term. But the demarcation between observable and unobservable is not a big deal here. I am using the distinction because it is the salient one in the literature. My proposal should work no matter how one specifies the precise demarcation.

[52] Why causation and not explanation? It is a theory that explains, not a fact, object, or relation. Consider someone who says that exposure to sunlight causes sunburn because the sun ignites tiny fires under the skin. They have the cause right in the first instance, but their explanation is wrong.

We're now prepared to articulate the first version of my proposal. Note the following stipulations (i – iv):

For some model, *M*,

i. P is a phenomenal fact presented[53] in *M*,
ii. There exist in *M* certain relations instantiated by certain relata such that a certain structure S is satisfied.
iii. The fact described in (ii) underlies the fact that P in *M*.
iv. P represents a phenomenal fact in reality, P'.

Our claim is as follows:

1. There exist in reality certain relations instantiated by certain relata such that the structure S is satisfied.
2. The fact described in (1) underlies the fact that P'.
3. For any logical connection found between the facts mentioned in (2), approximately the same logical connection can be found between the facts mentioned in (iii).

By design, this formulation gives us great freedom in picking and choosing what structures we are committed to and how we want to represent them. S is expected to be satisfied by some part of *M* which is meant to represent the unobservable world – the realist can freely pick which part they want to commit to according to their reasoning – and there is no expectation that it must represent the complete structure of whatever system the model is presenting. *S* itself might be represented in whatever form is convenient (whether linguistic, mathematical, graphical, etc.).

The point of (3) is to block the problem of background relations discussed above. Depending on the theory of causation in use, there might be many background relations which

[53] I follow Katherine Brading's and Elaine Landry's helpful distinction between the system that a model *presents* and the system that it *represents* (Brading and Landry 2006, p.573; Brading 2011, p.48). The former is the system which is described by the model and which does not necessarily exist beyond it; the latter is the real system that can be identified as what the model has described.

satisfy both (1) and (2), but it is not feasible that the logical connections between the facts presented in the model are preserved between the facts in reality. Let's look at an example to see how this all works.

Say there is some solid substance (we'll just call it 'the substance') which mostly dissolves when submerged in a certain liquid ('the liquid'). It is observed that when any quantity of the substance is exposed the liquid, the remaining solids (which look like the substance) always have a quarter of the mass and volume of the original quantity.

There is a model of a successful theory which explains this fact. In the model, the substance is identified with a certain molecule, and the liquid is identified as a catalyst which reacts with the 'substance-molecule' to ultimately produce five products: four new molecules and the catalyst itself. Let's label these new molecules A-products, B-products, C-products, and D-products. They are produced in equal measure and are equal in mass. A-products in close proximity form strong bonds with each other very quickly, making it almost impossible to keep them in a homogenous mixture with a liquid. B-, C-, and D-products, however, have no such tendency and readily mix with the catalyst. According to the model, then, the solids which remain when the substance is submerged in the liquid are A-product particles. They are the only product of the reaction to end up as a solid because, among the products, A-particles are the only molecules which quickly pack together to form a solid.

Here, our phenomenal fact is “when some quantity of the substance is submerged in the liquid, only one quarter of that quantity remains solid”. To explain it, a theorist who subscribes to the model we have just described might point to two unobservable relations as underlying this fact: A-products *solidify much more readily than* B-, C-, or D-products, and A-, B-, C-, and D-products are *produced in equal measure to* each other. These generate the structure in Figure 4, which we have represented in the diagram style used by Russell and Newman.[54] In the structure, the solid lines represent the *solidify much more readily than* relation and the dotted lines represent the *produced in equal measure to* relation.

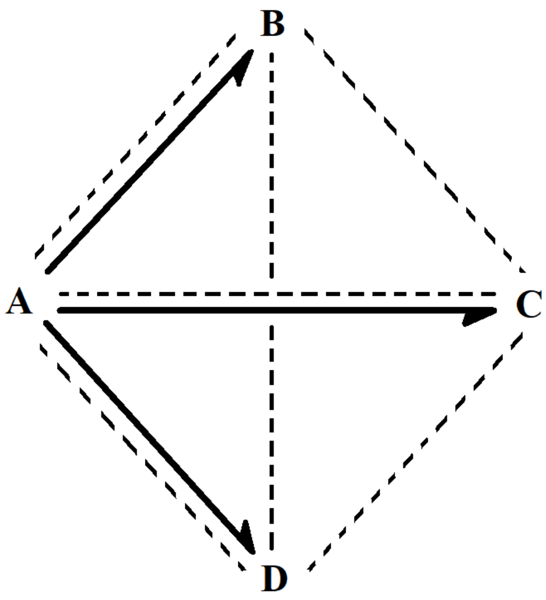


*Figure 4*

Now let us say that one of the theorists is a structural realist. They are unconvinced of the existence of any of the molecules we have referred to so far, and any of the relations which ostensibly hold between them. But they do commit to the belief that there exist relations which underlie the fact in question, that these relations instantiate the abstract structure depicted in Figure 4, and that they share approximately the same logical connection to the phenomenal fact that the two relations presented in the model do. It is difficult to imagine relations which satisfy all of these criteria, says the structural realist, and not remotely trivial to say that any such relations exist, so the fact that *there are* relations which satisfy the criteria is the knowledge that theory has given us via this model.

[54] The structure formed by the relation represented by the solid line is the same structure that Newman uses as an example in his paper.

Now let us say that the structural realist's suspicions were correct. A-, B-, C-, and D-products do not exist. The liquid is not a catalyst but a reagent which only reacts with three-quarters of any given quantity of the substance. This is because the substance is composed not of one kind of molecule but of four isomers. Isomers are molecules which contain exactly the same number and type of atoms as one another, but have those atoms arranged in different structures.[55] We will call the isomers which constitute the substance A-isomers, B-isomers, C-isomers, and D-isomers. Each one contains one each of the atoms W, X, Y, and Z. Figure 5 shows the structure of each isomer.

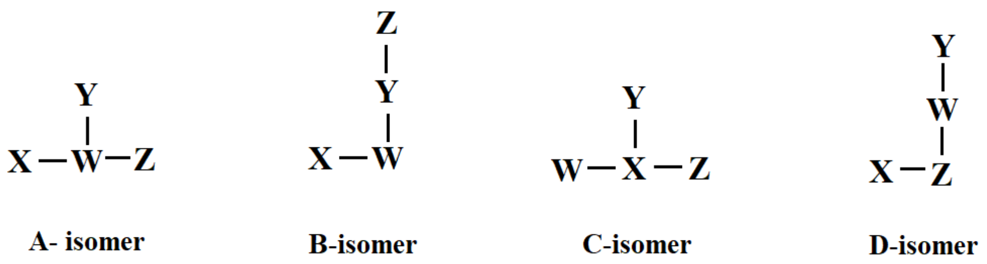



*Figure 5*

The process which creates the substance generates these isomers in equal quantities and in even distribution. They are *all* almost impossible to keep in a liquid state – they solidify rapidly when in close proximity to each other. What is happening is that the liquid is reacting with B-, C-, and D-isomers to produce different molecules altogether – ones which are happy to exist in a liquid mixture. A-isomers, however, happen to be much more stable than the others. They do not react with the liquid at all. When the substance is submerged in the liquid, all of the B-, C-, and D-isomers in the substance react with the liquid and the resultant products form a

[55] This describes a 'structural' or 'constitutional' isomer, which is only one type. There are also stereoisomers which are connected in the same arrangement but differ in how their atoms are spatially oriented. For the sake of simplicity, I'm ignoring stereoisomers in the main body of the paper and using the word 'isomer' to mean only structural isomers.

homogenous mixture with the liquid. Since the A-isomers do not react, they quickly bond with each other and form solids.

Here, the relation *solidifies much more rapidly than* does not exist between any of the isomers; they all solidify just as rapidly as one another. However, we still have relations between them which underlie the phenomenal fact and instantiate the structure in Figure 4. The relations are: *is much more stable than* (solid line) and *constitutes the substance in equal proportion with*[56] (dotted line). The letters in the diagram denote the correspondingly labelled isomers. It is *these* relations which really underlie the fact that "when some quantity of the substance is submerged in the liquid, only one quarter of that quantity remains solid" and they do indeed instantiate the structure in Figure 4. So far, the structural realist is vindicated. But what of the logical connection criterion?

We can see from Figure 5 that the atoms in an A-isomer also instantiate the structure in Figure 4. The bonding relation between atoms instantiates the substructure drawn by the solid lines and the relation *constitutes an A-isomer in equal proportion with* instantiates the substructure drawn by the dotted lines. Given that it is this particular structure which has resulted in the A-isomer's relative stability, it could be reasonably argued that these relations also underlie the phenomenal fact. These relations therefore also satisfy the first two criteria of my proposal – but they do not satisfy the third. That only a **quarter** of the original quantity remains solid is a logical consequence of the fact that there are **four** isomers which instantiate the relations *is much more stable than* and *constitutes the substance in equal proportion with*. That logical connection does not exist between the proportion of solids remaining and the number of *atoms* in an A-isomer. However, in the original (incorrect) model, this logical

[56] Technically, this should be *is an instance of the type of isomer of which instances constitute the substance in equal proportion with instances of the type of isomer instantiated by*, but this level of precision would make the following discussion difficult to read. I hope the reader will forgive the sloppier abridged version and appreciate my intent.

connection *does* exist between the proportion of solids remaining and the number of (A-, B-, C-, D-)products thought to instantiate the *solidify much more readily than* and *produced in equal measure to* relations. The fact that the structural realist is targeting therefore, in this respect, share the same logical relation to the phenomenal fact as their counterparts in the incorrect model. It is by claiming that connections of the sort pointed out here (i.e., four in one fact – a quarter in the other fact) are preserved between the facts in the model and the facts in reality that the third criterion is satisfied and the problem of background relations is deflected.

I hope that it is now clear how my proposal works. By identifying a phenomenal fact and making reference to a theoretical model, a structuralist can commit to the instantiation of some specific relations with some specific structure as a significant discovery, while admitting to no knowledge whatsoever of the nature of those relations or their relata. These relations are identified based on their status as causally underlying the phenomenal fact, which blocks Newman's objection since it is not remotely trivial to identify relations whose instantiation underlies some specified fact.

### 5.4 The formal account

Now that I have walked through the fundamentals of my proposal in informal language, it is time to put it into formal language for precision. As I mentioned before, Ramsey sentences are often used as formal means of making claims about the structures of unobservable systems. I will take Ramsification as a point of departure how we can rework some of the resources it gives us to formulate the kind of claim we want.

Until now, my talk of 'relations' has been consistent with the non-technical, English-language notion of a 'relation', such that a relation always is instantiated by at least two relata. When we move on to the technical logical notion of a relation, though, relations are described by predicates over $n$ elements (e.g., T$xy$: $x$ is taller than $y$; this is a binary relation because $n$=2). And, for our purposes, there is little relevant logical difference between predicates for which $n > 1$ and those for which $n$=1. So, one-place predicates, which in natural language would be called properties, are also considered 'relations' in the technical sense. It's this technical sense that we will be employing from now on.

The first thing we need to do is to, with reference to some model, identify a proposition which we take to underly some phenomenal fact. Let us do this now, using the example in the previous section.

> **Phenomenal fact:** "When some quantity of the substance is submerged in the liquid, only one quarter of that quantity remains solid."
>
> **Underlying proposition:** "A-products form solids very quickly while B-, C-, and D-products do not, and each kind of product constitutes a quarter of the new molecules produced under catalysis."

Now we will formulate the underlying proposition using a sentence, θ, in a second-order language.

| | |
|---|---|
| A$x$: | $x$ is an A-product. |
| B$x$: | $x$ is a B-product. |
| C$x$: | $x$ is a C-product. |
| D$x$: | $x$ is a D-product. |
| S$x$: | $x$ forms solids very quickly. |
| Qx: | $x$ is an instance of one of the four types of molecule which each represent a quarter of the new molecules produced under catalysis when the substance is exposed to the liquid. |

θ: $\forall w \forall x \forall y \forall z((Aw \wedge \mathrm{B}x \wedge \mathrm{C}y \wedge \mathrm{D}z) \rightarrow (Sw \wedge \neg Sx \wedge \neg Sy \wedge \neg Sz \wedge \mathrm{Q}w \wedge \mathrm{Q}x \wedge \mathrm{Q}y \wedge \mathrm{Q}z))$

I will borrow from Smithson (2017) his notation for "the fact that…". $\llbracket\theta\rrbracket$ means "the fact that θ". Furthermore, we will refer to the phenomenal fact listed above as P (there's no need to put this in a second-order language). With that, we will define:

$\mathbb{U}\llbracket\theta\rrbracket\llbracket P\rrbracket$: the fact that θ underlies the fact that P

For traditional realists, $\mathbb{U}\llbracket\theta\rrbracket\llbracket P\rrbracket$ is all that needs to be said. Structural realists, however, will want to Ramsify θ to $R(\theta)$:

$$R(\theta):\quad \exists X_1\exists X_2\exists X_3\exists X_4\exists X_5\exists X_6\forall w\forall x\forall y\forall z((X_1w \wedge X_2x \wedge X_3y \wedge X_4z) \rightarrow (X_5w \wedge \neg X_5x \wedge \neg X_5y \wedge \neg X_5z \wedge X_6w \wedge X_6x \wedge X_6y \wedge X_6z))$$

This, of course, is trivial given that there are at least four things in the world. It will not do to state that a triviality underlies P. But, while it might be a triviality to say, as $R(\theta)$ does, that some predicates satisfying the relevant structure exist, the triviality can go away if we say a bit more about the specific predicates we want to pick out. This must be done without interpreting them. Let's start by writing a new sentence, $\not{R}(\theta)$, which is just $R(\theta)$ without existential quantification:

$$\not{R}(\theta):\quad \forall w\forall x\forall y\forall z((X_1w \wedge X_2x \wedge X_3y \wedge X_4z) \rightarrow (X_5w \wedge \neg X_5x \wedge \neg X_5y \wedge \neg X_5z \wedge X_6w \wedge X_6x \wedge X_6y \wedge X_6z))$$

We can then make use of $\not{R}(\theta)$ and the other tools we have developed to make the following claim, which is not trivial:

$$\exists X_1\exists X_2\exists X_3\exists X_4\exists X_5\exists X_6\mathbb{U}\llbracket\not{R}(\theta)\rrbracket\llbracket P\rrbracket$$

Based on the explanation of the example in the previous section, this turns out to be true even though $\mathbb{U}\llbracket\theta\rrbracket\llbracket P\rrbracket$ is false. We can see that via the following possible interpretation of terms:

$X_1x$: $x$ is an A-isomer.

$X_2x$: $x$ is a B-isomer.

$X_3x$: $x$ is a C-isomer.

$X_4x$: $x$ is a D-isomer.

$X_5x$: $x$ is very stable.

$X_6x$: $x$ is an instance of one of the four types of isomer which each represent a quarter of the molecules which constitute the substance.

This strategy is enough to defeat Newman's objection, but there is still the problem of background relations. This can be solved in a similar way as before: additionally claim that any logical connection found between the facts picked out by the terms $\mathcal{R}(\theta)$ and P in $\exists X_1 \dots \exists X_6 \mathbb{U}[\![\mathcal{R}(\theta)]\!][\![P]\!]$ can also be found between the facts picked out by the terms $\theta$ and P in $\mathbb{U}[\![\theta]\!][\![P]\!]$.[57] I do not see how this notion of logical connection could be represented formally except as a primitive. Including it in our formal claim as a primitive does not seem helpful or elegant, so I suggest leaving it as an informal addendum which only needs to be invoked to handle the problem of background relations.

This informal addendum might be avoided under these conditions: $R(\theta)$ contains some observational predicates (unlike in our example) and specifies appropriate logical relations between them and relevant non-observational predicates. When these conditions are met, the Ramsey sentence itself might effectively preserve at least some of the 'logical connection' between phenomena and non-observational theory, removing the need for an addendum to handle the problem of background relations. That being said, the qualification "appropriate" points to potential difficulties. It is not immediately clear, on this approach, what requirements or limitations, if any, there are on the observational predicates or the logical relations that must be included in the Ramsey sentence to ensure that the problem of background relations does not come up. Can the predicates be too vague to solve the problem, or not relevant enough?

[57] $\mathcal{R}(\theta)$ is an open formula, but the fact that $[\![\mathcal{R}(\theta)]\!]$ can be parsed in the full expression proves that it picks out a fact in that context even though it is not a declarative sentence in isolation.

How closely related must they be to the specified phenomenal fact? Can the relations be too easily realisable? My intuition is that strict rules cannot be drawn here, but I will not investigate this now. The informal solution should handle the problem of background relations where this formal solution might not. The important point for now is that these solutions to that potential difficulty exist.

To sum up, where:

$\theta$ is a sentence describing an unobservable system,

$R(\theta)$ is the Ramsification of $\theta$,

$\exists X_1 \ldots \exists X_n$ are the existential quantifiers binding the predicate variables in $R(\theta)$,

$\not{R}(\theta)$ is $R(\theta)$ with those existential quantifiers removed,

P is a phenomenal fact,

$[\![\theta]\!]$ means 'the fact that $\theta$', and

$\mathbb{U}[\![\theta]\!][\![\varphi]\!]$ means 'the fact that $\theta$ underlies the fact that $\varphi$',

We can make a non-trivial claim about the structure of that unobservable system using a sentence of the form $\exists X_1 \ldots \exists X_n \mathbb{U}[\![\not{R}(\theta)]\!][\![P]\!]$.

## 5.5 A structural realist in a vat

Now that we have seen how to make claims about the structure of an unobservable system, let us finally return to the question of how to make claims about the structure of a system without relying at all on knowledge about the observable world. We won't have to change too much. After I have explained how to do it, I will illustrate it by showing how a person who is unwittingly a brain in a vat experiencing a simulation created by mad scientists can still be a successful structural realist.

First, let's specify an alternative to Ramsey sentences. What we want are sentences exactly like Ramsey sentences except that *all* of the predicates of the original sentences have been replaced by variables, not just the 'theoretical' ones. In these sentences, the predicate variables are not existentially bound. Call them "Quietened Ramsey sentences" or "Q-Ramsey sentences" because, being open formulae, the do not actually say anything, and even if we did existentially bind the predicate variables, the sentences could say virtually nothing of interest on their own. We will represent the Q-Ramsification of θ as Q(θ). Note that in cases where θ does not contain observational predicates (as in the previous discussion), Q(θ) and $\mathcal{R}$(θ) will be identical.

Next, we need to define an operator, $\mathbb{a}$. $\mathbb{a}$(P): 'it appears that P'. With this and the definitions given above, and where $\exists X_1 ... \exists X_n$ are quantifiers binding the predicate variables in $Q(\theta)$, we can make use of statements of the form $\exists X_1 ... \exists X_n \mathbb{U}[\![Q(\theta)]\!][\![\mathbb{a}(P)]\!]$. This is a non-trivial statement describing the structure of a system which relies on no knowledge of reality besides what appears in experience and that there is a system of relations underlying what appears in experience. As a demonstration, I will show that this can enable even a brain in a vat to say something true (and non-trivial) about reality outside of their simulated experiences.

Jalesa is the scientist who created the model which explains the interaction between the substance and the liquid in terms of A-products, etc. A bit of a sceptic, she does not want to commit herself to knowledge of the nature of the unobservable objects presented in her model, she chooses to believe the structural realist claim $\exists X_1 ... \exists X_6 \mathbb{U}[\![\mathcal{R}(\theta)]\!][\![P]\!]$ instead of $\mathbb{U}[\![\theta]\!][\![P]\!]$. However, in this version of the thought experiment, there are no A-isomers, etc., either. In fact, neither the substance nor the liquid exists as a physical object. In this version of the thought experiment, Jalesa is – in the real world – a brain suspended in a vat of life-sustaining fluid who is (and has always been) experiencing a simulation that is running on a computer and is transmitted to her neurons through electrical impulses indistinguishable from those that would normally come from sense faculties. Thus, one might reasonably think that $\exists X_1 \ldots \exists X_6 \mathbb{U}[\![\mathcal{R}(\theta)]\!][\![P]\!]$ turns out to be false because the key terms in P fail to refer. If the reader is inclined to think that P remains true in these circumstances because 'the substance' and 'the liquid' will refer, here, to the simulated substance and liquid, not to imagined mind-external physical objects, then so much the better for me – my proposal will work as is with the interpretation of terms given below. But we can skip this potential point of contention by attending to the following strategy.

One day, Jalesa is convinced, after some philosophical reflection, that she should be even more sceptical than she normally is. She decides that she will no longer trust that the objects of her perceptions exist except as mere appearances in her consciousness. She notes that, while the relations presented in her model underlie P in the model, those same relations also underlie $\mathbb{a}(P)$ in the model according to our definition of 'underlies': In the world presented in the model, $\mathbb{a}(P)$ is true, θ is true, and $\mathbb{a}(P)$ is true because θ is true. In her model, then, one may pick either $\mathbb{U}[\![\theta]\!][\![P]\!]$ or $\mathbb{U}[\![\theta]\!][\![\mathbb{a}(P)]\!]$ as the point of departure for their structural realism. Having now opted for $\mathbb{U}[\![\theta]\!][\![\mathbb{a}(P)]\!]$, Jalesa can take the steps we outlined above to weaken this claim into the structuralist $\exists X_1 \ldots \exists X_6 \mathbb{U}[\![\mathcal{R}(\theta)]\!][\![\mathbb{a}(P)]\!]$. Satisfied, she decides that *this* is what she believes about the real system targeted by her theory.

On the present version of the thought experiment, the following is what is really going on. There is a team of mad scientists responsible for this unethical brain-in-a-vat research project. In a meeting, the mad principal investigator declared that the substance (which has no real counterpart outside of the simulation) must have four distinct tastes if Jalesa puts it in her simulated mouth. No one is sure why, especially since the substance is not a food, but such is the nature of this project. Unfortunately, the simulation engine only allows one to code one flavour per virtual object. The mad graduate student who had the task of coding the substance came up with the following workaround: every instance of the substance would actually be a continuous, even distribution of four objects, with the filenames sub1, sub2, sub3, and sub4, which should have identical properties other than their tastes. Incidentally, in the simulation, every object is assigned a ‘dissolution resistance value’ which, along with some other parameters, determines what liquids it will dissolve in when in a solid state. The substance is supposed to have a dissolution resistance value of 9, but the overworked mad graduate student accidentally hit the ‘9’ key twice when coding sub1, resulting in a value of 99 for that file only. The result in the simulation was that, when the substance (which is really a uniform distribution of sub1, sub2, sub3, and sub4) comes into contact with the liquid, the objects sub2, sub3, and sub4 dissolve as they are supposed to while sub1 does not.

We can also say that, at some point, for whatever reason, Jalesa tastes the substance before it has been exposed to the liquid and the residue of it left after exposure to the liquid. She finds that the substance originally has an overpoweringly complex taste, but the residue just tastes something like cherries. She finds that her theory goes a long way in explaining this fact: the residue is really made up of A-products, which makes it a distinct substance that might be expected to have other properties which differ from the original substance. Why not taste? Jalesa considers this a modest novel prediction[58] of her theory, which strongly adds to its credibility.

Jalesa’s claim, $\exists X_1 \ldots \exists X_6 \mathbb{U}[\![\mathcal{R}(\theta)]\!][\![ⓐP]\!]$, comes out true because of the following possible interpretation of terms:

$X_1x$: $x$ has the filename sub1.

$X_2x$: $x$ has the filename sub2.

$X_3x$: $x$ has the filename sub3.

$X_4x$: $x$ has the filename sub4.

$X_5x$: $x$ has a very high dissolution resistance value.

[58] Leplin (1997, 77) gives an account of novel prediction that can be used here.

$X_6x$: $x$ is an instance of one of the four virtual objects which constitute every instance of the virtual substance in even distribution.

The problem of background relations is deflected with the same strategy as before: an informal addendum stating that the logical connection between the facts picked out by $\mathcal{R}(\theta)$ and $\mathbb{a}$P in $\exists X_1 \ldots \exists X_6 \mathbb{U}[\![\mathcal{R}(\theta)]\!][\![\mathbb{a}(P)]\!]$ is approximately the same as the one between the facts picked out by θ and $\mathbb{a}$P in $\mathbb{U}[\![\theta]\!][\![\mathbb{a}(P)]\!]$. (Again, the link between the number of objects (4) and the proportion of the substance that appears to be left in residue (a quarter) satisfies this condition.)

We have here solved the harder problem first. Jalesa's structuralism has gone much further than Russell's original proposal. Russell's proposal was that there are, in objective reality, direct counterparts of the appearances given to us by perception. And we can know that those objective counterparts instantiate relations which satisfy the same structure as the relations instantiated by our appearances. Jalesa's true claim, however, goes beyond mere counterparts to appearances. She has figured out the structure of relations instantiated by the real system which underlies, but is otherwise utterly alien to, her experience. This demonstrates that, in order to answer perceptual antirealism, we are not forced to restrict the scope of structural realism. We can be structural realists about scientific theories in all their depth even if we are also merely structural realists about the rest of the external world.

That Jalesa can help herself to Russellian structuralism, too, is easy enough to see. If she sees a duck and thinks $\mathbb{a}$(D): "it appears that there is a duck", then she can claim $\exists X \mathbb{U}[\![\exists x(Xx)]\!][\![\mathbb{a}(D)]\!]$. We can do this with whatever appearances we like. Once we can report the appearances in a sentence, we can replace that sentence's predicates and elements with variables and feed these, along with the logical relations between them, into the first place of the two-place $\mathbb{U}$ predicate while the original sentence takes the second place, while including the appropriate quantifiers and operators as has been explained at length. This is how one delivers on the intuition behind Russellian structuralism without succumbing to Newman's objection. Whether that intuition is entirely true (is there an 'objective counterpart' to every object that appears in subjective experience? Do we have a good grasp on what that suggestion means?), this is the way to begin expressing it without fear of triviality.

To sum up as we did before, where:

$\theta$ is a sentence describing a system,

$Q(\theta)$ is $\theta$ in a second-order language in which existentially bound variables have been substituted for *all* predicates,

$\exists X_1 \ldots \exists X_n$ are existential quantifiers binding the predicate variables in $Q(\theta)$,

P is a phenomenal fact,

$\mathbb{a}(P)$ means "it appears that P",

$[\![\theta]\!]$ means 'the fact that $\theta$', and

$\mathbb{U}[\![\theta]\!][\![\varphi]\!]$ means 'the fact that $\theta$ underlies the fact that $\varphi$',

We can make a non-trivial claim about the structure of a system, even if our senses do not provide objective knowledge of the nature of reality, using a sentence of the form $\exists X_1 \ldots \exists X_n \mathbb{U}[\![Q(\theta)]\!][\![\mathbb{a}(P)]\!]$.

**5.6 Distinguishing this from the closest alternatives**

Now that my proposal is fully articulated, I expect that the reader will want to know how my proposal differs from other formulations of structural realism and what the consequences of these differences are for their potential to answer perceptual antirealism. I will distinguish my view from the closest alternative proposals in the literature. We will be looking at the proposals of Melia and Saatsi (2006) and Smithson (2017). Both of these proposals, according to their authors, are attempts to deflect Newman's objection by targeting an unobservable system via its causal connection to the observable world, so they are similar to mine in at least that respect.

On the question of whether their proposals could be adapted from answering scientific antirealism to answering perceptual antirealism as mine was (in moving from §4 to §5), I will show that this seems possible in principle, but there are some details and challenges that would need to be settled in a proper proposal. It will be made clear that these issues do not arise in my proposal and why. I will also indicate how the unique properties of my proposal give it some advantages over theirs even when confined to the arena of scientific antirealism.

Melia and Saatsi's proposal is that Ramsey sentences be augmented with intensional modal operators which have the Ramsey sentence declare a law-like relation between its predicates, such as $L_P$: "it is physically necessary that…". If these modal operators are allowed into Ramsey sentences, they can enable the articulation of some of the important causal/explanatory implications of the target system according to the original theory. By doing so, they deflect Newman's objection.

Smithson's proposal is based on the idea that realists want to claim that "(a) Certain unobservable relations satisfy a certain structure S and (b) The instantiation of these relations is supported by the No Miracles Argument" (2017, 1004). He proposes that this be articulated formally by augmenting Ramsey sentences with the inclusion of a statement of the form $\mathcal{N}[\![P]\!]$: 'the fact that P is such that the NMA provides (direct) evidence for the fact that P' in conjunction with P, where P is an expression with Ramsified predicates.

These proposals are meant to avoid Newman's objection and answer scientific antirealism. They were not designed to be applied to perceptual antirealism, so new issues can be expected to arise if we invoke them in this context. Since both of these proposals employ modified Ramsey sentences, the first step will be to modify a sentence in a second-order language so that its claims about observational predicates only amount to claims about appearances. But this quickly reveals difficulties in truthfully using material implication or equivalence to relate such a sentence's claims about appearances to any other part of the sentence.

I will illustrate the problem starting with a simple (don't mind whether it is true) theoretical claim, "stones are constituted only by atoms", restated in a sentence, S, in a second-order language:

T$x$: $x$ is a stone.

A$x$: $x$ is a set of atoms.

N$xy$: $x$ constitutes $y$.

(S) $\forall x \forall y((Tx \wedge Nyx) \rightarrow \mathrm{A}y)$

The natural first effort to make this about appearances will be to replace T$x$ with $\mathbb{a}$(T$x$) so that we get $\forall x \forall y((\mathbb{a}(Tx) \wedge Nyx) \rightarrow \mathrm{A}y)$. But there's immediately a problem. We don't want to say that something is constituted by atoms when it *appears* to be a stone. So, as I indicated in the previous section when explaining my own proposal, we should rephrase the original claim to be about appearances within a model of the same theory. The reader will recall that this was quite easy to do using 'underlies' as I have defined it. It's less clear how to do so without relying on that predicate.

How else might we link appearances with the objective system of relations described in our theory or model using a sentence like this? We cannot achieve this by simply saying $\mathbb{a}(\exists x Tx) \rightarrow \exists x Tx$ because it's not true that the appearance of the existence of a stone implies that such a stone exists. This also does not work if we reverse the direction of the implication, because the mere existence of a stone does not imply that it is appearing to someone as an existent stone. We could try instead $\exists x(\mathbb{a}(Tx) \rightarrow Tx)$. This seems reasonable enough and consistent with our theory, but it won't work for the purpose we have in mind. The perceptual antirealist must preserve the possibility that their perceptions are deceptive. It will not do to state that what appears to exist does in fact exist, which means we cannot state outside of what $\mathbb{a}$ is operating on that any element it operates on is part of the domain. That is, if $\mathbb{a}$ operates on any variable, it must also operate on any quantifiers binding that variable.

If one tries to work around this issue by introducing a second element along with an identity relation between the two elements for the model alone, $\mathbb{a}(\exists xTx) \rightarrow \exists zTz \wedge (x = \mathrm{z})$, then we will run into problems when the expression is Ramsified and the identity relation $x = \mathrm{z}$ is abandoned (as it must be, so that we don't encounter the problem we were just trying to avoid). $\exists \mathrm{X}(\mathbb{a}(\exists xTx) \rightarrow \exists zXz)$ does not say what we want it to say. We don't want to claim that the target relation is instantiated by some object whenever it appears that there is a stone.

Now, though, there is a promising way forward. It might work to say that the target relation is instantiated by some object whenever it appears that there is a stone *under the right conditions*. To go this route, we say that, in the model, there is a set of conditions C under which the appearance of a stone implies the existence of a stone. I am not going to attempt to suggest how the conditions in C might be stated, because that is not obvious to me, and this is not a proposal that I am developing. The point is that *if* one wanted to use a Ramsey-sentence-centric approach with material implication to answer perceptual antirealism in this example, one would need to list, in a second-order language, the members of C, such that if every member of C is true, it is true that $\mathbb{a}(\exists xTx) \rightarrow \exists zTz \wedge (x = \mathrm{z})$. We can then say the following (in a real proposal, C would need to be spelled out in a second-order language and its elements quantified over):

(S') $(C \wedge \mathbb{a}(\exists xTx)) \rightarrow \exists z \forall y(Tz \wedge (x = \mathrm{z}) \wedge (Nyz \rightarrow \mathrm{A}y))$

Next, we need to replace all predicates outside of expressions operated on by $\mathbb{a}$ with existentially bound variables. This is technically neither Ramsification nor Q-Ramsification as I have defined it, but I will call it "Ramsification" in scare quotes so that I can write about it for just the next page or so. The predicates involved in whatever the members of C are will need to be "Ramsified" away as well, of course, because the perceptual antirealist will not have knowledge of these objective conditions. I will represent "Ramsified" C as $C^X$ (the details elude us because I have no proposal for how C should actually be formulated). We also need to "Ramsify" the mixed relation N*yz* because the antirealist cannot be sure that *constituted by* is the true nature of the target relation. Finally, the identification of x with z is removed, as the perceptual antirealist will not assume that it is the very thing which appears to exist which instantiates the relevant relation. We get:

R(S') $\exists X_1 \exists X_2 \exists X_3 ((C^X \wedge 𝕒(\exists x Tx)) \rightarrow \exists z \forall y (X_1 z \wedge (X_2 yz \rightarrow X_3 y)))$

Now we can say, with Melia and Saatsi:

(MS) $\exists X_1 \exists X_2 \exists X_3 \mathbf{L}_{\boldsymbol{P}} ((C^X \wedge 𝕒(\exists x Tx)) \rightarrow \exists z \forall y (X_1 z \wedge (X_2 yz \rightarrow X_3 y)))$

Or, with Smithson:

(SM) $\exists X_1 \exists X_2 \exists X_3 ((C^X \wedge 𝕒(\exists x Tx)) \rightarrow \exists z \forall y (X_1 z \wedge (X_2 yz \rightarrow X_3 y)) \wedge \boldsymbol{\mathcal{N}} [\![ (\boldsymbol{C^X} \wedge 𝕒(\exists \boldsymbol{x} \boldsymbol{T} \boldsymbol{x})) \rightarrow \exists \boldsymbol{z} \forall \boldsymbol{y} (\mathbf{X_1} \boldsymbol{z} \wedge (\mathbf{X_2} \boldsymbol{yz} \rightarrow \mathbf{X_3} \mathbf{y})) ]\!])$

I have bolded their respective augmentations.

The following are issues I see arising from these approaches. I offer no opinion on whether they can or can't be resolved satisfactorily:

- For either approach, it is not clear what level of specificity or rigor is needed when identifying the set of conditions, C, under which an appearance materially implies the instantiation of a set of theoretical relations satisfying a certain structure, or vice-versa. Is it enough, in most cases, to simply specify that some *x* is being observed by someone with normal sense faculties? Or would one need a complex and detailed list of conditions in order to make the sentence true? For material implication in one of these sentences where the elements are under universal quantification, it only takes one scenario in which the antecedent is true and the consequent is false for the sentence to come out false. So, C needs to be airtight. If these approaches can be effectively used without relying on material implication/equivalence as the connection between appearances and the target relations (and without adopting the form of my proposal instead), this will have to be shown.

- For Melia and Saatsi's approach, there will need to be some thought put into what law-like relation between predicates is proposed and what the implications of that are when applied to appearances. For example, if one sticks with $L_P$: 'it is physically necessary that', there will need to be some commentary on appearances in the context of physical necessity.
- For Smithson's approach, all knowledge beyond appearances would be restricted to what the no miracles argument (NMA) provides direct evidence for. There would need to be a discussion on the parameters of that body of knowledge.
- Are either of these approaches capable of effectively supporting Russellian structuralism? We have seen that my approach does so naturally. It is less clear that law-like relations or the NMA are generally well-suited to make Russell's claim about objective counterparts to apparent objects instantiating relations forming the same structure as the apparent objects. The ability to make this sort of claim without triviality is highly significant from the standpoint of answering perceptual antirealism.

That concludes my estimation of the suitability of Melia and Saatsi's or Smithson's proposals to be adapted to perceptual antirealism. My proposal does not require anything like C because it is not making a claim about the conditions under which a certain appearance arises *in general*. Instead, the proposal is saying that for some *specific* appearance, that the instantiation of some specific relations underlies it (which is not a logical relation between the two, but an interpreted two-place predicate). This distinction creates instant relief from the pressure associated with the material implication route. For, even though there surely is a vast field of conditions that must be met for any given appearance to arise – and indistinguishable appearances may possibly arise in other conditions – we do not need to account for all that before expressing our belief that *this* appearance (say, of a stone) is one which is underlaid by the instantiation of *those* relations (say, quantum physics) rather than some other ones (say, neurological ones in dreaming). In the other direction, we may say, informed by theory, that the instantiation of *these* relations underlies *some* appearances fitting some description.

Before concluding this section, I would like to briefly note two advantages (one for each rival) that my proposal has over Melia and Saatsi's and Smithson's proposals even when the conversation is limited to answering scientific antirealism alone. This is to help make it clear that the proposal I have developed is not only useful for bringing structural realism into the realm of perceptual antirealism, but it has value as a novel approach to structural realism even when we limit our concerns to those of the contemporary scientific structural realism literature.

First, I will identify an advantage it has over Melia and Saatsi's proposal as scientific structural realism. I follow Smithson's (2017) own criticism of Melia and Saatsi here. Smithson points out that important claims made using modal operators such as "it is physically necessary that…" have been shown to be *lost* in theory change even in instances where important structure is preserved. Smithson's key example is gravitation in the move from Newtonian mechanics to general relativity. While Newton's equations approximate limiting cases of relativistic equations, "this is a case where claims about nomic necessity were actually *discarded* across theory change" (p.999. Emphasis in original). Smithson points out that someone presenting the content of Newton's theory in the way that Melia and Saatsi suggest would have regarded it as physically necessary that any two massive bodies obey Newton's law of gravitation. However, "we now understand that Newton's law breaks down when we consider extremely massive bodies" (ibid.). Given that this is a paradigmatic example of the preservation of structure across theory change, this is a worrying observation for Melia and Saatsi's proposal.

Using my proposal, however, this example of theory change is easy to handle. Take any phenomenal fact involving gravitation $\mathrm{P}_G$ that is explained by Newton's theory (we are bracketing perceptual antirealism). We can take a model of Newton's theory and identify a sentence based on it in a second-order language, $\theta_N$, that underlies $\mathrm{P}_G$ based on that model. A structural realist in Newton's time would say $\exists \mathrm{X}_1 \ldots \exists \mathrm{X}_\mathrm{n} \mathbb{U}[\![\mathcal{R}(\theta_N)]\!][\![\mathrm{P}_G]\!]$ based on the proposal I defined at the end of §4. Now, stipulating that $\mathrm{P}_\mathrm{G}$ pertains to a context in which Newton's (ultimately false) theory is a limiting case of General Relativity (GR), we know, by virtue of this fact, that there is a model of GR which subsumes the relations mentioned in $\theta_N$, and that we can therefore identify another sentence based on the new model, $\theta_R$, which is isomorphic to $\theta_N$, and underlies $\mathrm{P}_G$ according to the new model. Based on the GR model, a structural realist will say $\exists \mathrm{X}_1 \ldots \exists \mathrm{X}_\mathrm{n} \mathbb{U}[\![\mathcal{R}(\theta_R)]\!][\![\mathrm{P}_G]\!]$, but since $\theta_N$ and $\theta_R$ are isomorphic, this sentence will be identical to $\exists \mathrm{X}_1 \ldots \exists \mathrm{X}_\mathrm{n} \mathbb{U}[\![\mathcal{R}(\theta_N)]\!][\![\mathrm{P}_G]\!]$. So, the structural realist from Newton's time will still be right according to the new theory.

As for Smithson's proposal, it is only able to make non-trivial claims about the structures of unobservable systems insofar as it claims that the NMA provides direct evidence for the instantiation of the relations which satisfy those structures. This has the potentially unfortunate consequence that the claim will be entirely without interest to anyone who thinks that the NMA is unconvincing in general, or anyone who disagrees with the structural realist on what the NMA gives evidence for. This will not likely perturb someone who thinks that the NMA is the one and only reason one could have to make claims about the structure of an unobservable system. In that case, it might even be helpful to formulate one's claims this way so that focus is maintained on the crux of the matter. However, I think it is valuable to have a way to make these claims independent of one's dominant motivation for believing them. This potentially enables more productive engagement with people who have different commitments (whether dramatically or only slightly different) to one's own and opens the door to the potential development or discovery of new kinds of evidence which support the claim. My proposal allows for this.

It should now be clear what separates my proposal from the closest existing alternatives for scientific structural realism. I have indicated how one might attempt to begin adapting these to answer perceptual antirealism and identified some issues that one taking on that project would need to settle. These issues do not arise for my proposal in the context of perceptual antirealism, and I have shown that my proposal offers some advantages as an answer strictly to scientific antirealism as well.

### 5.7 Conclusion

Newman's objection was published nearly a century ago. Since that time, I know of no one who has defended the plausibility of Russell's original idea – that knowledge of the external world is limited to its structure and its role in causing appearances. I have argued not only that Russell's idea can be formulated without triviality but that it is capable of answering even the most radical scepticism. I have shown, too, that this idea can be taken well beyond the scope of Russell's original suggestion: Russell suggested only to infer the structure instantiated by the objective entities corresponding to objects of perception, but we do not need to heed any such condition. It is possible to make non-trivial claims about the structure of any real system that underlies experience; there is no need for this system to be constituted even partly by counterparts to objects of perception.

Russell's original language suggested that he intended his structuralism to answer European idealism in the tradition of Kant. One will naturally wonder, then, whether my proposal here can work within the framework of Kant's system to make possibly true claims about the structure of the noumenal world. This is, of course, too big a question to be adequately treated here, but the paper would be incomplete without at least a few words on the matter. I would say that, *insofar as* noumena are regarded as objective causes or conditions for some of the features of experience, I share Russell's intuition that, simply by virtue of this causal role, they must, in some transcendent way, satisfy in the relevant way at least some of the structures that can be identified over experiences. We will discuss this further in Chapter Seven, where I work at figuring out how this type of sentence might be compatible with Madhyamaka despite seemingly asserting the existence of worldly things. Before that, I will settle, in the next chapter, why I think that ontic structural realism is not a viable alternative.

CHAPTER SIX

# Neither Individuals nor Relations? A criticism of ontic structural realism from a naturalistic stance

I argue that the positive metaphysical content of ontic structural realism – the claim that structure is ontologically fundamental – is indefensible from a naturalistic perspective. My strategy is to show that the evidence and arguments advanced by ontic structural realists to motivate their positive thesis underdetermine the choice between it and another, contradictory thesis. I argue that there is no apparent way to break this underdetermination without adopting an anti-naturalistic approach to the human mind. The negative metaphysical content of ontic structural realism, however – the elimination of fundamental individuals with intrinsic nature – is untouched by this criticism and may be defended independently.

## 6.1 Introduction

The structural realism discussed in the previous chapter is of the sort described as epistemic by Ladyman (1998). In that paper, Ladyman introduced the idea of ontic structural realism. Where epistemic structural realism claims that we can only know the structure of target systems (whether that is some or all unobservable systems, or all experience-external systems, as I argued in the previous chapter), ontic structural realism (OSR) is the view that all of reality, at the fundamental level, *is* ontologically subsistent structure. This is a dramatically different thesis and needs to be treated on its own terms. Ladyman and Ross's *Every Thing Must Go* (2007), which we discussed in Chapter Three for its strong position on naturalism, is a defence of OSR. Steven French (2014) has another influential book-length defence of the idea, with his own spin on it. And there are various other permutations of the view in the literature with papers discussing and developing them (e.g., Esfeld and Lam 2010, Tegmark 2008, Thébault 2016, McKenzie 2017).

OSR is a particularly tricky subject from a Madhyamaka point of view. Nāgārjuna's argument – however one interprets it – can clearly apply in principle to worldly things in general. However, there is a presumption that worldly "things" means worldly *objects* (the kinds of things that one would think might have intrinsic nature), and OSR has already eliminated all objects with intrinsic nature from its fundamental ontology. It is not at all clear what purchase Nāgārjuna's arguments could have here. But OSR advances a detailed characterization of the objective nature of ultimate reality, something that definitely runs afoul of any interpretation of the Madhyamaka philosophical-soteriological project. Some work needs to be done here, and I think there are four possibilities:

1. Argue that existing arguments within the Madhyamaka tradition can be retooled to refute OSR.
2. Attempt to refute OSR with an argument or arguments that have not previously appeared in the tradition but are compatible with it.
3. Argue that OSR is actually compatible with Madhyamaka.
4. Argue that OSR is incompatible with Madhyamaka and cannot be successfully refuted from a Madhyamaka point of view (OSR wins).

I do not think that (3) is workable – OSR provides a firm realist ontology according to which we can say that things exist and which presents itself as an accurate description of the nature of reality (*tattva*). I don't have any ideas for how (1) would work. And I do not think I am forced into (4) because I believe that (2) can be pursued. Pursuing (2) is what I will do in this chapter. I will argue that OSR does not work as a positive metaphysical thesis. The value of the OSR project, in my opinion, lies in the *negative* metaphysical thesis that is characteristic of it.

I will argue that all of the potentially effective arguments advanced by OSRists are in support of this negative thesis alone, while the positive thesis they put forward has no plausible naturalistic defence. Accordingly, the conclusion which I hope the reader takes away from this paper is that the positive metaphysics of OSR is a dead-end even if its negative metaphysics is worthy of close attention.

### 6.2 Underdetermined ontology

I will proceed by reconstructing OSR into three distinct theses which I take as essential perhaps to any position calling itself OSR, but certainly at least to any OSR meant to be kindred to the eliminative varieties defended by James Ladyman and Steven French, whose works will be the points of reference for my argument:

**Epistemic Thesis**: We have at least approximately represented the structures of certain real, unobservable systems in the models of our best scientific theories.

**Eliminative Thesis**: Belief in the existence of ontologically fundamental individuals with intrinsic nature is poorly motivated and should be abandoned.

**Ontological Thesis**: Real systems ultimately consist of relations alone.

In this paper, I will be taking the Epistemic and Eliminative Theses as given and arguing that we do not have a good reason, from a naturalistic point of view, to accept the Ontological Thesis as true.

First, a note on the definition of structure and the relationship between structure and relations. In the relevant literature, "structure" sometimes seems to be used simply as a collective noun for a network of relations[59], and sometimes has the more technical sense of referring to an abstract pattern which may or may not be instantiated by any given network of relations. In any case, the notion of "structure" is inseparable from that of "relation". While there are different shades of OSR, what unites them is the belief that structure is ontologically primary (see McKenzie 2017). This means that, for OSRists, all real systems have their ultimate metaphysical ground in relations. If they allow for the existence of objects at all, those objects are ontologically dependent on relations. This is what gives us the Ontological Thesis. On this view, the distinction between the mathematical structures used by scientists and the physical systems which they are supposed to represent is tricky to clarify (see French 2014, Chapter 8) but that issue need not concern us here. I will be referring to real systems, which all hands agree instantiate structures via their internal relations. The difference between OSRists and their opponents is that, for OSRists, to represent the structure of a real system in fundamental physics is to represent that system in its entirety, while their opponents think that there is more to what a real system fundamentally *is* than just that.

Let us touch briefly on the supporting arguments for the first two theses. The Epistemic Thesis can be defended with the no miracles argument, which would have it that the success of our best theories would be a miracle if it were not true. The Eliminative Thesis has been defended through appeal to physics. It's been argued that the putative entities which appear in fundamental physics are too ambiguous in their metaphysical status to be regarded as real individuals with their own intrinsic nature. If we want to insist that there are individuals in fundamental reality, the argument goes, there isn't an unproblematic way to identify them when

[59] An example from French and Ladyman (2003, p.46): "…these relations are not supervenient on the properties of unobservable objects and the external relations between them, rather this structure is ontologically basic."

interpreting modern physical theories. OSRists therefore argue that an individual-less ontology is more appropriate for the interpretation of those theories (Ladyman 1998; Ladyman and Ross 2007, Chapter 3; French 2014, Chapter 2).

Again, we are not going to ask whether these are good arguments or not. Let's just assume that they work and that the Epistemic and Eliminative Theses are true. What, then, motivates the Ontological Thesis?

Taking a look at that seminal paper of Ladyman's (1998), here's the key passage where he moves from the Eliminative Thesis to the Ontological Thesis:

> *It is an ersatz form of realism that recommends belief in the existence of entities that have such ambiguous metaphysical status.* What is required is a shift to a different ontological basis altogether, one for which questions of individuality simply do not arise. Perhaps we should view the individuals and nonindividuals packages, like particle and field pictures, as different representations of the same structure. There is an analogy here with the debate about substantivalism in general relativity. Recently it has been suggested that this issue also calls for a different fundamental ontology within which to assess the reality of spacetime. Robert DiSalle (1994) has suggested that the structure of spacetime be accepted as existent without being supervenient on the existence of spacetime points. This is a restatement of the position developed by Stein in his famous exchange with Grünbaum, according to which spacetime is neither a substance, nor a set of relations between substances, but a structure in its own right.
>
> So we should seek to elaborate structural realism in such a way that it can diffuse the problems of traditional realism, with respect to both theory change and

underdetermination. *This means taking structure to be primitive and ontologically subsistent.* (Ladyman 1998, p.420. Emphasis mine.)

I have put in italics the part representing the Eliminative Thesis at the start of the passage and the part representing the Ontological Thesis at the end (if real systems are fundamentally composed of primitive and subsistent structure, then they ultimately consist of relations alone). Between them, Ladyman states that we need an ontology which accommodates the Eliminative Thesis, points to some earlier work suggestive of the Ontological Thesis, and says that adopting the Ontological Thesis enables structural realism to diffuse the problems of underdetermination and discontinuity in traditional realism. The idea seems to be that we want to eliminate objects from our fundamental ontology due to underdetermination in physics, and a fundamental ontology of only structure achieves that elimination while accounting for theory change,[60] so we ought to adopt such an ontology.

But there is a problem here if this choice of ontology is itself underdetermined by the motivations provided for it. I will argue that this is indeed the case. For while the Ontological Thesis is consistent with the Eliminative Thesis, so too is this thesis which I call the Neither-Nor Thesis:

**Neither-Nor Thesis**: Real systems ultimately consist neither of individuals and their relations nor of relations alone.

On this thesis, whatever ultimately realises the structures of real systems cannot be adequately captured by human conceptual schema.[61] OSR requires us to believe that the Ontological Thesis

[60] It is supposed to account for theory change because, according to structural realists, the structure of successful scientific theories is preserved across even the most radical theoretical changes.

[61] I am assuming that human conceptual schema must necessarily employ individuals or relations to represent systems. I am open to the possibility that this is not true, but exploring that possibility would be a major digression which could not affect my case. My arguments are designed to show that real systems might not ultimately consist of individuals or relations and might not be cognizable. If they also might not ultimately consist of individuals or relations and yet be cognizable, then that can only be a bonus.

is true and the Neither-Nor Thesis is false. I will argue that there are no good naturalistic reasons to believe that.

There is a point of clarification that I cannot emphasise too much: I am not trying to convince the reader that the Neither-Nor Thesis is true. My goal is rather to show that we should not be committed to its falsity – not if naturalism is important to us – and therefore should not be committed to the truth of the Ontological Thesis (since they are contradictory, holding one of these theses to be true is committing to the falsity of the other). I will argue that we do not have good naturalistic reasons to determine which of these theses is true and which is false, and that we therefore ought not to make a determination one way or the other. Even if we agree with the Epistemic and Eliminative Theses, when asked what real systems ultimately consist of, the truthful answer is "we don't know". That, of course, brings us back to something which could be described as an epistemic form of structural realism (if we want to call it a form of realism at all). It offers no positive metaphysical description of fundamental reality, though it would contain the negative metaphysical claim of the Eliminative Thesis.

I will proceed by looking at potential objections and at a couple of the key principles used by OSRists to motivate their position. I'll show that none of these should motivate us to think that the Ontological Thesis is true and the Neither-Nor Thesis false.

### 6.3 Two-headed gerbils and the burden of proof

I begin with what I suspect might be many or most OSRists' first reaction to the case I am making here, which is informed by this passage from *Every Thing Must Go*, Ladyman and Ross's book-length defence of OSR:

> Should we reject the existence of things in which we could have no reason to believe, or suspend judgement about them? Perhaps the latter is the more

> enlightened option, but then we ought to be agnostic about a literal infinity of matters—whatever anyone can conceive without contradicting physics. Should we be agnostic about the existence, somewhere, of two-headed gerbils that sing the blues? If the agnosticism a philosopher advises us to take up towards the unknowable noumenal objects is strictly analogous to this then we are sanguine about agreeing to it. (Ladyman and Ross 2007, p.131)

This is a concession they are making for the Eliminative Thesis. They are saying that they are happy to accept that they're actually agnostic about the existence of unknowable individuals, but they take this as strictly analogous to agnosticism about the existence of two-headed, blues-singing gerbils. The effect is that they are not declaring *certainty* that there are no individuals with intrinsic nature out there, but are saying that they don't see a reason to take seriously the claim that there are such individuals.

Some might want to adopt the same attitude toward the Neither-Nor Thesis. They might be happy to admit that it's *possible* that the Neither-Nor Thesis is true and the Ontological Thesis false, but would add that this is not a possibility that they are willing to take seriously in their metaphysics because the Neither-Nor Thesis is concerned with unknowable realities which we couldn't even begin to conceptualise in order to ask whether they exist or not. So, the response would be, "technically we're agnostic about the Neither-Nor Thesis, but practically we'll reject it – we won't take it seriously".

But this is an inappropriate response, particularly from a naturalistic point of view. There are two basic reasons why it is inappropriate:

1. There are good naturalistic reasons to take the Neither-Nor Thesis seriously in metaphysics (even though I don't think we should be committed to its truth).
2. There are no good naturalistic reasons to place the burden of proof on the Neither-Nor Thesis when contrasted with the Ontological Thesis.

The good naturalistic reasons for taking the Neither-Nor Thesis seriously come from a naturalistic consideration of human cognitive capacities. That means treating the human mind as a subject of scientific inquiry and aspect of the natural world which enjoys no special *a priori* privileges over any other. As natural objects, we take human minds to fall within a fixed range of constitutions and abilities, and our beliefs about that range is informed by our best science primarily if not exclusively. Please note that by "natural object" I do not mean "physical object" or any such metaphysical description – I just mean that the mind is a part of nature (however "nature" is construed) and we should draw conclusions about it accordingly. One of those conclusions is that science is the best way of learning about the mind because science is the best way of learning about nature.

From this point of view, human cognitive abilities are a subset of the abilities of a kind of animal which evolved on the African continent. Every kind of animal in existence gained its entire range of abilities, as far as we know, from millions of years of mutations in combination with natural and sexual selection pressures. Given this context, my charge is that it is unjustifiable to *expect* that this or any animal is capable of representing every aspect of the fundamental nature of reality. We certainly have no reason to think that the range of human cognitive abilities *necessarily* corresponds to the range of possible ways that fundamental reality might be. Nor do we have good reasons to think that we have already successfully represented what fundamental reality is like.

The no miracles argument cannot make any difference here. It can't convince us that we have represented the fundamental nature of actual systems because the Epistemic Thesis adequately explains the success of science and is equally compatible with both the Ontological and Neither-Nor Theses. This is part of the original selling point of structural realism as proposed by Worrall (1989): it is supposed to solve the tension between the no miracles argument and the pessimistic meta-induction (Laudan 1981) by abandoning any pretence of describing the fundamental nature of the world and cleaving to structure instead. OSR is an entirely different story because it claims that structure *is* the fundamental nature of the world.[62] This metaphysical claim is clearly not itself an explanation of what makes any particular scientific theories successful and therefore cannot be directly supported by the no miracles argument (nor do OSRists claim that it can be). The relevant argument is rather that this claim is unfazed by the pessimistic meta-induction since structure is preserved across theory change. As I have indicated, the Ontological Thesis is positively motivated simply by the want for a positive metaphysical thesis consistent with both the Epistemic and Eliminative Theses. My argument is that this motivation underdetermines the choice between the Ontological and Neither-Nor Theses.

Returning to our naturalistic reasons to take the Neither-Nor Thesis seriously, we can ask: if it were the case that the fundamental nature of reality is beyond our cognitive abilities, what might we expect physics to look like? I would suggest that we would expect physics to be pretty baffling. It would be full of results which violate our intuitions, and the metaphysical interpretation of theories would be extremely difficult as we attempt to apply our conceptual schema to mathematical results which just don't fit easily within them. This, of course, is just the situation we are in with physics.

To summarize: on a naturalistic approach to the mind, we would not expect that the human mind matches up perfectly with the fundamental nature of reality. It would be quite the coincidence – or, you could say, a miracle – if our minds just happen to be capable of grasping the fundamental nature of reality. And secondly, empirically, physics looks very much like what

[62] One advantage that this claim gives OSRists over Worrall is that it neutralises Newman's (1928) objection for them. That is because OSRists do not shy away from describing the nature of the relations which they take to be forming the worldly structures described in successful scientific theories. There is therefore no question of the existence of those worldly structures being trivial.

we would expect it to look like if our minds aren't capable of grasping the fundamental nature of reality.

Here, of course, the OSRists might point out that physics also looks just like how we would expect it to if the Ontological Thesis were correct. On their view, we have represented the fundamental nature of reality because it turns out that all we needed in order to do that was the ability to do mathematics. According to them, fundamental reality matches various mathematical structures up to isomorphism, and when we have represented reality with those structures there is nothing that we're missing, nothing more there that we have failed to represent.

So, the key question here is whether we should simply assume that this one species of animal, by virtue of being able to do mathematics, has everything it needs to represent the subtle fabric of reality in its entirety, missing nothing. Or, whether we should actively suspect that, while mathematics has enabled the animal to represent the structures of real systems, the animal might not have the cognitive tools to pick up on every aspect of those systems' natures. Because the latter option fits comfortably with naturalistic expectations, it cannot be considered analogous to the idea that we should actively suspect that there are two-headed, blues-singing gerbils somewhere. If the OSRist wants us to reject it, they must give us some good reason to do so.

This brings us to the second part of my present case. Given that the truth values of the Ontological and Neither-Nor Theses are underdetermined by the evidence, an OSRist will want to place the burden of proof on the Neither-Nor Thesis. But I see no legitimate reason to do this, from a naturalistic stance.

We should acknowledge that both of the theses in question are very weird. OSRists rightly don't let critics dismiss their view on the grounds that it is weird, since our intuitions shouldn't be made the arbiters of what is acceptable metaphysics or not. Consider the following quotations:

> [I]maginability must not be made the test for ontology. The realist claim is that the scientist is discovering the structures of the world; it is not required in addition that these structures be imaginable in the categories of the macroworld (McMullin 1984, p. 14).
>
> The demand for an individuals-based ontology may be criticised on the grounds that it is the demand that the structure of the mind-independent world be imaginable in terms of the categories of the world of experience. (Ladyman 1998, p.422)

The first, from Ernan McMullin, has been quoted approvingly in multiple places to help vindicate OSR (Ladyman 1998, p.422; Ladyman and Ross 2007, p.132) and the second is Ladyman applying the idea to OSR.[63] Clearly, then, the OSRist has no grounds on which to be biased against the Neither-Nor Thesis because it is weird or counter-intuitive. However, we could imagine someone arguing that while both the Ontological and Neither-Nor Theses are weird, the former is significantly less weird than the latter in that it at least preserves a portion of our core metaphysical intuitions. Moreover, it might be argued that there is at least a sense in which the Ontological Thesis is intuitable while the Neither-Nor Thesis seems impossible to intuit whatsoever. Because of these points, it might be argued that the burden of proof should lie on the Neither-Nor Thesis because of its *relative* counterintuitiveness.

How does that work? It might be argued that humans intuitively believe the following conjunction:

Conjunction: Reality is fundamentally composed of individuals AND relations.

This, of course, is rejected by OSRists. However, OSRists maintain and rely on the truth of a closely related proposition – this one disjunctive:

[63] McMullin was not himself defending OSR in the original source.

Disjunction: Reality is fundamentally composed of individuals OR relations.

The Ontological Thesis follows from this disjunction plus the Eliminative Thesis. The negation of the disjunction gives us the Neither-Nor Thesis. Since this disjunction is much more in line with our default outlook (the conjunction) than its negation is, it could be argued that the burden of proof should lie with someone who negates the disjunction, not with someone who upholds it.

First of all, we can see that this is not a naturalistic argument. It still relies entirely on appeal to intuition and that, from a naturalistic perspective, should not be the deciding factor in what we believe the world is like. But moreover, when we approach the same controversy (i.e., whether to pick the Ontological Thesis or Neither-Nor Thesis) from a strictly naturalistic perspective, the landscape starts to look very different. Keeping in mind the naturalistic approach to the human mind I presented before, we can restate the Ontological and Neither-Nor Theses in painfully naturalistic terms as follows:

a) Real systems ultimately consist only of aspects of reality whose nature corresponds to the concept of 'relation' as found in human minds.
b) Real systems are ultimately constituted in ways that human minds are unable to comprehensively represent.

Note that (b), the Neither-Nor Thesis, does not deny that relations are involved in realising structure; what it denies is that relations are all there are at the ultimate level of analysis. And here, (a) strikes me as a weirder commitment to have than (b). We might want to choose (a) because it preserves our ability to understand reality where (b) does not, but what we *want* to be the case should be irrelevant in good metaphysics. We should also note that (a) is the stronger of the two claims in the sense that it specifies the ultimate nature of reality (ruling out all possibilities except one) while (b) only rules out a fixed range of possibilities, allowing for indefinite others. One might want to place the burden of proof on (a) for this reason. I am not going to press that point because I’m not attempting to put the burden of proof on either of them. I mention this only to show that there is at least as much reason to be suspicious of the Ontological Thesis on *a priori* grounds as there is the Neither-Nor Thesis.

All of that serves to show that adopting a dismissive agnostic attitude to the Neither-Nor Thesis in line with Ladyman and Ross’s two-headed gerbil analogy is inappropriate. The Neither-Nor Thesis is something that a reasonable person might expect to be true on naturalistic grounds, and while I don’t think we should be committed to its truth, we do have a responsibility to take the possibility of its truth seriously when forming our metaphysical commitments. After all, we don’t want a situation wherein metaphysicians simply dismiss without argument reasonable possibilities which contradict their own positions. So, we are still owed some reason from the OSRists to believe that the Ontological Thesis is true and the Neither-Nor Thesis false.

We will now move onto some key principles used by OSRists to which they might hope to appeal in order to positively motivate the Ontological Thesis.

### 6.4 Faraday’s Principle and the Principle of Naturalistic Closure

One principle to which an OSRist might want to appeal in order to counter my argument is that metaphysics ought to be limited to the content of science. They might say, “relations appear in mature scientific theories, including in fundamental physics, while vague *stuff* that we can’t conceptualise does not. We are only willing to entertain ontologies which align with what is presented in science, so we refuse to entertain the Neither-Nor Thesis but will entertain the Ontological Thesis.”

This is similar to a principle that Steven French adopts from Faraday and puts to use as motivation for OSR. He quotes Faraday as saying: "Why then assume the existence of that of which we are ignorant, which we cannot conceive, and for which there is no philosophical necessity?" (French 2014, p.59) Let's restate this as a thesis instead of a rhetorical question and call it Faraday's principle:

> **Faraday's Principle:** We should not assume the existence of that of which we are ignorant, which we cannot conceive, and for which there is no philosophical necessity.

French makes use of this by arguing that what he calls "object-oriented realism" violates this principle. I have no quarrel with that. In fact, I agree entirely with this principle. That is why I don't think anyone should commit to the truth of the Neither-Nor Thesis. What I am arguing is that we should not be committed to its falsity. To argue against *that*, we would need a different, more aggressive principle:

> **FP intensified:** We should assume that nothing exists of which we are ignorant, which we cannot conceive, and for which there is no philosophical necessity.

Adopting *this* principle would commit us to the falsity of the Neither-Nor Thesis and counter my argument. But this intensified principle strikes me as alarmingly problematic in general, and especially so from a naturalistic stance. As I have already explained, a naturalistic philosophy of mind sees the human mind as a natural object without special *a priori* metaphysical privileges. What would possibly lead us to believe that this natural object *necessarily* contains within it the ability to accurately represent the fundamental nature of reality, no matter what that nature is? It is impressively anti-naturalistic to make the human mind the measure of reality as an *a priori* philosophical principle.

That point holds even if we believe that mathematics enables us to represent the structure of any possible system – a view to which Ladyman and Ross gesture (2007, p.2). Even if that is true, it doesn't follow that we're necessarily capable of representing anything non-structural which might have a role in the constitution of real systems.

So, while it's fine for a naturalist to say that they won't positively believe that anything exists which isn't attested to in science, it is blatantly anti-naturalistic for them to insist that nothing exists which can't appear in science. That is incompatible with a naturalistic approach to the mind, as I have explained. Doing science is a human ability, and we have absolutely no naturalistic reason to believe that the nature of reality is necessarily restricted to the cognitive abilities of any species of animal.

Finally, let us consider Ladyman and Ross's particular brand of naturalism within which they present OSR. It revolves around what they call the Principle of Naturalistic Closure (PNC):

> Any new metaphysical claim that is to be taken seriously at time *t* should be motivated by, and only by, the service it would perform, if true, in showing how two or more specific scientific hypotheses, at least one of which is drawn from fundamental physics, jointly explain more than the sum of what is explained by the two hypotheses taken separately… (Ladyman and Ross 2007, p.37)

They really do mean 'taken seriously' here: if a metaphysical claim does not satisfy this exact purpose, they think it should be dismissed. So, if the Neither-Nor Thesis doesn't satisfy this explanatory demand, then someone who adopts the PNC could reject the Neither-Nor Thesis for that reason alone.

I do not think that the Neither-Nor Thesis satisfies the demands of the PNC, but I observe that *the Ontological Thesis doesn't satisfy the demands of the PNC either.* If we look at how Ladyman and Ross utilize the PNC to defend OSR in Chapter 3 of their book, we find that the metaphysical claim that they are defending due to its explanatory value is the *Eliminative* Thesis. They discuss quantum and spacetime physics at some length, but their conclusion always takes the form 'this makes better sense if we eliminate objects' such as in this quote toward the end of their section on spacetime:

> [I]t seems that the insistence on interpreting spacetime in terms of an ontology of underlying entities and their properties is what causes the problems for realism about spacetime. (Ladyman and Ross 2007, p.143)

This, of course, is motivation only for the Eliminative Thesis. I have found only one passage where Ladyman and Ross attempt to defend the Ontological Thesis as distinct from the Eliminative Thesis:

> If science tells us about objective modal relations among the phenomena (both possible and actual), then occasional novel predictive success is not miraculous but to be expected. Furthermore, the fact that scientific theories support counterfactual conditionals is also explained. Provision of these explanations is not a matter of satisfying philosophical intuitions, but of unifying scientific practices and theories. We thus suggest that in addition to the negative thesis that physical theory should not be interpreted in terms of underlying objects and properties of which the world is made, we are motivated in accordance with the PNC to take seriously the positive thesis that the world is structure and relations. (Ladyman and Ross 2007, p. 153)

Here, they appeal to the no miracles argument to insist that science is teaching us about real, objective modal relations, and explain how they see this as justifying the Ontological Thesis in accordance with the PNC. But, on my formulation, the notion that science is discovering objective modal relations, and every implication of that notion – from the no miracles argument to "unifying scientific practices and theories" – is captured by the Epistemic Thesis and is consistent with the Neither-Nor Thesis. We can believe that science is teaching us about the objective modal relations of real systems without attempting to describe the ultimate nature of those systems. So, the choice between the Ontological and Neither-Nor Theses is underdetermined here, as well.

Perhaps someone would suggest that the PNC will have us prefer the Ontological Thesis because the Neither-Nor Thesis posits additional aspects of reality which stand in need of explanation themselves, so it could be considered a worse explanation in that sense. But the additional realities posited by the Neither-Nor Thesis, *ex hypothesi*, can't be explained. So, they don't stand in need of explanation. If someone were to counter that the Neither-Nor Thesis is a worse explanation precisely because it posits things that can't be explained, again we reply that we have naturalistic reasons to suspect that there may be things which can't be explained. Therefore, we certainly can't disqualify or be otherwise prejudiced against an explanation solely for suggesting that there are things which can't be explained. If it were a consequence of the PNC that we should be so prejudiced, then the PNC is not as naturalistic as it seems. But I don't think that we need to challenge the naturalistic credentials of the PNC here. The PNC, as written, simply doesn't give us any reason to believe either the Ontological or the Neither-Nor Thesis. This, I think, is a mark in its favour.

### 6.5 Conclusion

We don't have any reason, from a naturalistic point of view, to think that the Ontological Thesis is true and the Neither-Nor Thesis is false, nor can we responsibly ignore the Neither-Nor

Thesis. If we are convinced by the Epistemic and Eliminative Theses and are asked what real systems ultimately consist of, the honest and responsible answer is “we don’t know”.

This does not leave us with a structural realism that is purely epistemic. The Eliminative Thesis is a metaphysical claim – but it is a purely negative one: it tells us what metaphysical commitments we should *not* have. The Ontological Thesis is the positive metaphysical content of OSR. Without it, we are left with a structural realism which encourages a radical revision of our intuitive picture of the world’s constitution, but which does not replace this picture with another one. The question should not be whether this satisfies us, but whether we are justified in anything more.

CHAPTER SEVEN

# Conclusion: No Miracles – Sacred or Profane

## 7.1 Introduction

This final chapter finally attempts to show what I said I wanted to show in the Introduction, using all of the resources I have developed in this work. I want to show that both scientific antirealists and realists can account for the success of science in ways that are consistent with Madhyamaka. In Chapters One through Three, I explained what I take Nāgārjuna to be arguing for, how he argues for it, and what kind of philosophy I take his to be. The bottom line is that he is an ontological conventionalist: he does not think that any sentence claiming that a worldly thing exists has a truth value. Chapters Four through Six contained various discussions of efforts to connect theories to reality without having to be committed to the existence of worldly objects. I explained Poincaré's ontological conventionalism and phenomenological relationism and found that it was appropriate as an antirealist account of scientific knowledge consistent with Madhyamaka. A realist account was still to be found. I also rejected as unjustified the ontic structural realists' claim that relations are the fundamental nature of reality. I showed it is possible to make true claims about otherwise epistemically inaccessible systems via their causal connection to experience, and this is what I will pick up on here. Can a Mādhyamika use the sentence form I developed to make true claims about the structure of reality (*tattva*) insofar as it underlies experiences?

The difficulty inherent in the idea is that my proposal relies on quantifying over elements in some domain of discourse (using a Q-Ramsey sentence), but it seems obvious that it can't be true, from a Madhyamaka perspective, to say that any such element exists when the domain is *tattva* as such. The impossibility of such sentences having truth values is what

Nāgārjuna was trying to show. This does not look good for the prospect of using the kind of sentence I've developed for this purpose. But it would be too hasty to call the case closed here. I aim to show, first, that the no miracles argument serves as a good reason to try very hard to make this work. That will motivate a discussion in philosophical logic, seeking a way to use the form of sentence I've developed to make truth-valued claims about reality which do not commit one to the truth of any existential sentence.

### 7.2 The case for (a little bit of) realism

Nāgārjuna writes (MMK 24.8-10),

> The Dharma teaching of the Buddha rests on two truths:
> the worldly, obscuring truth and the truth concerning the ultimate.
>
> Those who do not understand the distinction between these two truths,
> they do not understand reality in accordance with the profound teachings of the Buddha.
>
> The ultimate is not taught without having relied on customary ways of talking and thinking.
> Not having arrived at the ultimate, nirvāṇa is not attained.

In this well-known section, Nāgārjuna is replying to an opponent who has accused him of negating the Buddha's teachings. Nāgārjuna claims that it is a feature of Buddha's teachings that one can only teach about what is ultimately the case using language which does not reflect that reality. In other words, the only way to lead someone to a final realization of the nature of reality is to instruct them using what are ultimately *fictions*. Nāgārjuna goes on to argue that the only way that the Buddha's teaching of dependent arising make sense is if things are empty (i.e., taken to exist or not exist only conventional stipulation). So, in order to accept dependent origination – which, in Nāgārjuna's intent here, represents all causal judgments – as a cogent teaching, we need to think that it applies only to a fictional ontology.

Here is the difficulty I want to raise. How could it be that some fictions are more efficacious than others in teaching the ultimate? Nāgārjuna says that "the ultimate is not taught without having relied on customary [fictional] ways of talking". If dependent arising can only be taught as fiction, but in teaching it the buddhas and their disciples are somehow still teaching about the ultimate, why could they not do the same thing with the fictional teachings of Jesus, or Barney the Dinosaur, or Ayn Rand? How is it that some fictions can have a privileged relationship with ultimate reality which others do not? This is a serious problem lest we conclude that Nāgārjuna is an irrational dogmatist: he is saying that the teachings of his religion should be followed in pursuit of the ultimate, even though those teachings are not ultimately true. As a philosopher, he owes at least some indication of what properties these teachings have, if not truth, which affords them these privileges over others. But, as far as I can tell, he doesn't seem to have provided one.

I argued in Chapter One that, for Nāgārjuna, sentences which merely report experiences are truth-valued. Following Poincaré, we can say that the relations embedded in those sentences, at least, are objective in the sense of intersubjectively stable. Buddha's teachings on dependent arising could be interpreted as teaching about such phenomenological relations. The most important example is the twelvefold chain of the *nidāna*s, which Nāgārjuna recites in perfectly orthodox form in MMK26. This is Nāgārjuna's example of how the ultimate is taught despite the fact that the ultimate truth of the teaching has been refuted in the rest of the text. Let's look at how Nāgārjuna represents this teaching in the first verse of MMK26:

> One who is enveloped in ignorance forms three kinds of volitions that lead to rebirth;
>
> and by means of these actions one goes to one's next mode of existence.

As Nāgārjuna explains in the first twenty-five chapters of the text, there ultimately isn't anyone in ignorance, nor are there volitions, nor actions, nor modes of existence. But what he is doing here is accepting these as conventions corresponding to certain experiences. Those experiences are really had, but it is impossible to give them a true metaphysical interpretation – conventions are all we have to describe them with. The empirical relations between those conventionally designated experiences are not themselves conventions. Once we have established which experiences are associated with which conventions, the relations which hold between them is a simple matter of fact. This is how conventional truth can connect with ultimate reality.

If we consider the twelvefold chain of dependent arising as a whole, the relations described in this teaching form a certain structure. It may be said, then, that what makes this fictional teaching an effective path to the ultimate where others are not is that the *structure* of the teaching correctly reflects the structure of experience even though the *metaphysics* of the teaching does not correctly reflect the *nature* of experience or reality more broadly. Nāgārjuna's point in MMK is that no metaphysics, or physics, or psychology, or anything can ever reflect the true nature of reality, so these things are essentially matters of convention. Our choice between them is guided by the convenience they provide in communicating the structure of the experiences of sentient beings.

If we want to take the Poincaréan antirealist path, this is where the conversation ends: efficacious fictions like Buddha's teachings are efficacious because they satisfy the structure satisfied by intersubjectively stable phenomenal relations. This is also why successful scientific theories and many other fictions are efficacious. The natural question will be, is there more to reality than experiences? Neither Poincaré nor Nāgārjuna would be reckless enough to answer 'no' to this question, though it is plausible that they would hesitate to answer 'yes'. For them, the nature of reality is transcendent (MMK 18.9; Poincaré 1905/2001, 347-8). Given that it would be incoherent and unmotivated of them to claim that reality and experience are utterly distinct domains, this means that the ultimate nature of experience is transcendent. Given this, it would make no sense for them to declare that reality is only experience, or whatever. If reality transcends our understanding, we can't impose those sorts of limitations on it. The Poincaréan antirealist concludes that nothing at all should be said about objective reality beyond reports of experiences, not even suppositions about it 'underlying' experiences.

The Poincaréan has legs to stand on here. I made the case in Chapter Two for reading Nāgārjuna as relying on the idea that causation is a fundamental category of our understanding. Reality's being transcendent implies that it is outside of the legitimate domain of causal (and other) judgments. So, judgments about real systems underlying experiences would seem to be illegitimate.

To begin my counterargument, I will note that the kind of judgment we are considering is not, in the first instance, about transcendent reality but about experience. We start with an experience (appearance, sensation) and are making a proposal about its conditions. It is true, based on the reasoning above, that this proposal might be empty. According to the Poincaréan, we need to take experiences as brute facts – not because this is what they are intrinsically, but because our powers of judgment cannot probe any 'deeper' into what they truly are and why they come about (those notions themselves might be without meaning). Let's grant that we cannot *expect* to be able to make meaningful judgments about the objective conditions of experiences. Do we have good reasons to think that it is *impossible* for such judgments to be made? I think not. Our faculty of understanding is undoubtedly strictly limited – where those limits consist in the fact that the understanding is particular in what it is and how it functions, and those particularities cannot be thought to correspond with the extent of reality. But correspondence is a more-or-less affair, and the understanding is limited in its ability to make true or meaningful judgments even in its ideal application to experience. That is what Nāgārjuna shows in his departure from Kant: he problematizes the possibility of our understanding making existential judgments (i.e., judgments using existential sentences) even about phenomena (see Chapter Two). No one knows the extent to which the understanding can or can't achieve correspondence with reality and there is no straightforward way to find out. So – no, it's not justifiable to say that it's impossible for judgments about what underlies experience to have some sense and verisimilitude; the problem is figuring out how one could ever know that they do.

By my estimation, there is at least one thing that should convince a reasonable person that some judgment has achieved some verisimilitude even in a problematic application like the one under discussion. That thing is a case where the theory on which the judgment is based has achieved a truly impressive degree of novel empirical predictive success based on positing the existence of certain unobservable systems. In such cases where people have achieved profound objective success making judgments about experiences based on what they take to underlie those experiences, it would seem indicative of a lack of humility to insist that this very form of judgment fails to connect with objective reality in the way it seems to simply because we cannot verify that it does. We may reasonably question the status of the theory's ontology and other aspects of its supposed connection to reality, but to pre-empt the issue by deciding based on philosophical argument that judgments about the objective conditions of experience are necessarily illegitimate flies in the face of the fact that some such judgments represent the greatest intellectual achievements of human civilization. We have established that it is not impossible that such judgments connect with reality in some way. What better evidence could we hope for to show that some, in fact, do, than this?

Some people, of course, reject the argument above, which can be identified as an instance of what is called the no miracles argument (NMA) following Putnam's statement, "The positive argument for realism is that it is the only philosophy that doesn't make the success of science a miracle" (Putnam 1979, 73). The NMA is the heart of contemporary realism in the philosophy of science. Antirealists may reject it by refusing the notion that regularities in experience predicted by theories stand in need of explanation (e.g., van Fraassen 1980), or by arguing that the inference from the success of theories to their truth (or the truth of other, structure-based claims like mine) is fallacious (e.g., Magnus and Callender 2004), and I am not seeking an end to the debate here. I am saying that, for the many people, including me, who see the NMA as a powerful reason to believe that there must be some connection between at least some theories and reality, the NMA does not lose its force if we think, with Poincaré and Nāgārjuna, that the nature of reality transcends our faculties of understanding. In fact, it gains the additional power of assuring us that our ability to make causal judgments is effective when used to make at least some judgments about the objective transcendent causes of experiences. How far we can generalise that assurance beyond the exceptional cases is an open question, but this is now a straightforward epistemological question, not one demanding reflective metaphysics.

Not being able to provide a satisfying, at-least-a-little-realist account of the success of science would be enough to make many reasonable people cast Nāgārjuna's philosophy aside. Therein lies the urgency of figuring out how to make my epistemic structural realist proposal work with Madhyamaka. It is the only one existing which has any chance of working for a tradition so sceptical.

### 7.3 Transcendent satisfaction of structure

In Chapter Six, I rejected ontic structural realism on the basis that its positive metaphysical thesis is insufficiently motivated even if its negative one isn't. My argument relied on the intuition that it is possible for reality to instantiate relations in ways that we cannot understand. We need to take a closer look at this idea now, with Nāgārjuna's philosophy in mind. Considering my proposed sentence form, $\exists X_1 \ldots \exists X_n \mathbb{U}[\![Q(\theta)]\!][\![\mathbb{a}(P)]\!]$, it would be easy to say that the domain of $Q(\theta)$ is the set of transcendent objects and relations in objective reality. But this is problematic, for a transcendent object sounds like a kind of worldly thing – and Nāgārjuna says that neither existence nor nonexistence pertain to worldly things. What, then, are the relata being quantified over in $Q(\theta)$?

The first thing to note is that reality has some degree of complexity. We know this because our experiences are complex and experiences are not distinct from reality, so reality is at least as complex as the totality of phenomenal facts. One way to capture that complexity linguistically is to say that reality is a manifold of states of affairs. 'State of affairs' is meant to be a noncommittal designation. Such a state is not assumed to be an object possessed of intrinsic or extrinsic nature, or a relation, or anything else we could understand. It's merely some aspect of reality as such. Because this manifold is transcendent, there is no possible way to grasp how states of affairs can be differentiated from one another, but we know that there must be some sense of distinction because we know that there is complexity. This is what I propose we are quantifying over: transcendent states of affairs. The predicates $X_1 \ldots X_n$ represent relations instantiated by these states of affairs. We know that reality instantiates relations because experiences do, and we know that transcendent states of affairs instantiate certain relations (or do something somehow analogous to instantiating certain relations) because of the NMA.

Let's make sure that this isn't just smuggling problematic worldly things back in. The cleanest way to do that is to attend to Nāgārjuna's arguments and ask whether what I am proposing runs counter to what he was trying to establish. The explicit reason Nāgārjuna gives for rejecting the (non)existence of things is the absence of *svabhāva* (intrinsic nature) and *parabhāva* (extrinsic nature). That means there is no firm basis on which one may identify some worldly thing to say whether it exists or not. In other words, there are no precise truth conditions for any existential sentence pertaining to worldly things. This concern does not apply to the approximate[64] truth of sentences of our form $\exists \mathrm{X}_1 \ldots \exists \mathrm{X}_n \mathbb{U}[\![\mathrm{Q}(\theta)]\!][\![\mathbb{a}(\mathrm{P})]\!]$ for two main reasons:

1. Our reasons for thinking that worldly things need *svabhāva* or *parabhāva* as essential to the truth conditions of sentences asserting their existence does not apply to transcendent states of affairs (hereafter TSOAs).

Concerning the need for worldly things to have *svabhāva/parabhāva* to exist (MMK 15.4), the idea is that our understanding cannot make sense of how an existential sentence can be literally true unless its subject has an identity, and it cannot make sense of how a worldly thing could have an identity unless it has *svabhāva* or *parabhāva*. This latter problem does not apply to TSOAs because the T stands for transcendent – they, explicitly, are not expected to conform to our understanding's schema. Nor are they assumed to be 'things' by analogy to our ideas of worldly things like chairs, planets, or molecules, so I do not see why the inclination to think that they must have *svabhāva* or *parabhāva* would even arise.

2. The case for the existence of TSOAs satisfying the relevant structure does not rely on ascribing *svabhāva/parabhāva* to anything to be rigorously made.

[64] 'Approximate' truth is the standard goal for scientific realism. What precisely it means is controversial, but the guiding idea is that it would be uncharitable to the realist to say that their philosophical position is wrong unless their claims about the unobservable world are accurate to the last detail.

The case for the existence of these TSOAs is based on the incredible success of some theories. We take the models of these theories to be fictions. The objects they present are taken as conventions, not things thought to exist in objective reality. The nature of the relations presented are also taken as conventions, except perhaps for those which represent phenomenal relations. The argument for the existence of *some* TSOAs and some real relations which satisfy certain structures as explained in Chapter Five is the NMA. The argument is thus based entirely on fictions and the remarkable success of those fictions in navigating phenomena. At no point does anything need *svabhāva* or *parabhāva* for this to work.

A final concern I will discuss relates to Nāgārjuna's words. To establish that Nāgārjuna takes the nature of reality to be transcendent, I previously quoted MMK 18.9:

> Not ascertained from another, pacified, not fabricated by hypostatisation,
>
> Free from conceptualisation, not having different meanings – this is the nature of reality (*tattva*).

The potentially problematic part (for my argument) is his use of the word *anānārtham*, which I have translated as "not having different meanings". It's fine to say that TSOAs exist because to exist is just to be, and Nāgārjuna clearly thinks that reality *is*, given that he's talking about it. But saying that it is *anānārtham* suggests that it is illegitimate to analyse reality as having many different facets. I would seem to be doing this by representing reality as a manifold and quantifying over various TSOAs. I think that it is more or less clear why Nāgārjuna would have said something like this: his opponents spent a great deal of time and effort attempting to fully demarcate and categorise reality based on philosophy and scripture. As the rest of the verse makes clear, Nāgārjuna's point is that none of this conceptual activity can, even in principle, represent what reality ultimately is. I agree with him to that extent. But Nāgārjuna, in his time, would not have known any motivation, much less any detailed proposal, to speak about different aspects of reality at the level of abstraction that I have proposed. Nor could he have anticipated the incredible developments in our understanding of the world that are presently motivating me. So, I would argue that what we see here is not an incompatibility between Nāgārjuna's core insights and my proposal, but merely a mismatch between the application of those insights in his context compared to my own. This should be expected when a philosopher's work is being taken seriously roughly 1900 years after they wrote it. I don't expect to agree with the letter of everything Nāgārjuna has written.

## 7.4 Conclusion

I think that Nāgārjuna's ontological conventionalism is true, and I began researching this project because it was unclear to me how, given its truth, the incredible successes of science should be explained. I have come up with two solutions, one antirealist and one realist (the difference between these being whether they take theories to provide any knowledge of objective reality beyond the connections between phenomenal relations). The antirealist solution was discovered through a close reading of some relevant essays by Henri Poincaré, who came to ontological conventionalism through his own path of scepticism and transcendental idealism. I pursued the realist solution myself by first creating a way to make claims about the structures of unobservable systems which can work even in classic worst-case scenarios of radical scepticism. In this final chapter, I argued that this proposal can be used to make true claims about the structure of reality which are consistent with Nāgārjuna's philosophy.

I hope that this contribution can assist in better situating Nāgārjuna amidst contemporary philosophical concerns and the history of philosophy. He was a genius of the highest order. I pursued this topic in particular because I believe that humility is an intellectual virtue second to none, and that humility in the face of the accomplishments of natural science is of special importance for philosophers. If there were no way to square Nāgārjuna's philosophy with facts like Einstein's prediction that stars would appear in slightly different positions than usual when observed 'near' to the sun, then Nāgārjuna's philosophy should be rejected. I have shown that we are not in this situation and made some interesting findings along the way.

BIBLIOGRAPHY

Ainsworth, Peter M. 2009. "Newman's Objection." *The British Journal for the Philosophy of Science* 60 (1): 135–71.

Allison, Henry E. 2004. *Kant's Transcendental Idealism*. Revised and Enlarged edition. New Haven, Conn. ; London: New Haven, Conn; London : Yale University Press.

Ameriks, Karl. 2000. *Kant's Theory of Mind: An Analysis of the Paralogisms of Pure Reason*. Oxford University Press.

———. 2012. *Kant's Elliptical Path*. Oxford University Press.

Arnold, Dan. 2012. "The Deceptive Simplicity of Nāgārjuna's Arguments Against Motion: Another Look at Mūlamadhyamakakārikā Chapter 2." *Journal of Indian Philosophy* 40 (5): 553–91.

Ben-Menahem, Yemima. 2006. *Conventionalism: From Poincare to Quine*. Cambridge: Cambridge University Press.

Bird, Graham. 2006. *The Revolutionary Kant: A Commentary on the Critique of Pure Reason*. Chicago: Open Court.

Bitbol, Michel. 2003. "A Cure for Metaphysical Illusions: Kant, Quantum Mechanics, and the Madhyamaka." In *Buddhism and Science: Breaking New Ground*, edited by B. Alan Wallace, 325–64. New York: Columbia University Press.

———. 2010. "Reflective Metaphysics: Understanding Quantum Mechanics From a Kantian Standpoint." *Philosophica* 83: 53–83.

Brading, Katherine. 2011. "Structuralist Approaches to Physics: Objects, Models and Modality." In *Scientific Structuralism*, edited by Alisa Bokulich and Peter Bokulich, 43–65. Dordrecht: Springer Netherlands.

Brading, Katherine, and Elise Crull. 2017. “Epistemic Structural Realism and Poincaré’s Philosophy of Science.” *HOPOS: The Journal of the International Society for the History of Philosophy of Science* 7 (1): 108–29.

Brading, Katherine, and Elaine Landry. 2006. “Scientific Structuralism: Presentation and Representation.” *Philosophy of Science* 73 (5): 571–81.

Bueno, Otávio. 1999. “What Is Structural Empiricism? Scientific Change in an Empiricist Setting.” *Erkenntnis* 50 (1): 55–81.

Carnap, Rudolf. 1950. “Empiricism, Semantics and Ontology.” *Revue Internationale de Philosophie* 4 (11): 20–40.

———. 1928/2003. *The Logical Structure of the World and Pseudoproblems in Philosophy*. Open Court edition. Chicago: Open Court.

Daston, Lorraine, and Peter Galison. 2007. *Objectivity*. New York: Zone Books.

Deguchi, Yasuo, Jay L. Garfield, and Graham Priest. 2008. “The Way of the Dialetheist: Contradictions in Buddhism.” *Philosophy East and West* 58 (3): 395–402.

Demopoulos, William, and Michael Friedman. 1985. “Bertrand Russell’s The Analysis of Matter: Its Historical Context and Contemporary Interest.” *Philosophy of Science* 52 (4): 621–39.

Disalle, Robert. 1994. “On Dynamics, Indiscernibility, and Spacetime Ontology.” *The British Journal for the Philosophy of Science* 45 (1): 265–87.

Esfeld, Michael, and Vincent Lam. 2011. “Ontic Structural Realism as a Metaphysics of Objects.” In *Scientific Structuralism,* edited by Peter Bokulich and Alisa Bokulich, 143–59. Boston Studies in the Philosophy of Science 281. Springer.

Fraassen, Bas. C. van. 1980. *The Scientific Image*. Clarendon Library of Logic and Philosophy. Oxford: Oxford University Press.

———. 1996. “Structure and Perspective: Philosophical Perplexity and Paradox.” In *Logic and Scientific Methods*, edited by Maria Luisa Dalla Chiara, Kees Doets, Daniele Mundici, and Johan van Benthem. Springer Netherlands.

French, Steven. 2014. *The Structure of the World: Metaphysics and Representation*. Oxford University Press.

French, Steven, and James Ladyman. 2003. “Remodelling Structural Realism: Quantum Physics and the Metaphysics of Structure.” *Synthese* 136 (1): 31–56.

Garfield, Jay L. 1995. *The Fundamental Wisdom of the Middle Way: Nāgārjuna’s Mūlamadhyamakakārikā*. New York ; Oxford: Oxford University Press.

Geach, P. T. 1972. *Logic Matters*. Oxford: Basil Blackwell.

Giedymin, Jerzy. 1982. *Science and Convention: Essays on Henri Poincaré’s Philosophy of Science and the Conventionalist Tradition*. Foundations & Philosophy of Science & Technology. Oxford: Pergamon.

———. 1991. “Geometrical and Physical Conventionalism of Henri Poincaré in Epistemological Formulation.” *Studies in History and Philosophy of Science Part A* 22 (1): 1–22.

Guyer, Paul. 1987. *Kant and the Claims of Knowledge*. Cambridge: Cambridge University Press.

Holder, Justin P. 2023. “Poincaré’s Radical Ontology.” *HOPOS: The Journal of the International Society for the History of Philosophy of Science* 13 (1): 151–79.

Inwagen, Peter van, Meghan Sullivan, and Sara Bernstein. 2023. “Metaphysics.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta and Uri Nodelman, Summer 2023. Metaphysics Research Lab, Stanford University.

Ivanova, Milena. 2015. “Conventionalism, Structuralism and Neo-Kantianism in Poincaré′s Philosophy of Science.” *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* 52 (November): 114–22.

Kant, Immanuel. 2003. *Critique of Pure Reason*. Translated by Paul Guyer and Allen W. Wood. Repr. w. corr. Kant, Immanuel, 1724-1804. Works. English. 1992. Cambridge: Cambridge University Press.

———. 2004. *Prolegomena to Any Future Metaphysics That Will Be Able to Present Itself as a Science: With Two Early Reviews of the Critique of Pure Reason*. Translated by Günter Zöller and Peter G. Lucas. Oxford Philosophical Texts. Oxford: Oxford University Press.

———. 2005. *Notes and Fragments*. Edited by Paul Guyer. Translated by Paul Guyer, Curtis Bowman, and Frederick Rauscher. The Cambridge Edition of the Works of Immanuel Kant. Cambridge: Cambridge University Press.

Kellner, Birgit, ed. 2010. "Special Issue on Buddhist Theories of Self-Awareness (Svasaṃvedana) - Reception and Critique." *The Journal of Indian Philosophy* 38 (3).

Ketland, Jeffrey. 2004. "Empirical Adequacy and Ramsification." *The British Journal for the Philosophy of Science* 55 (2): 287–300.

Ladyman, James. 1998. "What Is Structural Realism?" *Studies in History and Philosophy of Science Part A* 29 (3): 409–24.

Ladyman, James, and Don Ross. 2007. *Every Thing Must Go: Metaphysics Naturalized*. Oxford University Press.

Langton, Rae. 1998. *Kantian Humility: Our Ignorance of Things in Themselves*. Oxford : New York: Clarendon Press ; Oxford University Press.

Laudan, Larry. 1981. "A Confutation of Convergent Realism." *Philosophy of Science* 48 (1): 19–49.

Leplin, Jarrett. 1997. *A Novel Defense of Scientific Realism*. Oxford, New York: Oxford University Press.

Longuenesse, Béatrice. 1998. *Kant and the Capacity to Judge: Sensibility and Discursivity in the Transcendental Analytic of the Critique of Pure Reason*. Princeton, N.J.: Princeton University Press.

———. 2005. *Kant on the Human Standpoint*. Modern European Philosophy. Cambridge: Cambridge University Press.

———. 2006. “Kant on a Priori Concepts.” In *The Cambridge Companion to Kant and Modern Philosophy*, edited by Paul Guyer, 129–68. Cambridge University Press.

———. 2011. “Kant’s Standpoint on the Whole: Disjunctive Judgment, Community, and the Third Analogy of Experience.” In *Kant and the Concept of Community*, edited by Charlton Payne and Lucas Thorpe, 17–40. NAKS Studies in Philosophy 9. New York: University of Rochester Press.

Macdonald, Anne. 2015. *In Clear Words: The Prasannapadā, Chapter One*. Vol. 2. Beiträge Zur Kultur- Und Geistesgeschichte Asiens ; Nr. 86. Wien: Verlag der Österreichischen Akademie der Wissenschaften.

Magnus, P. D., and Craig Callender. 2004. “Realist Ennui and the Base Rate Fallacy*.” *Philosophy of Science* 71 (3): 320–38.

May, Jacques. 1959. “Kant et Le Mādhyamika a Propos D’un Livre Récent.” *Indo-Iranian Journal* 3 (June): 102–11.

McKenzie, Kerry. 2017. “Ontic Structural Realism.” *Philosophy Compass* 12 (4): e12399.

McMullin, Ernan. 1984. “A Case for Scientific Realism.” In *Scientific Realism*, edited by Jarrett Leplin, 8–40. Berkeley: University of California Press.

Melia, Joseph, and Juha Saatsi. 2006. “Ramseyfication and Theoretical Content.” *The British Journal for the Philosophy of Science* 57 (3): 561–85.

Morrison, Margaret, and Mary S. Morgan. 1999. “Models as Mediating Instruments.” In *Models as Mediators: Perspectives on Natural and Social Science*, edited by Margaret Morrison and Mary S. Morgan, 10–37. Ideas in Context. Cambridge: Cambridge University Press.

Murti, T. R. V. 1980. *The Central Philosophy of Buddhism: A Study of the Mādhyamika System*. 2nd ed. London: Unwin Paperbacks.

Myrvold, Wayne C. 2020. “‘— It Would Be Possible to Do a Lengthy Dialectical Number on This;’” *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* 71 (August): 209–19.

Newman, M. H. A. 1928. “Mr. Russell’s ‘Causal Theory of Perception.’” *Mind* 37 (146): 137–48.

Oetke, Claus. 2011. “Two Investigations on the Madhyamakakārikās and the Vigrahavyāvartanī.” *Journal of Indian Philosophy* 39 (3): 245–325.

Oizerman, T. I. 1981. “Kant’s Doctrine of the ‘Things in Themselves’ and Noumena.” *Philosophy and Phenomenological Research* 41 (3): 333–50.

Paz, María de. 2021. “Poincaré, Le Roy, and the Nouveau Positivisme.” *HOPOS: The Journal of the International Society for the History of Philosophy of Science* 11 (2): 446–60.

Poincaré, Henri. 1902/1905/1908/2001. *The Value of Science: Essential Writings of Henri Poincaré.* Modern Library Science Series. New York: Modern Library.

———. 2013. *La Science Selon Henri Poincaré*. Paris: Dunod.

Psillos, Stathis. 2014. “Conventions and Relations in Poincaré’s Philosophy of Science.” *Methode - Analytic Perspectives*, no. 4: 98–140.

Putnam, Hilary, ed. 1975. “The Refutation of Conventionalism.” In *Philosophical Papers: Volume 2: Mind, Language and Reality*, 2:153–91. Cambridge: Cambridge University Press.

———. 1987. “Truth and Convention: On Davidson’s Refutation of Conceptual Relativism.” *Dialectica* 41 (1/2): 69–77.

Russell, Bertrand. 1927/992. *The Analysis of Matter*. London: Routledge.

———. 1919/1993. *Introduction to Mathematical Philosophy*. 2nd ed. New York: Dover Publications.

Schmaus, Warren. 2020. "From Positivism to Conventionalism: Comte, Renouvier, and Poincaré." *Studies in History and Philosophy of Science Part A* 80 (April): 102–9.

Siderits, Mark. 2003. "On the Soteriological Significance of Emptiness." *Contemporary Buddhism* 4 (1): 9–23.

———. 2013. "Does a Table Have Buddha-Nature?" *Philosophy East and West* 63 (3): 373–86.

———. 2014. "Causation, 'Humean' Causation and Emptiness." *Journal of Indian Philosophy* 42 (4): 433–49.

Siderits, Mark, and Shōryū Katsura. 2013. *Nāgārjuna's Middle Way: The Mūlamadhyamakākarikā*. Classics of Indian Buddhism. Boston: Wisdom Publications.

Smith, Joel. 2020. "Self-Consciousness." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Summer 2020. Metaphysics Research Lab, Stanford University.

Smithson, Robert. 2017. "Newman's Objection and the No Miracles Argument." *Erkenntnis* 82 (5): 993–1014.

Stcherbatsky, Fedor Ippolitovich. 1968. *The Conception of Buddhist Nirvāṇa: Along with Saṅskṛta Text of Madhyamaka-Kārikā*. [Rev. & enl. Ed.]. Varanasi: Bharatiya Vidya Prakashan.

Tegmark, Max. 2008. "The Mathematical Universe." *Foundations of Physics* 38 (2): 101–50.

Thébault, Karim P. Y. 2016. "Quantization as a Guide to Ontic Structure." *The British Journal for the Philosophy of Science* 67 (1): 89–114.

Tsongkhapa, Jay L. Garfield, Geshe Ngawang Samten, Tsongkhapa, Jay L. Garfield, and Geshe Ngawang Samten. 2006. *Ocean of Reasoning: A Great Commentary on Nagarjuna's Mulamadhyamakakarika*. Oxford, New York: Oxford University Press.

Tuck, Andrew P. 1990. *Comparative Philosophy and the Philosophy of Scholarship: On the Western Interpretation of Nāgārjuna*. Oxford: Oxford University Press.

Van Cleve, James. 2002. *Problems from Kant*. New York: Oxford University Press.

Votsis, Ioannis. 2003. “Is Structure Not Enough?” *Philosophy of Science* 70 (5): 879–90.

Westerhoff, Jan. 2009. *Nāgārjuna’s Madhyamaka: A Philosophical Introduction*. Oxford: Oxford University Press.

———. 2010. *The Dispeller of Disputes: Nāgārjuna’s Vigrahavyāvartanī*. New York ; Oxford: Oxford University Press.

Worrall, John. 1989. “Structural Realism: The Best of Both Worlds?” *Dialectica* 43 (1/2): 99–124.

———. 2007. “Miracles and Models: Why Reports of the Death of Structural Realism May Be Exaggerated.” *Royal Institute of Philosophy Supplements* 61: 125–54.

Zahar, E. G. 1997. “Poincarés Philosophy of Geometry, or Does Geometric Conventionalism Deserve Its Name?” *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* 28 (2): 183–218.

Zahar, Elie. 2001. *Poincaré’s Philosophy: From Conventionalism to Phenomenology*. Chicago: Open Court.

———. 2007. *Why Science Needs Metaphysics: A Plea for Structural Realism*. Chicago: Open Court.

Zahar, Elie, and John Worrall. 2001. “Appendix 4.” In *Poincaré’s Philosophy: From Conventionalism to Phenomenology*. Chicago: Open Court.