

Aporia

Undergraduate Journal of the St Andrews Philosophy Society

VOLUME XVIII — ISSUE No. 1

Aporia is funded by the University of St Andrews Philosophy Society, which receives funds from the University of St Andrews Department of Philosophy, the University of St Andrews Students' Association, and independent benefactors.

Contents

Harmful Dysfunction and Mental Illness	1
<i>Leilan Nishi</i>	
De Jure Rigidity	9
<i>Nicolien Janssens</i>	
Layers of Logical Consequence	19
<i>Grace Field</i>	

Acknowledgements

We would like to thank the individuals and institutions that continue to support the journal. Special thanks go to Kirsty Hardwick, Oscar Westerblad, Matthew Green, Adrian Kreutz, Paul Conlan, Makan Nojournian, and Dr Kevin Scharp for their helpful contribution to the review process, and Jackie Proven and Kyle Brady from the Online Journal Hosting Service at the University of St Andrews for their technical support.

Kimon Surlas-Kotzamanis
Editor-in-chief

Max Waller
Deputy editor

Harmful Dysfunction and Mental Illness: Why the latter is not the former

Leilan Nishi*

lnishi@scu.edu

Abstract

In his essay ‘The Concept of Mental Disorder’ (1992), Jerome C. Wakefield puts forth a hybrid account of mental disorder that relies on the concept of ‘harmful dysfunction’, wherein ‘harmful’ is a subjective value term determined by social norms, and ‘dysfunction’ is the objective value-neutral counterpart that denotes the failure of a mechanism to perform as evolutionarily intended. In this paper, I begin by laying out the kind of commitments Wakefield is wedded to, which will demonstrate that Wakefield’s ‘harmful dysfunction’ account of illness critically fails to unify accounts of physical and mental illness. I claim this because the concept of mental dysfunction itself is not value-neutral like Wakefield assumes and needs it to be, which makes the view unworkable when applied to mental disorder. In its place I propose a model of human flourishing that will account for many different models of mental functioning.

1 Introduction

Jerome C. Wakefield (1992) undertakes a grand venture when he seeks to come up with a comprehensive account of disorder that can explain what physical and mental disorders have in common. He defines disorders as ‘harmful dysfunctions’: ‘harmful’ being a value term determined by social norms, and ‘dysfunction’ a scientific term referring to the failure of an evolutionarily designed mental mechanism to function as intended (1992, 373). In certain ways, he succeeds. By combining the apparent opposites of value-laden and scientific approaches, he avoids the issues that come with defining disorders as either entirely normative or entirely descriptive.

However, his concept of mental disorder as harmful dysfunction contains a fatal problem: it cannot be adequately applied to mental disorders. In this paper, I will argue that there is reason to doubt, and ultimately reject, aetiological dysfunction as a marker of disorder, because

*Leilan Nishi is currently finishing the last year of her undergraduate degree in Philosophy at Santa Clara University. While her main interests are in philosophy of mental illness and existentialism, she has yet to encounter a philosophical field she doesn’t like. She will be starting her PhD at CUNY Graduate Center in the Autumn.

it is in fact not value-neutral in the way that Wakefield assumes. I will argue that Wakefield's appeal to aetiological function and rational agency contains commitments to two implicit norms: biological determinism and a normatively loaded conception of rationality. I end by gesturing at an alternative model of function, which accounts for the multitude of varying ways in which people function and promotes a life of flourishing.

2 The harmful dysfunction model

In this section, I will briefly lay out the commitments embedded in Wakefield's 'harmful dysfunction' account of disorder. The harmful dysfunction (HD) analysis is a hybrid account of disorder. The dysfunction component is the scientific half of the formula and is grounded on the concept of etiological (i.e., evolved) function: a mechanism is dysfunctional when its present state of operation has deviated from its intended function as determined by evolution (1992, 374). The harmful component is a value-laden concept that refers 'to the consequences that occur to the person because of the dysfunction and are deemed negative by sociocultural standards' (Wakefield 1992, 374). Only those mechanism malfunctions that are socially disvalued can be considered disorders. The HD analysis provides a concept of disorder that is meant to apply to all conditions we call disordered, thereby unifying them under a single category definition. The result is intended to be a cohesive account of disorder that explains why certain conditions are not merely environmental issues or socially disvalued conditions.

Wakefield defines a function more broadly as an effect that explains its cause: the heart's effect of pumping blood, for example, enters into the explanation of its cause for existing in our bodies, which makes pumping blood the function of the heart (he also uses the examples of seeing as the function of the eyes, and mobility as the function of legs, using this same reasoning). Aetiological explanations for natural functions provide a causal story for how the mechanisms possessed by an organism that contributed to its reproductive success become naturally selected and present in us today. He wants to extend such etiological explanations for physical mechanisms like pumping blood, seeing, and walking to *mental* mechanisms:

Considering that mental processes play important species-typical roles in human survival and reproduction, there is no reason to doubt that mental processes were naturally selected and have natural functions, as Darwin himself often emphasized [(Boorse 1976)]. Because of our evolutionary heritage, we possess physical mechanisms such as livers and hearts; that same heritage gave us mental mechanisms such as various cognitive, motivational, affective, personological, hedonic, linguistic, and behavioral dispositions and structures. Some mental conditions interfere with the ability of these mental mechanisms to perform the functions that they were designed to perform. In such cases, there is a part dysfunction of the particular mental mechanism. (Wakefield 1992, 375)

Wakefield assumes, without justification, that there must be an etiological story that explains

the existence of mental mechanisms like there are for physical ones; he seems to imply that because evolutionary heritage explains the physical, it is necessary that it explains the mental, because, he argues, it couldn't be merely a happy accident that we developed structures and dispositions that work so intricately and harmoniously together so well to provide the ultimate remarkable benefit of rationality. I argue that, even if we grant him this claim¹, his account of mental dysfunction is unworkable due to the normativity inherent in what is supposed to be a descriptive account of mental mechanism aetiology.

3 The problem of rational agency

Wakefield's account of the etiological function of mental mechanisms relies on his own unjustified conception of rational agency as value-free. When Wakefield discusses dysfunctional mental mechanisms, he almost universally appeals to his own conjecture that mental mechanisms were evolved to make us rational. Consider the following two examples of dyslexia and depression.

Wakefield (2000) uses dyslexia as a supposedly obvious example of a dysfunctional mental mechanism. He says, in regards to individuals who cannot learn how to read, that with those who 'seem incapable of learning to read even under optimal learning conditions, we infer that there is something wrong with some internal neurological mechanism that, when functioning as designed, supports the capacity to read (although it supports reading accidentally, not by design)² (2000, 256). Again, his statement is rife with assumptions about what this 'some neurological mechanism' is that is malfunctioning, which betrays his own lack of clarity on whether or not there even is a mechanism, and if there is, what it is. More notably, there is another assumption at work: that there exists an evolutionary dysfunction in this instance at all. He simply takes for granted that in the case of an individual who couldn't learn to read under conditions conducive to learning how to read, we would automatically infer that there was 'some internal neurological mechanism' that is failing to function as designed. The implication is that we evolved with the capacity to read, which is 'an invented way of exploiting our selected mechanisms for our own purposes' (2000, 255), the purpose being rationality. For Wakefield,

1. In addition, to move forward we must also ignore two extremely damaging epistemic concerns resulting from his etiological commitment that specific mental mechanism have specific evolved functions: (1) how do we identify the mechanism we should be looking at when a potential mental malfunction occurs for a specific condition? (2) even if we could solve that problem, how do we identify the evolutionary function of these mental mechanisms? Wakefield seems to realise that these are issues (2000, 263–64), but provides no way of solving them.

2. Wakefield here is talking about spandrels, or un-designed side effects of design, which he argues can malfunction in a harmful way when they are inevitable by-products of design, in which case the failure of the spandrel implies the failure of some intended function. He does not explain how this could apply to mental illnesses, but the following example would be in line with his theory: suppose there is an evolved mechanism for symbol recognition, and the inevitable spandrel of that design is the capacity to read written language. He would argue that a failure of the spandrel of reading is indicative of a failure of the evolved mechanism of symbol recognition. However, he would still be faced with the two epistemic issues mentioned in the first footnote, which remain just as practically insurmountable.

not being able to read is not merely a difference, but a dysfunction, because illiteracy marks an obstacle to exploiting our selected mechanisms towards this purpose of making us better able to reason and interact with our world to logically pursue our ends.

In Wakefield's particularly damning treatment of depression (2000, 266), we find another example of his belief that rationality is an evolutionarily selected function. He calls depression disordered when the 'loss [response is] extremely disproportionate to experienced loss,' which, as he acknowledges, assumes that 'sadness as a designed response to loss [could] turn out to be incorrect.' He explains away cultural values and culturally defined expectations for expressing sadness by arguing that depressive behaviour disvalued in one place and valued in another could have normal and abnormal sources; in other words, while 'there may be experiences when growing up (or, for all we know, genetic dispositions) that cause people in those cultures to express sadness more readily or intensely or more enduringly than in our culture, [someone] in our culture who has not had those experiences and yet reacts with a similarly high level of intensity may be doing so for entirely different reasons, including possibly a dysfunction' (2000, 266). Furthermore, he says that we should adjust our judgements and attributions of disorder according to other circumstances regarding the loss that could explain why the individual is reacting so strongly: 'if the individual's personality, the special meaning of the loss, or other circumstances suggest that a more intense or enduring response than the usual is due to non-dysfunction factors, we refrain from attributing disorder.'

However, Wakefield does not explain where those explanations end; he gives no guide for how we can differentiate between a personality that is inclined to dramatic expressions of emotion and a disorder, or how we know if we have correctly assessed the correct significance of the loss to the individual. He merely says that 'we try to judge when the reaction goes so far out of the usual bounds that it seems unrelated to any possible coping benefits', in which case '[w]e then become more persuaded that there is a possible dysfunction' (2000, 266). He assumes that if there is no identifiable culturally learned explanation for why someone is responding with extreme depression to a certain loss, we would not be unreasonable to infer a dysfunction. Similarly, he assumes that in lieu of personal circumstances related to the loss that could contribute to a 'disproportionate' loss response, we could potentially infer a dysfunction as well. He is assuming that, if there are no external or personalised circumstances that could explain the extreme sadness, we can infer a dysfunction. There is a notable and damaging implication from these assumptions: both of these limitations on attributing disorder are limits to determine when 'the reaction goes so far out of the usual bounds that it seems unrelated to any possible coping benefits', which would indicate that 'the disorder can be recognised by the fact that there is great sadness for no apparent reason'. He concludes by saying that 'none of this complex, contextually anchored reasoning, however speculative and fallible it may be, has anything inherently to do with local values' (2000, 266).

The problem is that it does in fact have to do with values, and that is the value of rational agency that underlies his assessment of how and when we recognise depressive disorder. When he refers to sadness that is 'disproportionate' and outside of 'usual bounds' as disordered, he cannot be saying they are statistically deviant modifiers, as those would not be aetiological cate-

gories. Wakefield is referring to rational agency. The sadness response is disproportionate in the sense that is unreasonable: the individual is failing to assess the magnitude of the loss correctly, and so reacts in a way that is unresponsive to the facts, much like the case of the chronically low and unresponsive self-esteem. The rational response would be to change one's sadness to match the magnitude of the loss, so when that does not happen, there is a failure to be rational—and for Wakefield, a disorder. In other words, feeling overwhelming sadness is a disorder when the magnitude of the sadness is beyond the rational range of response in regards to the loss, thereby constituting a reduction in the ability to be rational. We can make sense of Wakefield's view on dyslexia, depression, and self-esteem only through this notion of mental dysfunctions as undermining rational agency.

Rationality as the metric would not be an issue if it could reliably differentiate disorder from non-disorder while being value-neutral, as Wakefield believes. However, neither of these demands are met by the kind of rational agency that Wakefield assumes. The conditions that are disorders are those that are indicative of an impingement on rational agency—specifically, a condition is an impingement when it '[tends] to cause a person to act contrary to their interests without an adequate reason for doing so, and [impairs] a person's ability to decide competently and voluntarily, for example, by disrupting one's cognitive abilities' (Edwards 2009, 79). If we ask how we differentiate between a character predisposition towards 'intermittent but massive and harmful lapses in rationality' (2009, 80) and a mental illness, we find that the distinction is made on normative judgements, not descriptive ones. This distinction cannot be attributed to statistical distribution in a population, environmental circumstance, or severity (2009, 80).

In this same way, rationality is normative. There is no blueprint for what rationality should look like, both in mode of function and degree of function. To champion one model of rationality over another is to make a normative judgement; the parameters and applications of this rationality are arbitrary. Consider two of Edwards' examples: first, the description of 'a couple living in a war zone during a bombing raid, frozen in fear under a precarious cover just a few meters from a bomb shelter that would greatly increase their chances of survival' (2009, 80), who, by not running to the bomb shelter, are failing to think rationally; second, of a youth who engages in behaviour that could not be described as rational, such as committing robberies where the 'potential takings could not possibly justify the risk' or committing assaults where there is 'no chance of avoiding arrest' (2009, 80). Edwards calls the bomb raid couple irrational, but not disordered, and the youth irrational, but morally responsible for their actions. This is because there are many mental conditions that impair our rational agency and can have a negative impact to our wellbeing, but are not considered mental illnesses, and because the way we differentiate between a character predisposition towards 'intermittent but massive and harmful lapses in rationality' (2009, 80) and a mental illness is based on morality. Now contrast that with Wakefield's self-esteem example: he infers a dysfunction and attributes disorder when a person has chronically low self-esteem that is not aligned with the facts about themselves (facts that should make them feel otherwise) and is unresponsive to those facts—in other words, when it is irrational. Why is his version of rational agency any more objective than the kind being impaired in these two examples? There is no framework to which he is appealing that can explain

why certain lapses in rationality, even serious ones, are indicative of a dysfunction, while others are understandable, acceptable, or a matter of character. At best, he is making a normative value judgement about the ‘correct’ understanding of the concept of rational agency, and at worst, an appeal to his own personal intuition.

Physical illnesses have none of these problems, because physical illnesses are not identified by their effects on rational agency. As a result, the HD analysis applies without trouble to examples of physical illness, which Wakefield realises, as he uses many physical examples to better explain his position. However, what he seems not to fully realise is that the extension to the mental realm cannot be made, as the examples he gives to defend his model for mental illnesses are overwhelmingly examples of physical illness: human chin and jaw (2009, 255), appendicitis (2009, 256), fever and morning sickness (2009, 259, 262), and sickle cell anaemia (2009, 260), *et al.* This means that the extension of the harmful dysfunction concept from physical illnesses to mental illnesses is utterly broken, and therefore it cannot apply to mental illnesses.

4 Why aetiology for mental mechanisms?

Suppose now that somehow it were possible to solve all these issues and maintain the harmful dysfunction model for mental illnesses. While impressive, there still looms the question we should have asked first: why should we use aetiology to determine the function of mental mechanisms? As previously stated, aetiology does not pose practically intractable problems for physical illnesses. This is because we can differentiate between difference and dysfunction by referring to a basic account of biological determinism that is not value-laden. This is not functional determinism, which is the idea ‘that functions take place in a uniform mode at a relatively uniform performance level by a statistically distinctive portion of the members of a species’ (Amundson 2000, 36)—this is problematic because considering functional mode-how an organ functions-fails to account for the fact that a comparable performance level can be achieved through a different mode of function. I propose instead a model of biological determinism to understand physical illness: the idea that certain organs evolved with respect to other organs within a biological system in order to create an optimally functioning organism, and that when the organ does not develop as intended, there is a biological dysfunction, which does not necessarily have to result in reduced performance level or ability. By adhering to aetiology for physical conditions, we get the benefit of the unification of conditions under the harmful dysfunction model as a complete and comprehensive way of understanding and classifying conditions that stays faithful to the value-laden and value-neutral formula.

There is no such benefit conferred from applying an aetiological account of mental function. This is because we are persons, and ‘the interesting thing about *persons*, and possibly other things that have sophisticated mental lives, is that we value things other than survival and reproduction, and for the most part we evaluatively judge that other people *should* value things other than survival and reproduction (e.g., happiness or fulfillment)’ (Edwards 2009, 77). Aetiological

accounts of function not only fail to capture what it is we care about; they stifle such understandings. I have already shown this with rationality, but the same is true of a misapplication of biological determinism to the mental realm. This is because there are no blueprints for mental mechanisms like there are for physical mechanisms; there is no objective standard for what makes a self-esteem mechanism (assuming there is one) that is low more or less dysfunctional than one that is high. Even that mode of function that is non-typical ‘is not broken by its failure to comply with some imagined blueprint [...] It will function anyhow, in spite of its atypicality’, because there is no blueprint for mental function. Even if there was a blueprint, that blueprint would only be one that brings about functional integration, so that various mechanisms develop together and adjust to one another in order to function, as evidenced by the incredible multitude of ways in which people function (Amundson 2000, 39). Functional and biological determinist accounts have the same problem as models of rationality: though touted as descriptive, what is really happening is a normative judgement about the desirability of certain modes of mental mechanism function over others.

The only type of function we can employ to adequately encompass all these different kinds of mental functioning without sacrificing variation is a Cummins function (Woolfolk 1999 665–67). A Cummins or ahistorical function is one that focuses on ‘the causal relations among systems and their component parts, such that “the function of a part of a system is its causal contribution to some specified activity of the system” [(Walsh and Ariew 1996, 493)]’. This means that the Cummins function of a certain mechanism is defined in relation to a particular designated purpose of the system as a whole, which could be entirely arbitrary; it is ‘interest relative,’ and so ‘many different systems can be posited that concurrently contain the component’ (1999, 665). Because the designation of the system is arbitrary, ‘no background context of inquiry is privileged over any other, as is the case with the privileging of an evolutionary account by historical functional analysis’ (1999, 666). A mechanism is functioning if it is concurrent with whatever framework of interest is being applied, and failing to function if it is not; this means that the same mechanism can also be dysfunctional according to one account while being functional in another. For example, to take Woolfolk’s example of the heart (1999, 665–66), in the context of explaining human physiology, the Cummins function of the heart is to pump blood; in the context of an electrocardiogram, to produce electrical signals that result in EKG tracings; in the context of assassination, to bleed and lead to death.

For mental mechanisms, we can privilege certain background contexts, because the one we privilege is going to be that which best explains the interest we have. This means understanding their functionality in different contexts according to certain frameworks of interest—and these frameworks are not going to be aetiological ones. Aetiological frameworks could give us explanations only if our interest is in the biological background and history of mental mechanisms (granting for the moment that it is even possible to determine); such concerns are in the realm of theory, and so remain practically removed from our current environment, giving us no understanding of how to value these mechanisms in a practical manner. In Wakefield’s own words, ‘[t]he mental health theoretician is interested in the functions that people care about and need within the current social environment, not those that are interesting merely on evolutionary

theoretical grounds' (1992, 384). Human flourishing is the interest.

5 Conclusion

Wakefield's harmful dysfunction analysis, while an excellent model for classifying physical disorders, cannot be applied to mental disorders as he believes. The HD analysis of mental disorder cannot escape from the fact that the idea of a mental mechanism dysfunction is grounded in the normative concept of rationality, as are the concepts of aetiology and biological determinism. We therefore must abandon such historical concepts of mental function and adopt a model that can account for mental variability. This model will be an ahistorical account of interest-relative function that accurately captures the interest that really matters to us: how to live our best and most meaningful life. Because this will look different for each person depending on who they are, this new framework will allow for any number of different lifestyles, none of which are dysfunctional merely because they are different. It will make room for these alternative modes of being while understanding that mental conditions, including serious ones, are not pathological but struggles to be overcome on the path to leading the most fulfilling life.

References

- Amundson, Ron. 2000. "Against Normal Function." *Studies in History and Philosophy of Science Part C* 33 (1): 33–53. doi:10.1016/s1369-8486(99)00033-3.
- Boorse, Christopher. 1976. "What a Theory of Mental Health Should Be." *Journal for The Theory of Social Behaviour* 6 (1): 61–84. doi:10.1111/j.1468-5914.1976.tb00359.x.
- Edwards, Craig. 2009. "Ethical Decisions in the Classification of Mental Conditions as Mental Illness." *Philosophy, Psychiatry, & Psychology* 16 (1): 73–90. doi:10.1353/ppp.0.0219.
- Wakefield, Jerome C. 1992. "The Concept of Mental Disorder: On the Boundary Between Biological Facts and Social Values." *American Psychologist* 47 (3): 337–88. doi:10.1037/0003-066x.47.3.373.
- . 2000. "Spandrels, Vestigial Organs, and Such: Reply to Murphy and Woolfolk's "The Harmful Dysfunction Analysis of Mental Disorder"." *Philosophy, Psychiatry, & Psychology* 7 (4): 253–69.
- Walsh, Denis M., and André Ariew. 1996. "A Taxonomy of Functions." *Canadian Journal of Philosophy* 26 (4): 493–514. doi:10.1080/00455091.1996.10717464.
- Woolfolk, Robert L. 1999. "Malfunction and Mental Illness." *Monist* 82 (4): 658–70. doi:10.5840/monist199982429.

De Jure Rigidity

Nicolien Janssens*

nicolien.j@gmail.com

Abstract

The rigid designation of proper names and natural kind terms is the most well-known doctrine of Kripke's *Naming and Necessity* (1981). On the basis of rigidity, Kripke has shown that proper names and natural kind terms do not refer via a description as argued by descriptivists. In response to Kripke several people have argued that all general terms could be interpreted rigidly, which would make the notion of rigidity trivial. This leads to the 'rigidity problem': the notion of rigidity cannot be used to argue against descriptivism anymore. I will show that the rigidity problem appears on a larger scale: firstly, because it appears independently of the trivialisation problem, secondly, because it appears for descriptions acting like singular terms as well. I will argue, however, that proper names and natural kind terms differ in an important manner from rigid descriptions. While the first are *de jure* rigid, the latter are *de facto* rigid. I will show that the rigidity problem indeed appears for *de facto* rigidity, but not for *de jure* rigidity, with the result that Kripke's argument against descriptivism can withstand.

1 Introduction

In *Naming and Necessity* (1981), Kripke introduces his well-known notion of rigid designation. Kripke defines a rigid designator as a designator that designates the same object in every possible world in which the object exists (1981, 48). In other words, a rigid designator is a referential expression that in every possible situation refers to the same object. According to Kripke, a specific kind of singular term, namely proper names (1981, 48), and a specific kind of general term, namely natural kind terms (1981, 134), are rigid designators. Kripke uses the notion of rigidity mainly to argue against descriptivism, which holds that the meaning of a term is determined by a description (1981, 6–15).

The notion of rigidity, however, has only been defined by Kripke for singular terms, which designate a single individual. General terms, on the other hand, designate more individuals

*Nicolien Janssens is a second year undergraduate philosophy student at the University of Amsterdam. Currently, she is particularly interested in philosophy of language, metaphysics and logic. However, her interests are broad and therefore she hopes in the future to be able to combine philosophy with other disciplines.

at the same time. While Kripke says that natural kind terms are rigid designators, he does not define *what* they rigidly designate. Several philosophers have suggested that a rigid general term designates the same property or kind in every possible world (LaPorte 2000; Salmon 2005; Orlando 2014). I will call this theory henceforth the ‘kind-theory’. Some pointed out that the kind-theory would make *all* general terms rigid and thus would make the notion of rigidity trivial (Macbeth 1995; Schwartz 2002). For every description that acts like a general term, it would be possible to argue that it designates a certain abstract property or kind in every possible world¹. Hence, they argue, the concept of rigidity could be extended to all general terms, and thus would cease to be a useful concept to describe the different semantic behaviour of different kinds of terms. As a result, we encounter what I will call the ‘rigidity problem’: since it turned out that, apart from natural kind terms, every general term could be interpreted rigidly, rigidity cannot be used to argue against descriptivism anymore, as the other general terms do refer via a description.

Some hold that the trivialisation problem simply cannot be solved (Schwartz 2002; Soames 2002; Haraldsen 2017). Therefore, they argue that the notion of rigidity should not be applied to general terms at all. The result of this is that we cannot argue by means of the notion of rigidity that natural kind terms do not refer via a description. Others wanted to defend the claim that natural kind terms are rigid designators. They therefore tried to overcome the trivialisation problem either by showing that not all general terms are rigid (LaPorte 2000; Salmon 2005; Orlando 2014), or by trying to give another theory of what a general term designates (Devitt 2005).

Solving the trivialisation problem, however, will not suffice, since the rigidity problem appears on a larger scale. Firstly, because there are general terms that turn out to be rigid independently of the kind-theory. Secondly, because there are singular terms that turn out to be rigid as well (Kaplan 1978). The description ‘that woman—the one that has won the Nobel prize—is brilliant’, for example rigidly designates the winner of the Nobel prize. The rigidity of this singular term does not refute descriptivism either. As a result, the rigidity problem has been enlarged: the problem not only arises for general terms, but for the notion of rigidity in general.

To solve the rigidity problem, it is thus needed to show that the notion of rigidity can still be used to refute descriptivism for certain terms. It is my aim in this paper to show that this is possible and thus that the rigidity problem can be solved. In order to do this, I will show in §2 that Kripke’s reason to draw attention to the notion of rigidity is that it is required in his argument against descriptivism. In §3, I will show that the rigidity problem consists in the fact that the critique of descriptivism is undermined. In §4, I propose my own solution for the rigidity problem which is to make a distinction between *de jure* rigidity and *de facto* rigidity. I

1. Apart from descriptions acting like general terms, it is sometimes argued that all artificial kind terms can be interpreted rigidly according to the kind-theory. Meanwhile, there are several philosophers who hold that artificial kind terms cannot be interpreted rigidly due to the kind-theory (Orlando 2014). The rigidity of artificial kind terms is a matter of dispute that is too big to discuss here. However, it will not affect my main point (see footnote 5).

will show that *de jure* rigidity can be used to argue against descriptivism, with the result that Kripke's argument against descriptivism can withstand.

2 Rigidity to argue against descriptivism

In *Naming and Necessity* (1981) Kripke argues against descriptivism of names as advocated by Russell and Frege. Kripke offers different objections to descriptivism, which we can roughly divide into three categories: modal arguments, semantic arguments and epistemic arguments. In this section, I will present Kripke's modal argument, since this argument is interconnected with the notion of rigidity, as I will show.

Descriptivism holds that a name has an associated description or cluster of descriptions that gives the meaning of the name (Kripke 1981, 27)². Assume that we associate the description 'the last great philosopher of antiquity' with the name 'Aristotle'. We should then analyse sentence (1), according to descriptivism, as sentence (2) (Kripke 1981, 6–7):

- (1) Aristotle was fond of dogs.
- (2) The last great philosopher of antiquity was fond of dogs.

That descriptivism is erroneous becomes apparent when we compare the truth-values of sentences (1) and (2) in different possible worlds. We will see that sentence (1) is true in other worlds than sentence (2), from which follows that sentence (1) and (2) have different modal profiles. Sentence (2) is true in the actual world if Aristotle was actually fond of dogs, since in the actual world Aristotle satisfies the description 'the last great philosopher of antiquity'. That Aristotle was the last great philosopher of antiquity, is however not a necessary property of Aristotle. It might have been the case that someone else was the last great philosopher of antiquity. In such a situation, sentence (2) will be true *if that other person*, not Aristotle, was fond of dogs. In contrast to sentence (2), sentence (1) will always be true only if Aristotle, and *not some other person*, was fond of dogs. In this way, Kripke has shown that, contrary to what descriptivists argue, sentences (1) and (2) have different modal profiles and hence different meanings.

To clarify the different modal behaviour of the name 'Aristotle' and the description 'the last great philosopher of antiquity', Kripke introduces the notion of a rigid designator. A rigid designator designates the same object in every possible world in which the object exists (1981, 48). Kripke argues that all proper names are rigid designators (1981, 49). Although Aristotle might not have been the last great philosopher of antiquity, it is not the case that Aristotle might not have been Aristotle. Thus in every possible world the proper name 'Aristotle' designates the same person. The description 'the last great philosopher of antiquity' is on the other hand not a rigid designator. If we think of a situation in which the last great philosopher was a woman, we are not thereby thinking of a situation in which Aristotle was a woman. More generally, when we use a description to describe another possible world, we refer to whom—or whatever satisfies

2. There are also description theories that hold that a description only determines the reference of a name. I will not discuss this kind of descriptivism here, since Kripke's argument only refutes description theories of meaning (Kripke 1981, 5).

that description in that other world, independently of whom or what satisfies the description in the actual world. Thus a description does not designate the same object in every possible world, which makes it a non-rigid designator. Since proper names are rigid designators and definite descriptions are non-rigid designators, it follows that names cannot be equivalent to descriptions as descriptivists argue. In this way, Kripke has shown that we have a direct intuition of the rigidity of proper names, which is exhibited in our understanding of the truth conditions of particular sentences (1981, 14).

Kripke argues that we have the same intuition for natural kind terms, hereby showing equally that these terms are not synonymous with a description (1981: 134–35). We can demonstrate this by means of the following example:

- (3) Water is used to make tea.
- (4) A transparent liquid is used to make tea.

That descriptivism is erroneous for natural kind terms, becomes again clear when we compare the truth-values of sentences (3) and (4) across possible worlds. We will see, as in the case of sentence (1) and (2), that sentence (3) and (4) are true in different possible worlds. Sentence (3) is true if and only if water, *and not any other substance*, is used to make tea. Sentence (4), on the other hand, is true if a transparent liquid, *which might be another substance than water*, is used to make tea. Consider a situation in which vodka is used to make tea. In such a situation, sentence (3) is false, but sentence (4) is true. If ‘water’ means ‘a transparent liquid’, then sentence (3) and (4) could not differ in truth-value in any counterfactual situation. We see, however, that sentence (3) and (4) have different modal profiles and hence different meanings. Kripke has shown in this way that we do not refer to natural kinds by means of a description. In conclusion, the reason that Kripke draws attention to the notion of rigidity is that it is needed to argue against descriptivism.

3 The rigidity problem

In what follows, I will show how the trivialisation problem emerged from the kind-theory. Due to this theory, every description that acts like a general term can be interpreted rigidly, which makes the notion of rigidity trivial. The consequence of this is the rigidity problem: rigidity cannot be used to argue against descriptivism anymore. I will show, however, that not only the trivialisation problem undermines Kripke’s critique of descriptivism. Firstly, because there are descriptions acting like general terms that turn out to be rigid independently of the kind-theory. Secondly, because there are descriptions acting like singular terms that turn out to be rigid as well. The consequence of this is that Kripke’s argument against descriptivism is not only at stake for general terms, but for the notion of rigidity in general. The rigidity problem thus appears on a larger scale.

As pointed out earlier, the definition of a rigid designator only applies to singular terms, not to general terms. Therefore some have developed the kind-theory, according to which a

rigid general term designates the same property or kind in every possible world (LaPorte 2000; Salmon 2005; Orlando 2014). In response to this, others pointed out that as a result, every description acting like a general term could be interpreted rigidly (Martí and Martínez-Fernández 2011). For example, the description ‘the colour of the sky’ in a sentence like ‘My true love’s eyes are the colour of the sky’, can be interpreted in two ways. Firstly, it can be interpreted as designating the colour that is the colour of the sky, which in our world is blue, but which in other worlds might be another colour. In this case, ‘the colour of the sky’ is a non-rigid general term. It is non-rigid, because it designates different colours in different worlds. It is a general term, because a colour can apply to more individuals at the same time. Secondly, ‘the colour of the sky’ could also designate the ‘property of being the colour of the sky’. This is a description that acts like a general term, since a property can apply to more individuals at the same time. In this case, ‘the colour of the sky’ is a rigid designator, since it designates the same property in every possible world. In this way, it turned out that due to the kind-theory all general terms can be interpreted rigidly, which makes the notion of rigidity trivial.

The consequence of the trivialisation problem is that Kripke’s argument against descriptivism is undermined. Kripke’s argument showed that rigid terms do not refer via a description. Due to the trivialisation problem, however, it turned out that there are rigid terms that do refer via a description, since they simply *are* descriptions. ‘The property of being the colour of the sky’ refers to that property, via the description ‘the property of being the colour of the sky’. So it is not the case here that the notion of rigidity demonstrates that descriptivism is not adequate for these terms. For this reason, the notion of rigidity cannot be used to argue against descriptivism anymore. This is the rigidity problem. The rigidity problem can only be solved by showing that the notion of rigidity can still be used to argue against descriptivism. To solve the problem, it is therefore not right just to show that not all general terms are rigid, as some have tried (LaPorte 2000; Salmon 2005; Orlando 2014). We need to show that those and only those terms that do not refer via a description are rigid.

The other given solution to formulate another theory of what a rigid general term designates (Devitt 2005), will not solve the rigidity problem either. There are descriptions acting like general terms that are rigid, independently of the kind-theory. Firstly, this is the case with essentialist descriptions (Kaplan 1978). Essentialist descriptions are descriptions that pick out their object by properties that the object necessarily satisfies uniquely. An example of an essentialist description is ‘the substance with molecular formula H_2O ’. Secondly, there are rigidifying operators that make descriptions rigid. Kaplan has argued that this is the case with demonstratives (1978). For example, in the sentence ‘that liquid—the one with the molecular structure H_2O —is transparent’, the demonstrative ‘that’ rigidly designates water. Apart from demonstratives, the word ‘actual’ is a rigidifying operator. In every possible world the word ‘actual’ causes the description to refer to the actual world, so that in every possible world the description will refer to the same object. For example, ‘the actual liquid that is filling the seas’ will in every possible world refer to the liquid that is actually filling the seas. So it will refer in every possible world to water.

Essentialist descriptions not only make descriptions acting like general terms rigid, but also

descriptions acting like singular terms. This is for example the case with the description ‘the smallest prime number’. This description necessarily picks out the number two, since it is the only number that has the property of being the smallest prime number. Also rigidifying operators make descriptions that act like singular terms rigid. For example in the sentence ‘that woman—the one that has won the Nobel prize—is brilliant’, the demonstrative ‘that’ rigidly designates the winner of the Nobel prize. The same holds for the rigidifying operator ‘actual’: ‘the actual prime minister of the Netherlands’ will in every possible world refer to the person who is the prime minister of the Netherlands in the actual world. So apart from general terms, there are singular terms that turn out to be rigid. For these singular terms, rigidity cannot be used to argue against descriptivism either. The rigidity problem is thus a problem just as much for singular terms as for general terms.

In sum, it turned out that some descriptions that act like singular terms and some descriptions that act like general terms are rigid. The rigidity of these descriptions did not show us that these terms do not refer via a description, as it did for proper names and natural kind terms. In this way, the rigidity problem arose: rigidity cannot be used to argue against descriptivism anymore.

4 *De jure* and *de facto* rigidity

Kripke is aware that some philosophers argue that there is a rigid sense of definite descriptions (1981, 6n8). He says, however, that he is not convinced of this, ‘but if these philosophers are right, my principal thesis is not affected’ (1981, 6n8). I have just shown that these philosophers are right. In what follows, I will argue that Kripke is right to claim that his principal thesis, that proper names and natural kind terms do not refer via a description, is nonetheless not affected. This becomes clear if we acknowledge that there is an important distinction between proper names and natural kind terms on the one hand, and rigid descriptions on the other hand. While the first are *de jure* rigid, the latter are *de facto* rigid. I will show that *de jure* rigidity refutes descriptivism for certain terms, while *de facto* rigidity cannot be used to argue against descriptivism.

A rigid designator can designate the same object in every possible world in two different ways, which leads to a distinction between two different kinds of rigid designators. The first kind is a *de jure* rigid designator: a *de jure* rigid designator designates the same object in every possible world purely in virtue of its semantic kind. In this case, the semantical rules of the language directly link the term to the object. These semantical rules come about by means of stipulation. The second kind is a *de facto* rigid designator: a *de facto* rigid designator designates the same object in every possible world in virtue of expressing a description that happens to designate the same object in every possible world. Hence, *de facto* rigid designators are rigid not because of semantical rules, but because of non-linguistic facts³.

3. Kripke admits that he has ignored the distinction between *de jure* and *de facto* rigid designators, but he does give a rough sketch of what the distinction is (1981, 21n21). According to Kripke, in the case of a *de jure* rigid

The term ‘two’ for example is a *de jure* rigid designator of the number two. The term ‘two’ is a proper name with which we associate particular semantical rules. These semantical rules determine that the term ‘two’ will designate the number two in all possible worlds. So the term ‘two’ is rigidly linked to the number two because of semantical rules, which makes it a *de jure* rigid designator. ‘The smallest prime’, on the other hand, is a *de facto* rigid designator of the number two (Kripke 1981, 21n21). It is not possible that another number would be the smallest prime. This follows not from semantical rules, but from the metaphysical fact that mathematical facts are true in all possible worlds. ‘The smallest prime’ is thus via a non-linguistic fact rigidly linked to the number two, which makes it a *de facto* rigid designator.

Proper names and natural kind terms are *de jure* rigid. When we give someone a name *x*, we stipulate that from now on when using that name *x*, we refer to *that* person⁴. The same holds for natural kind terms. When introducing a natural kind term, we equally stipulate that it will refer to *that kind of thing*. We may use a description to determine the reference of a natural kind term, but this description will never be the meaning of the natural kind term. As an example, Kripke points out that we may determine the reference of the term ‘light’ by the fact that it affects our eyes in a certain way (1981, 130). However, it might have been the case that all people were blind and that light would not affect our eyes. In such a situation, we would not say that light did not exist. Rather, we would say that light existed in that case, although our associated description with light would not be adequate. In this way, Kripke shows that ‘light’ does not mean the same as ‘that which affects our eyes in a certain way’. From this it follows that natural kind terms cannot be *de facto* rigid designators. If they were *de facto* rigid designators, then light had to mean the same as its description, which on its turn would happen to be rigid because of non-linguistic facts. Since light is not equivalent to some description, it must be the case that the term ‘light’ is directly linked to the natural phenomenon of light, due to the semantical rules which arose from stipulation. So both proper names and natural kind terms refer rigidly because of semantical rules that arose from stipulation. This makes them *de jure* rigid designators⁵.

Definite descriptions are, if rigid, *de facto* rigid. Descriptions designate objects by means

designator ‘the reference of a designator is stipulated to be a single object’, while in the case of a *de facto* rigid designator ‘a description “the *x* such that *Fx*” happens to use a predicate “*F*” that in each possible world is true of one and the same unique object’ (1981, 21n21). I think my elaboration of the distinction coincides with this rough sketch.

4. Bostock (1988) and Evans (1979) have pointed out that some proper names refer via a description. Kripke admits that this might be true (1981, 79–80), but I cannot discuss the matter here. If true, however, my argument should be adjusted slightly: not *all* proper names are *de jure* rigid. My main point could withstand: *de jure* rigid designators could still be used to argue against descriptivism, although there might not be a specific group of terms that is *de jure* rigid.

5. As I pointed out in footnote 1, some hold that artificial kind terms are rigid designators in the sense that they refer to the same kind in every possible world due to semantical rules. This would make artificial kind terms belong to the *de jure* rigid designators. I also pointed out that others hold that artificial kind terms are not rigid. In either case, my main point, that *de jure* rigid designators do not refer via mediation of a description, will not be affected. The only doubt here is whether we should count artificial kind terms as *de jure* rigid designators or as non-rigid designators.

of a description as we saw earlier. If then descriptions are rigid, they have constant reference via mediation of some description that happens to designate the same object in every possible world. I illustrated this above by means of the description ‘the smallest prime’. Descriptions refer thus, if rigid, always via a description, which make them *de facto* rigid designators.

It is a more controversial issue whether a description that contains a demonstrative refers via a description. Earlier we saw the sentence: ‘that woman—the one that has won the Nobel prize—is brilliant’. Here it seems we are stipulating at this very moment that ‘that woman’ will always refer to the woman that has won the Nobel prize and thus that it is a *de jure* rigid designator. Kaplan (1978) for example holds this. However, what a demonstrative designates is always a matter of context. That ‘that woman’ refers rigidly to the winner of the Nobel prize is determined by the context of utterance, or in other words by the way that ‘that woman’ is described. Without a context of utterance ‘that woman’ refers to nothing. I cannot see therefore, how a demonstrative could designate an object purely in virtue of its semantic kind, since a context of utterance is always needed to determine the referent. Hence, I hold that definite descriptions containing demonstratives are *de facto* rigid as well⁶.

I have shown that the reason that Kripke draws attention to the notion of rigidity is that it is needed to argue against descriptivism. We can now say that *de jure* rigidity is the notion that is needed to argue against descriptivism. In the case of *de facto* rigidity, however, the reference of a term is determined by means of a description. From this it follows that *de facto* rigidity cannot be used to argue against descriptivism. Due to the *de jure–de facto* distinction, we have killed two birds with one stone. Firstly, we have justified Kripke’s claim that descriptivism is not adequate for proper names and natural kind terms, even though more terms turned out to be rigid. Secondly, we have solved the trivialisation problem with regard to *de jure* rigid terms. It might be the case that all terms can be interpreted as *de facto* rigid, but at least I have shown that not all terms can be interpreted as *de jure* rigid. Hence, *de jure* rigidity is an important, non-trivial notion which we can use to argue against descriptivism for certain terms, as Kripke justly has pointed out.

Devitt also holds that the notion of rigidity must refute descriptivism for proper names and natural kind terms (2005, 144)⁷. Devitt introduces the notion of ‘rigid application’ for general terms: ‘a general term “*F*” is a rigid applicer iff it is such that if it applies to an object in any possible world, then it applies to that object in every possible world in which the object exists’ (2005, 146). According to Devitt, this notion refutes *most, but not all* description theories for names and natural kind terms, since there are descriptions acting like singular terms (2005, 145) and descriptions acting like general terms (2005, 147) that are rigid. He gives the following

6. As in the cases pointed out in footnote 4 and 5, even if demonstratives would turn out to be *de jure* rigid, my main point would not be affected.

7. Schwartz (2002) also thinks that rigidity should have a theoretical work if it is not to be trivial. Schwartz, however, has another, and in my opinion wrong, conception of what the theoretical work is that the notion of rigidity should have. He namely thinks that rigidity is supposed to distinguish natural kind terms from artificial ones (2002, 273). According to Schwartz, general term rigidity fails to do this and therefore the notion of rigidity should only be applied to singular terms (2002, 275–76). As I showed in section 2, however, this is not the reason that Kripke draws attention to the notion of rigidity.

example of a singular term rigid applier that refers via a description: ‘the person who was *actually* the last great philosopher of antiquity’ (2005, 145). Equally, he gives the example ‘stuff with atomic number 79’ and points out that this is a general term rigid applier, although it does refer via a description (2005, 147).

Because on Devitt’s solution rigidity *most, but not all*, of the times refutes descriptivism, the solution does not seem convincing. In claiming that rigidity should be able to be used to argue against descriptivism, the notion should be able to refute descriptivism consistently. Moreover, Devitt only provides a solution for general terms, while my solution applies to singular terms as well as general terms. The solution I propose is therefore favourable over the one from Devitt, since in my solution *de jure* rigidity *always* can be used to refute descriptivism for certain terms.

5 Conclusion

In this paper, I have shown that a term can be rigid in two ways. A *de jure* rigid designator designates the same object in every possible world in virtue of its semantic kind. A *de facto* rigid designator designates the same object in every possible world in virtue of expressing a description which happens to designate the same object in every possible world. Due to the *de jure–de facto* distinction, rigid terms that do not refer via a description, namely proper names and natural kind terms, are clearly separated from rigid terms that do refer via a description, namely rigid descriptions. Consequently, the notion of rigidity, understood as *de jure* rigidity, can be used to refute descriptivism for certain terms. In this way, I have shown how the rigidity problem could be overcome in a way that is compatible with Kripke’s claim that rigid terms, understood as *de jure* rigid, do not refer via a description.

There is some secondary work that the notion of rigidity fulfils that I have not discussed in this paper. For instance, Kripke has argued that an identity statement in which both designators are rigid must be, if true, necessarily true, even if the statement is *a posteriori*. I think my distinction between *de jure* and *de facto* rigidity coincides with this secondary work in the following way: an identity statement containing two *de jure* rigid designators is necessarily true, while an identity statement containing two *de facto* rigid designators is not always necessarily true. I am afraid, however, that future research should decide whether I am right about this. Besides that, it is still not clear whether the kind-theory applies to rigid general terms. I have shown that the kind-theory could apply to general terms, because my proposal to make a distinction between *de jure* and *de facto* rigidity shows that the trivialisation problem does not apply to *de jure* rigid general terms. But there might be other reasons to adhere or reject the kind-theory, which I cannot discuss here. What I wanted to point out in this paper is that rigidity, understood as rigidity *de jure*, is not a trivial notion, but an important and indispensable one to argue against descriptivism.

References

- Bostock, David. 1988. "Necessary Truth and A Priori Truth." *Mind* 97 (387): 343–79.
- Devitt, Michael. 2005. "Rigid Application." *Philosophical Studies* 125 (2): 139–65. doi:10.1007/s11098-005-8221-y.
- Evans, Gareth. 1979. "Reference and Contingency." *The Monist* 62 (2): 161–89.
- Haraldsen, Fredrik. 2017. "Kind Term Rigidity and Property Identities." *Philosophia* 45 (3): 1179–93. doi:10.1007/s11406-017-9822-3.
- Kaplan, David. 1978. "Dthat." In *Syntax and Semantics*, edited by Peter Cole, 221–43. Academic Press.
- . 1989. "Demonstratives." In *Themes From Kaplan*, edited by Joseph Almog, John Perry, and Howard Wettstein, 481–563. Oxford University Press.
- Kripke, Saul. 1977. "Speaker's Reference and Semantic Reference." In *Studies in the Philosophy of Language*, edited by Peter A. French, Theodore E. Uehling Jr, and Howard K. Wettstein, 255–296. University of Minnesota Press.
- . 1981. *Naming and Necessity*. Oxford: Wiley-Blackwell.
- LaPorte, Joseph. 2000. "Rigidity and Kind." *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 97 (3): 293–316.
- Macbeth, Danielle. 1995. "Names, Natural Kind Terms, and Rigid Designation." *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 79 (3): 259–81.
- Martí, Genoveva, and José Martínez-Fernández. 2011. "General Terms, Rigidity and the Trivialization Problem." *Synthese* 181 (2): 277–93. doi:10.1007/s11229-010-9802-7.
- Orlando, Eleonora. 2014. "General Terms and Rigidity: Another Solution to the Trivialization Problem." *Manuscrito* 37 (June): 49–80. doi:10.1590/S0100-60452014000100002.
- Putnam, Hilary. 1975. *Mind, Language, and Reality*. Cambridge: Cambridge University Press.
- Salmon, Nathan. 2005. "Are General Terms Rigid?" *Linguistics and Philosophy* 28 (1): 117–34.
- Schwartz, Stephen P. 2002. "Kinds, General Terms, and Rigidity: A Reply to LaPorte." *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 109 (3): 265–77.
- Soames, Scott. 2002. *Beyond Rigidity: The Unfinished Semantic Agenda of Naming and Necessity*. Oxford: Oxford University Press. doi:10.1093/0195145283.001.0001.

Layers of Logical Consequence: Logical consequence as epistemically model-theoretic and metaphysically proof-theoretic

Grace Field*

field_grace@bell.net

Abstract

Model-theory and proof-theory are two long-standing alternative descriptions of logical consequence. Proof-theory characterises truth in terms of logical implication. Namely, according to proof-theory, the statement ‘ A implies B ’ is true iff there exists a proof from A to B . In contrast, model-theory characterises truth based on possible states of the world. By model-theory, ‘ A implies B ’ is true iff for any model m , if m satisfies A then m satisfies B . In this paper I argue that we can reconcile the views, by making an appropriate distinction between epistemic nature and metaphysical nature. Namely, I will argue that we can view logical consequence as epistemically model-theoretic and metaphysically proof-theoretic.

1 Introduction

When we say that something *logically implies* something else, we are appealing to the relationship of implication between a premise and a conclusion. This relationship is described by the theory of logical consequence. But how can we describe logical consequence itself? On what grounds can we say something logically implies something else? In other words, how should we think about the technical relationship between a premise and a conclusion? As of yet, there is not a universally accepted answer to this question, although the nature of logical consequence has been a subject of debate for decades.

Model theory and proof-theory have prevailed in parallel since the early days of the debate, as alternative descriptions of logical consequence. A model-theoretic view frames logical consequence in terms of truth preservation across cases, while a proof-theoretic view holds that consequence reduces to the existence of a formally valid argument between the premises and conclusion.

*Grace Field is just finishing a joint specialist degree in Physics and Philosophy at the University of Toronto. Next year she will continue to study philosophy of science at the graduate level, either at Cambridge or at LSE.

It is difficult to refute either on purely technical grounds, and some authors have suggested adopting some combination of the two. It even seems reasonable to argue that we *should* be searching for a compromise: neither view seems entirely counter-intuitive, or technically flawed. In this paper I will identify the relevant compromise with a distinction between epistemic and metaphysical nature. My aim is to defend logical consequence as *epistemically* model-theoretic, and *metaphysically* proof-theoretic.

In §2, I will describe the technical frameworks of model-theory and proof-theory in more detail, describe existing compromise-style views, and emphasise that they tend to view proof-theory and model-theory as alternative but individually complete descriptions of logical consequence. In §3, I will present a view on which neither offers a complete description of logical consequence. Rather, I will show that we can take model-theory as a description of the epistemic side to logical consequence, and proof-theory as a description of its metaphysical side. These descriptions are individually incomplete—but when taken together, they offer a complete description of logical consequence. The aim of §3 is to show that it is possible to adopt such a view. The aim of §4 will be to show that it is plausible to adopt such a view, because of the link between epistemic access and the model-theoretic cases, and because we have reason to believe that these case-by-case relationships are instantiations of underlying deductive relationships.

2 Background

First of all, let me describe the existing frameworks of proof-theory and model-theory in more detail. Conceptually, we can think of model-theory as a case-by-case description: ‘The *model-centred approach* to logical consequence takes the validity of an argument to be *absence of counterexample*’ (Beall and Restall 2016). In technical terms, consequence is model-theoretic iff it is described by the following equivalence relation: B is a consequence of A iff for any model m , if m satisfies A then m satisfies B (Tarski 1983, 417).

Proof-theory instead relies on the existence of argument-based reasoning. Again appealing to Beall and Restall, ‘On the *proof-centred approach* to logical consequence, the validity of an argument amounts to there being a *proof* of the conclusions from the premises’ (2016). So a consequence is proof-theoretic on the condition that an argument from A to B is valid iff there is a proof of B from A .

As it is, the two theories tend to be seen as alternative rather than compatible views. This state of affairs is exhibited particularly clearly in Shapiro’s article, ‘Logical Consequence, Proof Theory and Model Theory’, where he summarises the current state of the logical consequence debate. He writes, ‘[m]odel theory and proof theory each provide for a notion of logical consequence, but the notions employed by these two branches are quite different from each other, at least conceptually’ (Shapiro 2007, 651). Later in the article he comments on the attitude philosophers commonly have towards these theories. He writes that there are, in general, two schools of thought. One takes model-theory to be the primary essence of logical consequence,

the other takes proof-theory to be the primary essence of logical consequence (2007, 668). Proof-theory and model-theory are taken to be ‘different, but closely related, notions of logical consequence’ (2007, 669).

Even within the diverse landscape of existing compromise-style theory, in which authors do not commit themselves to pure proof-theory or pure model-theory, the two views tend to be seen as individually complete descriptions of logical consequence. Logical pluralists hold that there are several correct notions of logical consequence; Beall and Restall, Cook and Shapiro, and Carnap have each defended a unique version of this view (Beall and Restall 2000; Cook 2002; Shapiro 2006, Carnap 1959). Relativists have suggested that the correct notion might instead be domain-relative—variable but determined by the field we are working in. Some versions of relativism even identify logical domain with culture, resulting in culture-dependent logic (Lokhorst 1998, Greiffenhagen and Sharrock 2006). In the pluralist version, proof-theory and model-theory are seen as two correct but individually complete descriptions. On the relativist view, each theory is a complete description of logical consequence, but only within a certain domain.

At first sight, it does seem natural to treat them this way, as alternative but complete descriptions—they propose two different ways to describe the same thing. However, in the following section I will argue that we are not forced to adopt this attitude. Instead, if we make an appropriate distinction between the metaphysical and epistemic nature of logical consequence, we can view proof-theory and model-theory as two separately incomplete components of a full description.

3 The technical details, and why they are possible

To address the epistemic and metaphysical nature of logical consequence, we must begin with a clear definition for each type of property. For my theory of logical consequence, a property is *epistemic* iff it concerns knowledge or justified belief. Metaphysics is slightly more abstract by nature, and this is sure to be reflected in any attempt at a definition. I will stick to the following admittedly elusive, but meaningful notion: a property is *metaphysical* iff it concerns the most abstract essence of a concept.

The general frameworks of model-theory and proof-theory, as described in the previous section, clearly establish a set of minimal conditions for my theory of logical consequence. At the very least, I must be able to identify the epistemic side of logical consequence with a set of possible models and a corresponding satisfaction relation. And I must be able to identify its metaphysical nature with existence, or absence, of proof. Here I merely aim to show that it is at least *possible* for logical consequence to fit the strict technical requirements of model-theory in its practical use, and fit the equivalent requirements of proof-theory in its metaphysical nature. In the next section I will argue that this is not only an option, but a *plausible* option.

So, how can model-theory fit with the epistemic side to logical consequence? What are the

models, and what is the satisfaction relation? As Varzi notes, logic aims to act as a universal requirement for theoretical reasoning. As a result, its models must be universally applicable in some sense; ‘logic is a uniquely ambitious theory [...] It aims to be the theory included in every other theory [...] its models want to include the models of every other’ (Varzi 2002, 199). I propose we can access this universality by appealing to Carnap’s distinction between individual and general concepts. An individual concept is specific to a certain space-time coordinate, while the corresponding general concept has many possible space-time representations. Carnap’s pet dog Luchs exists at a given location at a given instant. This individual concept is one among many possible representations of the corresponding concept *dog* (Carnap 1959, 247–48). The fact that Carnap’s dog does not have some property X at that specific place and time is a representation of the general concept of X *not obtaining*.

We may adopt this distinction as a backbone for the epistemic nature of logical consequence, partly in virtue of its universality. I have claimed that epistemically, logical consequence can be described by satisfaction of models. I propose we should make the following identifications: *model* with the above-described notion of *individual concept*, satisfaction with representation, and subject of satisfaction with the above-described notion of *general concept*. Then my model-theoretic proposal for the epistemic side of logical consequence amounts to the claim that we observe individual concepts to come to *know and/or believe* consequence relationships between general concepts. Note the emphasis on observation, knowing, and believing here—these are all epistemically-charged terms, which is just what we are asking for since we are talking about the epistemically accessible side of logical consequence. In technical terms, this proposal means: a relationship of logical consequence from A to B is epistemically valid if and only if for all individual concepts m , if m is a representation of the general concept A , then m is a representation of the general concept B .

An immediate objection to this view might be that it fails to include a huge set of states of the world—namely, states of the world that do not explicitly depend on space or time, and therefore cannot be represented in the way I have described above. For such states, we must add to Carnap’s conception. The individual concept in such cases cannot be the discrete state itself. Instead, we may identify the individual concept as our knowledge or experience of that discrete state. The fact that this involves connecting models with experience is not an issue, since we are only attempting to deal with the epistemic nature of logical consequence. Overall, the proposed individual-general concept distinction is universally applicable in an epistemic sense, if it includes this adjustment for discrete states without inherent space-time coordinates.

To get a further sense of how this works, consider $\neg(\neg X) \models X$, the statement that X is a logical consequence of $\neg(\neg X)$. Let’s examine our epistemic access to this consequence relationship, on the view I have proposed above. The general concept on the left hand side of the implication is ‘it fails to obtain that some property of the world, X , fails to obtain’. This general concept can be represented by individual concepts that fail to lack some property of the world, where individual concepts are specific in space and time, either themselves or through space and time-localised experience. An example of such an individual concept would be the space-time event of me looking out my window and failing to find a field of view in which there

is not a german shepherd on my front lawn. The right hand side of the implication $\neg(\neg X) \models X$ is the general concept ‘some property of the world, X , obtains’. This might be represented by any number of individual concepts in which some property of the world obtains, for example by my observation that I hear a helicopters in the sky. My proposal for the epistemic nature of logical consequence claims that we can see our epistemic access to relationships like $\neg(\neg X) \models X$ by running through these individual concepts, and checking that every space-time specific and experientially accessible individual concept that represents ‘it fails to obtain that some property of the world, X , fails to obtain’ also represents ‘ X obtains of the world’. It should be noted here that we clearly cannot run through the infinite number of possible individual concepts available. This is an epistemic limitation. But we can satisfy ourselves with the idea that we gain knowledge of relationships like $\neg(\neg X) \models X$ by having experienced some large number of its individual concept instantiations. And the greater the number of instantiations we have experienced, the greater our confidence in the validity of $\neg(\neg X) \models X$. For example, when I look out of my window and fail to find a field of view in which there is not a German shepherd on my front lawn, at that same instant of time and space I observe that there is a German shepherd on my front lawn. This individual experiential and space-time specific concept can be seen as one of many such concepts that give me epistemic access to $\neg(\neg X) \models X$. When I listen to the sky and fail to avoid hearing the noise as helicopter engine noise, I am always simultaneously hearing helicopter engine noise. Having this experience, in addition to the previous experience, gives me even more confidence in the validity of $\neg(\neg X) \models X$.

So we have seen that it is possible to describe the epistemic nature of logical consequence model-theoretically, according to an experience-based description of models, and a representation-based description of the satisfaction relation. Next I need to show that it is possible to describe metaphysical nature of logical consequence proof-theoretically. This requirement is more straightforward to satisfy. Since a concept’s metaphysical nature is its abstract essence, we merely need to show it is possible to think of logical consequence as being fundamentally represented by abstract proof from premises to a conclusion. But we can imagine the existence of an argument from a set of premises to the corresponding conclusion, in the same way we can imagine the existence of an abstract mathematical proof for $x + x = 2x$. The difficulty for the proof-theoretic side of my view is not whether we can *possibly imagine* a metaphysical relationship between proof and consequence. The difficulty will be to establish whether we can *plausibly assert* the existence of that relationship. That will be the aim of the next section.

A satisfactory proof must be both rigorous and formal. Beyond that basic requirement, I do not claim to subscribe to any particular type of proof when I suggest proof-theoretic metaphysical essence. Any discussion of that kind raises myriad problems of its own; and if I were to choose one type of proof, my entire proposal for logical consequence would become opaque for anyone unsympathetic to my choice. Instead I propose a view in which logical consequence is epistemically model-theoretic, in the sense of individual-general concept experience-based modelling outlined above, and metaphysically proof-theoretic, according to some rigorous and formal type of proof.

By specifying its epistemic and metaphysical nature, we must have provided a complete

description of logical consequence. If we have both a description for how we can come to know and believe about a concept, and a description for how that concept exists independently in the abstract world, what more could we ask for? Epistemic nature and metaphysical nature, taken together, exhaustively describe any concept. Therefore, the sub-descriptions I have offered come together naturally as individually incomplete descriptions that jointly offer a complete description of logical consequence. Of the two parts of logical consequence—its epistemic nature and its metaphysical nature—one can be described by model-theory and the other can be described by proof-theory. In this way, my proposal allows us to transition from a view of *alternative* theories to a view of *compatible* theories, each of which only describes one part of the full concept we are trying to characterise.

4 The technical details, and why they are plausible

Now that we have established that it is *possible* to provide a view on which logical consequence is epistemically model-theoretic and metaphysically proof-theoretic, we are in a position to establish whether this is a *plausible* view to adopt. I will argue that it is, for several reasons. First, I will examine our reasons for subscribing to the specific models and satisfaction relation I have proposed. Based on this technical framework, with the models and satisfaction relation as I have described, I will then explain why we should think logical consequence is epistemically model-theoretic in the way I have described. Finally, I will examine our reasons for believing that logical consequence is metaphysically proof-theoretic, according to some rigorous and formal kind of proof.

To begin with, why should we subscribe to the models and satisfaction relation I have described? Our epistemic access to logical consequence is defined by the limits on our actual inference process. And I claim that our actual inference process follows a representationalist model-theoretic program, where models represent possible states of reality (Sher 1996, 658–61). I make this claim in light of the intuitive dependency between experience and epistemic access to the world. Carnap notably supports such dependency, in his comprehensive work *The Logical Structure of the World*. In his words, ‘I can make an “epistemic evaluation” of any experience I have had by stating to what extent this experience has added to my (theoretical) knowledge. This addition consists not only of the theoretical content of the experience itself, but also of whatever I can infer from this content with the aid of my earlier knowledge’ (Carnap 2003, 309). As he notes, we come to know or believe through assessing the nature of the world based on real experience, or imaginary but believable extension of experience. And an experience is a snapshot of a possible state-of-the-world. Our models and satisfaction relation, if they are to describe our epistemic access to logical consequence, should respect this relationship between epistemic access and experience. And it does not take much to get from this requirement to the models and satisfaction relation I describe. I merely take a model to be a space and time-specific state of the world, accessible or hypothetically accessible via experience, and the satisfaction relation to be the relationship of representation that holds between these localised state of the world and unlocalised, general concepts. Thus, I argue that it is plausible

to accept the models and satisfaction relation I have proposed for the epistemic side of logical consequence.

Having granted these models and this satisfaction relation, why should we think logical consequence is epistemically model-theoretic in the first place? In other words, why should we think our knowledge and beliefs about logical consequence involve case-by-case analysis of representational relationships between different individual concepts? A point raised by Shapiro is particularly relevant here. He notes that, even if we are working within a proof-theoretic system, the only way we can intuitively check the correctness of that system is by checking that its rules of inference ‘do not lead from truth to falsehood’ (Shapiro 2007, 667). This checking is an epistemic endeavour—we are asking, how can we *know, and test the world* to conclude that our proof-theoretic system of logical reasoning is correct? There is only one clear way to do this test in an epistemically accessible way—namely, by running through possible hypothetical states of the world. This amounts to leaning on the model-theoretic version of logical consequence I have described. Thus, it *is* plausible to believe that epistemically, we are limited to treating logical consequence model-theoretically.

Logical consequence might still be *metaphysically* proof-theoretic. But why should we see it this way? To fully argue for the plausibility of proof-theoretic metaphysical nature for logical consequence, we need to firmly establish whether it even needs any kind of metaphysical description at all. This brings us to a challenge tied up with the ancient debate on realism versus antirealism towards the existence of abstract universals. Namely, if we come to know and believe using a certain notion of a concept, why should we think there exists anything more abstract or fundamental to that concept’s full nature? A sceptic might claim: logical consequence does not have any metaphysical content beyond its epistemic nature, which we have already described as model-theoretic. I will argue that such scepticism is unwarranted—that we *do* have reason to believe in a non-trivial metaphysical essence for logical consequence. Then I will explain why we have reason to believe that this metaphysical essence is proof-theoretic.

We will be in a good position to identify gaps in the sceptic’s reasoning if we first examine the source of their inspiration. Hume’s treatise on causation and induction forcefully encourages doubt about the existence of abstract metaphysical essence for causal law. In *An Enquiry concerning Human Understanding*, he argues that we are fundamentally limited in our ability to identify causal relationships *a priori*, since all knowledge stems from case-by-case examination of experience (Hume 2008, 18–23). This position has since been widely acknowledged by scientists and philosophers alike (Carnap 2003, 265). Hume then settles on a description of cause-and-effect as mere cosmic regularity, questioning the notion of cause-and-effect as manifestation of abstract causal law (Armstrong 1993, 438; 1983, 4). Causation is not equivalent to logical consequence; but I believe we can construct an analogy here. Epistemic reliance on case-by-case method in science has led to widespread scepticism about the existence of any abstract scientific law beyond functional dependency. It seems that similar doubt could apply to existence of metaphysical essence for logical consequence - assuming its epistemic nature is model-theoretic, and therefore reliant on case-by-case relationships. Definitely, this model-theoretic epistemic nature restricts our epistemic access to any proof-based abstract nature that

might exist.

However, we cannot directly conclude from this that an abstract proof-based nature does not exist. I will argue that it likely *does* exist, beginning with an appeal to work in logic-as-modelling. I have emphasised that, if epistemic and metaphysical nature do both carry weight, they are not equivalent properties. One useful way to characterise this difference allows us to draw from Shapiro's work on logic-as-modelling. Namely, we can easily view the epistemic nature of logical consequence as a model. Here, I mean model in the colloquial sense of approximate representation—I am not referring to the individual-general concept introduced earlier for my treatment of model-theory. In general, epistemic nature is defined by the boundaries of knowledge; therefore, it is in some sense a knowledge-based approximation of a concept that *might* have deeper metaphysical qualities. Key here is the approximate in approximate representation. A model is not necessarily a complete representation of the concept it stands for; 'there is almost always a gap between a model and what it is a model of' (Shapiro 2006, 50). This point helps Shapiro to develop his logic-as-modelling view, but I will use the same point to claim that there is likely an underlying metaphysical nature to logical consequence, above and beyond its model-theoretic epistemic nature.

The above reasoning only works if the epistemic nature of logical consequence is in fact a knowledge-based approximation. So how do we know it is an approximation of this kind? Here I can use the sceptic's point for my own benefit. As the sceptic notes, if the epistemic nature of logical consequence is model-theoretic, it is limited in its access to *a priori* reasoning. Therefore it is a limited description. And a limited description is an approximate description. It is an approximate model, in the same way that 'a collection of point masses is a model of a system of physical objects, and the Bohr construction is a model of an atom' (Shapiro 2006, 49). However, every model models something—every approximate representation must represent some non-trivial exact concept. I suggest that this exact concept is precisely the metaphysical nature of logical consequence. Epistemic nature covers everything we are able to know about logical consequence relationships. And everything we are able to know creates a model of the non-trivial abstract metaphysical reality, which is outside the realm of our direct knowledge. Therefore, the sceptic's emphasis on our epistemic constraints offers indirect support for my defence of non-trivial metaphysical nature.

Furthermore, and crucially, there is no fundamental inconsistency in a view that includes non-trivial essence from both sides, epistemic and metaphysical. Here we are encouraged by various influential philosophers, who subscribe to Hume's work on scientific method yet maintain realism on the existence of universal law. Again, we work with an analogy, comparing the Humean notion of scientific method with epistemic model-theoretic essence, and the existence of fundamental causal law with the existence of non-trivial metaphysical essence.

From the early 20th century, Carnap and Russell are among many explicit advocates for Hume's view on the epistemic limits of *a priori* reasoning. Still they do not deny the existence of fundamental law, as metaphysical nature. See the following excerpt from Carnap on the metaphysical essence of causal correlation: '[h]ere we do not simply ask between what object

the relation obtains, but what it is between the correlated objects, by virtue of which they are connected' (Carnap 2003, 35). And, even more explicitly, 'the essence problems belong to *meta-physics*' (2003, 35). Still, he sees our epistemic experience of specific causal relationships as a manifestation of mere functional dependency (2003, 264). Russell, whose position on epistemic empiricism is steadfast, still does not rule out the existence of causal law as an abstract mathematical formulation (Russel 1913, 14). He writes, 'there is no *a priori* category of causality, but merely certain observed uniformities' (1913, 24). However, '[i]n all science we have to distinguish two sorts of laws: first, those that are empirically verifiable but probably only approximate; secondly, those that are not verifiable, but may be exact' (1913, 16). He clearly acknowledges the possibility of exact abstract law, in spite of its inherently unverifiable nature.

From the more modern tradition, Armstrong and Davidson hold a similar position. Both endorse Humean views on method but accept abstract law (Armstrong 1993, 438; Davidson 1967, 701–02). In Armstrong's words, 'It is true that there appears to be no *a priori* argument that takes one from singular causation to law' (1993, 438). However, '[i]t may be noted that the unity of the space-time world is not constituted by the mere conjunction of the state of affairs [...] [t]he real *unity* is given by the fact that all the particulars are directly or recursively linked to each other by real, that is external, relations' (1993, 435). Davidson emphasizes that we are often limited in our epistemic access to external causal law—but notes: 'very often, I think, our justification for accepting a singular causal statement is that we have reason to believe an appropriate causal law exists, though we do not know what it is' (1967, 701).

Armstrong presents an explicit argument for his subscription to the separate existence of abstract law as a metaphysical basis for correlation. We either see correlation as instantiation of a universal and abstract entity, or instantiation of a mere regularity. The abstract entity in the former view acts as a basis for *explanation*. Mere regularity does not—any attempt at explanation would be circular, explaining correlation in terms of correlation. Armstrong believes the former view is inherently more desirable, simply because genuine explanation is desirable (Armstrong 1983, 40–41).

For logical consequence we have further reason to prefer the former view, even if we do not share Armstrong's belief in the intrinsic appeal of explanation. Genuine explanation implies a degree of modal transparency. A genuine explanation can describe why a particular consequence relation *must*, or *should* hold. Thus, by maintaining an anti-sceptic position on the existence of abstract metaphysical nature for logical consequence, we automatically ease modal problems associated with a purely model-theoretic view. These problems have been examined at length by Prawitz (2005).

Furthermore, we can appeal to an intuition raised by Prawitz. As he points out, purely model-theoretic consequence would be intuitively back-to-front. On such a pure view, 'we cannot really say that we infer the truth of the conclusion by the use of a valid inference. It is, rather, the other way around: we can conclude that the inference is valid after having established for all inferences of the same form that the conclusion is true in all cases where the premises are' (2005, 675). This is a crucial and telling point to acknowledge. Intuitively, we want some part

of logical consequence to be a fundamental property of relationships between abstract ideas. We lose this if the validity of an inference depends purely on running through all possible interpretations or states of the world. In some sense we are demoting logical consequence, from fundamental to dependent—in particular, to dependent on possible discrete states of the world. On the view I have proposed, logical consequence is instead metaphysically dependent on rules for argument and proof, from which possible discrete states are derived. Here we see a key benefit of my view, for solving Prawitz's intuitive problem with model-theory: namely, the ability to push logical consequence back to its intuitive status, without denying Humean limits on our *epistemic* access to *a priori* reasoning.

Thus, we have established that we do have grounds to assert that logical consequence has non-trivial metaphysical content. I still need to explain why that metaphysical nature should plausibly be seen as proof-theoretic. I argue that the past few paragraphs serve as a perfect explanation, in their association of metaphysical nature with *law-like* concepts. Implicitly, in my argument for the non-triviality of metaphysical nature for logical consequence, I have shown that this metaphysical nature is associated with some kind of law-like system of relationships. Proof-theory fits this description perfectly. It is a deductive system, based on rules that determine the relationship between premises and a conclusion. Therefore, given that we should not neglect the metaphysical nature of logical consequence, that metaphysical nature is plausibly proof-theoretic.

Overall, then, my proposal amounts to the idea that we never observe the underlying abstract proofs that exist, just like we never directly observe causal laws. We observe logical regularities that we take to be epistemically accessible instantiations of more abstract and general deductive relationships, just like we take physical observations to be epistemically accessible instantiations of more abstract and general causal laws. You might object to this by claiming that you *do* feel as if you reason using proof-theory, and that you *do* think you picture your knowledge in terms of proof theory. The response to this is clear: proof-theoretic reasoning is useful as a tool, just like reasoning in terms of general causal laws is useful for understanding. Nevertheless, we do not strictly have epistemic access to the underlying proof-theoretic relationships themselves, just like we do not strictly have epistemic access to causal laws.

5 Conclusions

It may seem as if the essence of logical consequence is model-theoretic, if we subscribe to a model-theoretic program when we come to know or believe consequence relationships. But it is possible to see logical consequence as metaphysically proof-theoretic even if we endorse a model-theoretic program for its epistemic nature. We can examine the potential for existence or dominance of fundamental *a priori* causal law as a basis for analogy. We have seen that on this issue, the historic rivalry between empirical versus *a priori* views is not in fact a strict rivalry. Limits on our epistemic access to abstract nature might apply, given a model-theoretic epistemic nature. But if such limits do exist, they seem to imply approximate epistemic representation of

some non-trivial exact metaphysical entity. Non-trivial abstract nature might still exist, and metaphysically dominate. And intuitively we understand logical consequence as a *fundamental* relation, from which cases should derive, not follow. I have proposed that we can adopt a hybrid view, one that will satisfy each of these appeals to analogy and intuition. By allowing for a distinction between epistemic and metaphysical nature, we can characterise logical consequence as epistemically model-theoretic but metaphysically proof-theoretic.

References

- Armstrong, D. M. 1983. *What is a Law of Nature?* Cambridge: Cambridge University Press. doi:10.1017/CB09781139171700.
- . 1993. “A World of States of Affairs.” *Philosophical Perspectives* 7 (Language and Logic): 429–40.
- Beall, JC, and Greg Restall. 2000. “Logical Pluralism.” *Australasian Journal of Philosophy* 78 (4): 475–493. doi:10.1080/00048400012349751.
- . 2016. “Logical Consequence.” In *The Stanford Encyclopedia of Philosophy*, Winter 2016, edited by Edward N. Zalta. <https://plato.stanford.edu/archives/win2016/entries/logical-consequence/>.
- Carnap, Rudolf. 1959. *The Logical Syntax of Language*. Paterson: Littlefield, Adams, & Co.
- . 2003. *The Logical Structure of the World*. Chicago: Open Court Publishing Company.
- Cook, Roy T. 2002. “Vagueness and Mathematical Precision.” *Mind* 111 (442): 225–47.
- Davidson, Donald. 1967. “Causal Relations.” *The Journal of Philosophy* 64 (21): 691–703.
- Field, Hartry. 2008. *Saving Truth From Paradox*. Oxford: Oxford University Press.
- Greiffenhagen, Christian, and Wes Sharrock. 2006. “Logical Relativism: Logic, Grammar, and Arithmetic in Cultural Comparison.” *Configurations* 14 (3): 275–301. doi:10.1353/con.0.0016.
- Hume, David. 2008. *An Enquiry Concerning Human Understanding*. Oxford: Oxford University Press.
- Lokhorst, Gert-Jan C. 1998. “The Logic of Logical Relativism.” *Logique et Analyse* 41 (161/163): 57–65.
- Prawitz, Dag. 2005. “Logical Consequence: A Constructivist View.” In *The Oxford Handbook of Philosophy of Mathematics and Logic*, edited by Stewart Shapiro. Oxford: Oxford University Press. doi:10.1093/0195148770.003.0022.
- Quine, W. V. O. 1948. “On What There Is.” *The Review of Metaphysics* 2 (5): 21–38.

- Russel, Bertrand. 1913. "On the Notion of Cause." *Proceedings of the Aristotelian Society* 13 (1): 1–26. doi:10.1093/aristotelian/13.1.1.
- Shapiro, Stewart. 2006. *Vagueness in Context*. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199280391.001.0001.
- . 2007. "Logical Consequence, Proof Theory, and Model Theory." In *The Oxford Handbook of Philosophy of Mathematics and Logic*, edited by Stewart Shapiro, 651–71. Oxford: Oxford University Press. doi:10.1093/oxfordhb/9780195325928.003.0021.
- Sher, G. Y. 1996. "Did Tarski Commit "Tarski's Fallacy"?" *The Journal of Symbolic Logic* 61 (2): 653–686.
- Smith, Peter. 2011. "Squeezing Arguments." *Analysis* 71 (1): 22–30.
- Tarski, Alfred. 1983. "On the Concept of Logical Consequence." In *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*. Indianapolis: Hackett.
- Varzi, Achille C. 2002. "On Logical Relativity." *Philosophical Issues* 12 (1): 197–219. doi:10.1111/j.1758-2237.2002.tb00067.x.