Contents

Research Articles

Organon F 32 (3) 2025: 262–299 ISSN 2585-7150 (online)
https://doi.org/10.31577/orgf.2025.32301

RESEARCH ARTICLE

Source of Reality/Causal Capacity: Outside of the Priority Chain?

Savvas Ioannou*

Received: 7 June 2024 / Revised: 23 March 2025 / Accepted: 31 March 2025

Abstract: I will consider two views about the structure of reality: metaphysical foundationalism (all grounded entities are fully grounded in ungrounded entities) and metaphysical infinitism (there are infinite chains of grounding that lack a foundation). Foundationalists motivate their view by appealing to an intuition that there must be a fundamental level that is the source of reality (Schaffer) or causal capacity (Trogdon) of grounded entities because a grounded entity inherits its reality or causal capacity from its ground and there would be a vicious infinite regress otherwise. I will argue that this argument is not successful as the source of reality or causal capacity of an infinite chain can be its cause. For example, this source can be the direct physical cause, the first physical cause or a Prime Mover. A vicious infinite regress is avoided this way too because there is no transference of the same status ad infinitum.

Keywords: Fundamentality; metaphysical foundationalism; metaphysical infinitism; vicious infinite regress.

* University of Cyprus
https://orcid.org/0000-0002-3930-0807
Department of Classics and Philosophy, University of Cyprus, 1 Eressou Street, First Floor, P.O. Box 20537, CY-1678, Nicosia, Cyprus
savvasioa@hotmail.com

1. Introduction

In this paper, I will consider two views: metaphysical foundationalism (all the chains of ontological dependence terminate in fundamental entities, or all derivative entities are fully grounded in fundamental entities) and metaphysical infinitism (there are infinite chains of grounding that lack a foundation). To begin with, I will explain the notion of 'grounding' by mentioning how it is used in the literature. Then, I will talk about an argument for the existence of a fundamental level. A foundationalist intuition is that there must be a source of reality (Schaffer, 2009; 2010; 2016) or causal capacity (Trogdon 2018) because a grounded entity inherits its reality or causal capacity from its ground. If infinitism was true, "[b]eing would be infinitely deferred, never achieved" (Schaffer 2010, 62). I find this argument unpersuasive as the source of reality or causal capacity of an infinite chain can be outside of it. This source can be the cause of the infinite chain. I will argue that the cause of a grounded object can also explain why the grounded object is real and has causal capacity; nothing is left unexplained if there is no fundamental level. A vicious infinite regress can be avoided this way too. So, if this foundationalist argument is the only reason why somebody is foundationalist, they should stop endorsing foundationalism.

2. Grounding

Metaphysical foundationalism and metaphysical infinitism are usually described by using the notion of 'grounding'. Therefore, examining the literature concerning grounding before going into more detail about the two aforementioned views is helpful.

Grounding accounts differ on what they take as the relata of the grounding relation. While some have taken the relata to be sentences or facts (Fine 2001; 2012; Rosen 2010; Audi 2012; Dasgupta 2014), Schaffer (2009; 2012; 2016) has taken the relata to be members of any two ontological categories. Following Heil (2003; 2012), I consider objects property-bearers (i.e., substances). Objects are things like elementary particles and living organisms that possess certain properties. A fact can also be described as a state of affairs or a particular way that the world is. Examples of facts are the fact

that a person is happy and the fact that a table is brown. I will assume that Schaffer's account is correct and present it here. One motivation for following Schaffer's account is that it allows grounding between objects. As one of the issues in this paper is to decide whether there is a fundamental level of objects, this is a very relevant account of grounding.

According to Schaffer (2009), what exists are the grounds/fundamental entities, grounding relations[1], and the grounded entities that are generated from fundamental entities through grounding relations. While there is an abundant roster of grounded entities, they are grounded on a sparse basis. Dependent entities are not fundamental. The grounded entities are an ontological free lunch.

Schaffer (2009) has defined the notions of a fundamental entity (primary, independent, ground entity) and derivative entity (secondary, dependent, grounded entity) in terms of grounding (ontological dependence, priority in nature). "Fundamental: x is fundamental $=_{df}$ nothing grounds x […]. "Derivative: x is derivative $=_{df}$ something grounds x" (Schaffer 2009, 373).

Schaffer (2010) has argued that the priority relations among concrete objects form a well-founded partial ordering. There is a partial ordering structure because priority relations are irreflexive, asymmetric, and transitive.[2] [3] There is also a well-founded ordering: all priority chains terminate

[1] Not everybody explains grounding as a relation. For Fine (2001, 16; 2012, 43) and Correia (2010), grounding is best regarded as a sentential operator that connects the sentences that state the ground to the sentence that states what is grounded. Dasgupta (2014) has also taken grounding to be a sentential operator but claimed that a plurality of sentences can ground a plurality of sentences.

[2] This is the most prominent way to understand grounding. Schaffer (2012; 2016) has changed his mind and given a contrastive treatment of grounding. The contrastive treatment involves viewing grounding not as a binary relation between two actual nonidentical entities but as a quaternary relation, including a non-actual grounding contrast and a non-actual grounded contrast. E.g., *"The fact that* φ *rather than* φ^* *grounds the fact that* ψ *rather than* ψ^**"* (Schaffer 2012, 130). Likewise, irreflexivity, asymmetry, and transitivity can be understood as holding between differences. We do not need to discuss the details here as they will not affect my argument.

[3] All of these are controversial though. For instance, Jenkins (2011) has argued that grounding is reflexive. Barnes (2018) has argued that grounding is nonsymmetric. Bennett (2017) has responded to them and defended the claim that grounding

in something fundamental.[4] There are basic actual concrete objects because priority chains terminate.

Grounding can also be explained by comparing it with causation. Schaffer (2012; 2016) has argued that grounding is analogous to causation. Both are relations of generation, partial orders, and back explanation. Both causal and grounding relations can be described similarly using terms such as 'generation', 'production', 'making', and 'dependence'. Concerning partial ordering, both of these relations are irreflexive, asymmetric, and transitive binary relations. Concerning explanation, we can explain why something is the case by giving a causal or grounding story.[5] [6]

Grounding has been considered as one among many building relations (Bennett 2011b; 2017). For Bennett (2017 32), "all building relations are

i. directed, in that they are antisymmetric and irreflexive,
ii. necessitating, roughly in that builders necessitate what they build, and
iii. generative, in that the builders generate or produce what they build. Built entities exist or obtain because that which builds them does".

According to Bennet (2017), building relations form a unified family. When we talk about building, we quantify over the unified class of building

relations are irreflexive and asymmetric. However, she has denied that grounding relations are transitive.

[4] How exactly we should define well-foundedness is controversial (see Rabin and Rabern 2016; Dixon 2016). I will present Dixon's view later in this paper. Tahko (2018) has noticed that while some foundationalists, such as Schaffer, use a set-theoretic notion of well-foundedness, others do not (e.g., Rabin and Rabern, Dixon). I will not choose between these notions, as my arguments in this paper do not rely on any specific notion of well-foundedness.

[5] Schaffer (2012; 2016) has argued that both causal and grounding relations are best formalised via structural equation models which incorporate contrastive information (these models were firstly introduced to explain causation). He has claimed that structural equation models for grounding provide more structure than the mere partial ordering mentioned by Schaffer (2009). This account will be relevant in section 5; so, it will be briefly described there.

[6] This is what Schaffer has believed. Still, it may be the case that some things may admit only one type of explanatory story, even if others might admit to either. The objection I will give does not need to agree with this view of Schaffer.

relations. There is not a single very abstract building relation that is either more fundamental than the specific building relations or the only building relation. For example, grounding, causation, and composition were conceived as building relations. Considering building relations a unified family will be relevant in section 5.

The aforementioned point (iii) will be especially relevant. Bennett (2017, 184) has claimed that "all building relations are generative in the sense that they license 'makes it the case' and 'in virtue of' talk". This is neutral on why such talk is licensed. It can be licensed because "some relations are generative, and some are not; it's just primitive" (Bennett 2017, 184). Otherwise, it may be the case that "there is no such relation, only the talk. It is just a matter of convention that certain relations license certain ways of talking" (Bennett 2017, 184). "On the first approach, the world decides which relations count as building relations" (Bennett 2017, 185).

Whether building talk is licenced by something metaphysical or is merely a matter of convention will be relevant in section 5. The objection to the foundationalist argument will be developed differently depending on how we explain building talk. But before we see this, metaphysical foundationalism and metaphysical infinitism will be explained.

3. Metaphysical Foundationalism and Metaphysical Infinitism

Grounding is used to describe two major views about the universe's structure: metaphysical foundationalism and metaphysical infinitism.[7] While foundationalism says that all grounded entities are ultimately grounded in ungrounded entities, infinitism does not.

According to metaphysical foundationalism (Cameron 2008b; Schaffer 2009; 2010; 2016; Bennett 2011a; Trogdon 2017; 2018), there is a fundamental entity (priority monism), or there are several fundamental entities (priority pluralism).[8] For example, a priority monist could claim that the whole

[7] Metaphysical coherentism (see Bliss 2014) is another view, but it is out of the scope of this paper.

[8] I presuppose that zero grounding is not the correct way to talk about objects. Objects are either grounded or ungrounded. According to Fine (2012, 47-48),

universe is an ungrounded object, and a priority pluralist could claim that elementary particles are ungrounded objects. Ungrounded objects are not grounded in anything else. More precisely, foundationalism claims that all grounded entities are directly or indirectly (i.e., by transitivity) grounded in ungrounded entities. For Schaffer (2009), all priority chains terminate. The ordering of grounding is well-founded: a priority chain is well-founded, iff it terminates in something fundamental/ungrounded. An ungrounded entity or entities can also be called a 'foundation'. A well-founded chain is not infinite at the fundamental end but may consist of infinite grounded entities. Priority pluralism can be represented this way: a grounded object is grounded in its grounds, and arrows represent grounding relations (e.g., Y_1 grounds X).

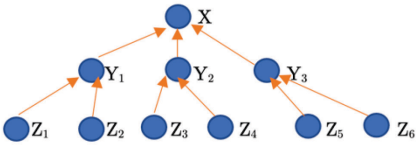

Figure 1: Priority Pluralism

X is a grounded object, Ys are its grounds, and Zs are the grounds of Ys.

On the contrary, metaphysical infinitism (Bohn 2018; Morganti 2009; 2014; 2015) claims that there are limitless chains of grounding that lack a foundation. An entity is grounded in another; the second is grounded in

something is zero-grounded, iff it is grounded in nothing. It is generated from a zero number of objects. Zero-grounded and ungrounded are different things. *"The case in which a given statement is zero-grounded, i.e. grounded in zero antecedents, must be sharply distinguished from the case in which it is ungrounded, i.e. in which there is no number of statements – not even a zero number – by which it is grounded. We may bring out the difference by means of an analogy with sets. Any non-empty set {a, b,. . .} is generated (via the "set-builder") from its members a, b,. . .. The empty set { } is also generated from its members, though in this case there is a zero number of members from which it is generated. An urelement such as Socrates, on the other hand, is ungenerated; there is no number of objects – not even a zero number – from which it may be generated"* (Fine 2012, 47).

An empty set is generated by its zero-ground. 'Socrates is identical to Socrates' is zero-grounded: it is grounded in nothing at all. However, an ungrounded truth or object is not generated at all.

another, and so on ad infinitum. The ordering of grounding is not well-founded as the priority chains do not terminate in something fundamental. The entities in question could be facts, objects, or other ontological categories. Concerning objects, different infinitist versions can be suggested. Whether grounding relations go from the larger to the smaller or the opposite is controversial. An infinitist may claim that the world is gunky: every object is a whole grounded in its proper parts. There is no bottom level. There is an infinite descent of levels, and each further level is grounded in the former. Another infinitist option is to claim that the world is junky: every object is a proper part of something, and the former is grounded in the latter. Alternatively, an infinitist could believe both and argue that the world is hunky (both gunky and junky) (e.g., Bohn 2018).[9] A gunky infinitist world can be represented this way, where the small circles represent that the chain continues ad infinitum:

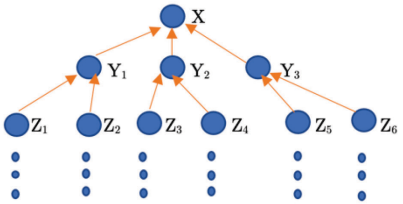

Figure 2: Gunky Infinitism

4. Source of Reality or Causal Capacity: Fundamental Objects

Metaphysical foundationalists (Cameron 2008b; Heil 2003; Schaffer 2010; 2016; Trogdon 2018) have argued that metaphysical foundationalism is true. It has been argued that there is a need for a source of reality or

[9] According to Bohn (2018, 175–76), in hunky worlds, grounding is asymmetrical. Still, someone could believe that hunky worlds result in violations of the asymmetry of grounding. The objection that I will give does not depend on whether grounding is asymmetrical, and so, I will not choose between these two views.

causal capacity[10] to avoid a vicious infinite regress and only foundationalism can provide us with one. In what follows, I will present this argument. In the next section, I will argue that it doesn't work.

Cameron (2008b) has endorsed an intuition that a fundamental layer of reality must exist. This intuition arises when we consider the whole as grounded in its parts. According to this intuition, grounding cannot go on ad infinitum: all the grounded objects must be grounded in a fundamental level. If metaphysical infinitism was true, "composition could never have got off the ground. If the existence of each complex object depends for its existence on the existence of the complex objects at the level below, and if we never reach a bottom level, then it is hard to see why there are any complex objects at all" (Cameron 2008b, 6). If everything were grounded in something else, "there would be no grounding to being: there would be no end to explanation when we try to explain why what there is exists" (Cameron 2008b, 6-7). "There must be a metaphysical ground, a realm of ontologically independent objects which provide the ultimate ontological basis for all the ontologically dependent entities" (Cameron 2008b, 8).

Schaffer (2010) has also shared this intuition.

> If one thing exists only in virtue of another, then there must be something from which the reality of the derivative entities ultimately derives" (Schaffer 2010, 37). If infinitism was true,

[10] Mentioned philosophers who use this terminology do not define what it is to be a source of reality or causal capacity. I take it to be something that can be an ultimate explanation: something that can be used to end an explanation of why something is real or has causal capacity. Given a source, we have ultimately explained why something is real or has causal capacity. This source can show why there is no need to keep asking why something is real or has causal capacity. A source of reality or causal capacity can be such an ultimate explanation because it is the reason that something is real or has causal capacity.

It could be asked why X is real or has causal capacity at time *t*. Later, I will argue that X's source of reality or causal capacity can occur earlier than *t*. By using the expression 'source of reality or causal capacity', I do not presuppose that the object that needs a source and its source exist simultaneously. I will argue that this presupposition is unwarranted.

I will mainly talk about a grounded object's source of reality or causal capacity. I will make it clear when this is not the case.

> "[b]eing would be infinitely deferred, never achieved (Schaffer 2010, 62).
>
> Grounding must be well-founded because a grounded entity inherits its reality from its grounds, and where there is inheritance there must be a source. One cannot be rich merely by having a limitless sequence of debtors, each borrowing from the one before. There must actually be a source of money somewhere. (Schaffer 2016, 95).

It has been argued that if infinitism is true, there is a vicious regress; so, we should reject infinitism. For Schaffer (2016, 95), "a regress counts as vicious if and only if there is an endless chain of dependency with transference of the relevant status". This sort of transference leads to the need for a source. In the grounding case, there is such a transference of reality: the grounded entity exists in virtue of its grounds. "That is why a source of reality is needed, in order for there to be anything to transfer" (Schaffer 2016, 96).

Instead of talking about the property of being real to develop the foundationalist intuition, some foundationalists have developed this intuition by talking about causal powers (Heil 2003; Trogdon 2018). Heil (2003, 19) has been sceptical about the possibility of a universe with no fundamental level. He has confessed ignorance as to how it is supposed to work given the dependence of higher levels on those below them: something, it seems, must ground the superstructure. If someone thinks that higher-level causal relations depend on lower-level causal relations, it is not clear that these could fail to bottom out. If the only unattenuated causal relations are those at the basic level, there had better be a basic level.

Trogdon (2018) has developed this kind of reasoning. According to him, instead of focusing on the property of being real, we should focus on the property of having the capacity for causal activity (causal capacity for short). "An entity has this property just in case it has causal powers, dispositions to enter into particular sorts of causal transactions" (Trogdon 2018, 191).

My objection against Schaffer's and Trogdon's arguments, which I will develop in the next section, does not depend on whether we talk about the source of reality or the source of causal capacity. So, I will remain neutral

between these two options and not mention Trogdon's objection against the foundationalist argument that mentions reality inheritance.

Trogdon (2018) has argued for causal foundationalism. "Causal foundationalism: necessarily, any nonfundamental entity with causal capacity is fully grounded by fundamental entities" (Trogdon 2018, 191). Causal foundationalism is true because the following three premises are true:

> The causal inheritance premise: necessarily, if A is nonfundamental and has causal capacity then A inherits its causal capacity from whatever fully grounds it.
> The source of causal capacity premise: necessarily, if A inherits its causal capacity then there are Δ that are a source of A's causal capacity (i.e. A inherits its causal capacity from Δ, and no entity among Δ inherits its causal capacity).
> The causality/fundamentality premise: necessarily, if Δ are a source of A's causal capacity then the entities among Δ are fundamental and Δ fully ground A" (Trogdon 2018, 192).

The causal inheritance premise is inspired by Kim's (2005) causal exclusion argument. This premise is supported by two principles. First, Kim's causal exclusion principle states that no property instance has simultaneous full causes. Second, the causal closure of grounding principle: if a property instance has a full non-fundamental cause, then whatever fully grounds that cause is also a full cause of the property instance. The premise can be supported by considering an instance of φ (lowercase Greek letters range over properties) that is non-fundamental and fully causes an instance of property ψ. Following the causal closure of grounding, there is a property instance that fully grounds the φ-instance and fully causes the ψ-instance. It follows that the ψ-instance has two simultaneous full causes. Following the causal exclusion principle, no event has two simultaneous full causes. Therefore, by reductio, it is false that there is an instance of φ that is both nonfundamental and fully causes an instance of ψ.

Concerning the source of causal capacity premise, Trogdon (2018) has believed that the rationale for it appeals to general considerations about inheritance. Specifically, it appeals to the following principle:

> The inheritance principle: necessarily, if A inherits φ then there are Δ that are a source of A's φ-ness (i.e. A inherits φ from Δ and no entity among Δ inherits φ) (Trogdon 2018, 186).

The source of causal capacity premise is motivated by the inheritance principle, which is assumed to be true. Trogdon (2018) has mentioned that Schaffer (2010) has presented some good reasons for the truth of the inheritance principle (reasons I have presented above).

Concerning the causality/fundamentality premise, Trogdon (2018) has given some reasons to believe it. Given the causal inheritance premise and that none of the entities among Δ has their causal capacity by inheritance, the entities among Δ are fundamental. Given that A inherits its causal capacity from Δ, A is fully grounded by Δ. We get this result because the following thesis is true: it is necessary that if A inherits its causal capacity from Δ, then the latter fully ground the former. This thesis can be supported by paradigmatic cases of causal capacity inheritance, such as when relevant property instances characterise the same entity (e.g., DNA molecules and a gene), when relevant property instances characterise distinct but materially coincident entities (e.g., a lump of clay and a statue), and when relevant property instances characterise objects at different levels of mereological aggregation (e.g., carbon molecules and a diamond). In these cases, the inheritors are fully grounded by the entities from which they inherit. This gives us a reason to believe that it is necessary that if A inherits its causal capacity from Δ, then the latter fully ground the former.

In the next section, I will present a new objection against Schaffer's and Trogdon's views.[11] According to it, even if we have an intuition or a belief that there must be a source of reality or causal capacity in order to avoid a vicious infinite regress, this source can be found outside the grounding chain. In other words, a vicious infinite regress of reality/causal capacity can be avoided without endorsing foundationalism.

[11] Different objections have been stated against the foundational intuition. Bliss (2013; 2014) has argued that Schaffer is begging the question. Bohn (2018) has claimed that he does not share the same intuition as Schaffer. Morganti (2014; 2015) has rejected the inheritance account. These objections are out of the scope of this paper.

5. Alternative Sources of Reality and Causal Capacity

I am not persuaded that our intuitions can be used to build a sound argument for foundationalism. What is behind our foundationalist intuitions is our search for a source of reality or causal capacity. This is shown explicitly in Trogdon's (2018, 192) "causal inheritance premise: necessarily, if A is nonfundamental and has causal capacity then A inherits its causal capacity from whatever fully grounds it". However, it seems that these sources can be found even if infinitism is true. In what follows, I will mention 'source of RCC', where 'RCC' means 'reality or causal capacity'. I will do so because independently of whether foundationalists choose to speak about reality or causal capacity, a similar objection against their view can be stated. According to the foundationalist argument, given the need for a source of RCC and given the truth of a specific account of grounding, foundationalism follows. I will argue that even if there is a need for a source of RCC and that particular grounding account is true, alternative plausible metaphysical views can be suggested. Instead of searching for the source of RCC inside the priority chain, we could search for this source outside the priority chain. The cause of a grounded entity can be its source of RCC. More specifically, alternative plausible sources can be the physical cause of each grounded entity (the direct[12] cause or the initial cause of the universe) or God.[13]

Different objections against the foundationalist argument can be stated depending on which account of causation is assumed to be true. As a result, I will examine different accounts of causation and argue that the foundationalist argument fails in every case. In the first part of this section, I will

[12] E_1 directly causes e_2, iff e_1 causes e_2 and there is not an event e_3 between e_1 and e_2 such that e_1 causes e_3 and e_3 causes e_2. E_1 indirectly causes e_2, iff there is an event e_3 between e_1 and e_2 such that e_1 causes e_3 and e_3 causes e_2.

[13] Nothing related to my argument relies on the existence of a God. If someone does not like to assume the existence of God, one cannot use this as an objection to my view. Alternative *physical* sources of RCC can be used and will be presented below to show that foundationalism is not the only way to avoid a vicious infinite regress. Examples that involve God's intervention are only given to clarify my claims.

assume the truth of a generative and power-conferring account of causation (such as the productive account of causation or the dispositionalist account of causation[14]). By 'generative', I mean that causes generate the effects: they bring about the existence of the effects. In this section, I will use 'generative' in its metaphysical reading; that is, causation talk is licensed because some relations are generative, and some are not. Building talk is not merely a matter of convention. By 'power conferring', I mean that the cause transmits causal powers to the effect (that is, the effect has its causal powers in virtue of its cause). For example, object Y has power P_2 because it was caused by object X with power P_1.

This account of causation also allows that the cause transmits reality to effect (i.e., the effect is real in virtue of its cause), but I will talk about causal powers in what follows just for ease of exposition. There is a similarity between causation and grounding: both are building relations (see Bennett, 2011b; 2017). If a generative, power-conferring account of causation is true, alternative sources of RCC can be the immediate/direct physical cause of a grounded object, the first physical cause of the universe, or God (I will give examples below).

In the second part, I will assume that a non-generative and non-power conferring account of causation is true: the structural equation models account of causation. This is the account that Schaffer endorses. It is a non-generative account of causation because according to it, causes do not bring about the existence of their effects. It is also non-power conferring because causes do not transmit causal powers to their effects according to it. What I will say will apply to other non-generative and non-power conferring accounts of causation (such as the counterfactual account of causation). If a non-generative, non-power conferring account of causation is true, a source of RCC may not be needed. Maybe we live in an eternal universe (i.e., a universe that always existed), and so, there is no need for a source of RCC. Every entity always had its RCC. There was no time when an entity came into existence, and causal powers were transferred into it. So, we should not search for an entity that is the source of RCC of a grounded entity.

[14] For productive accounts of causation, see Fair (1979) and Castaneda (1984). For dispositionalist accounts of causation, see Bird (2007), Heil (2003; 2012), and Shoemaker (1980; 1998; 2007).

Otherwise, God can be an alternative source of RCC. In this case, there is a kind of causation that is generative and power conferring (God's causation), even though physical causation is non-generative and non-power conferring. This God could be the Prime Mover, a four-dimensionalist God, or the occasionalist God. Maybe the Prime Mover created the first physical event, and everything happens without God's intervention after that, or God created everything from the beginning (a kind of four-dimensionalism), or the occasionalist God creates the universe every single moment (detailed examples will be given below).

I will not argue that one of these metaphysical views is correct. Instead, I will argue that different metaphysical views avoid vicious infinite regresses and therefore, we should not suppose that only foundationalism does so. Further arguments are needed to show why foundationalism should be preferred.

Let's examine the aforementioned metaphysical views in more detail through some examples. Consider the case of a physical cause being the source of RCC. It may be asked why a statue is real and has a specific causal capacity (being solid, being white, etc.). Two possible answers are: because of its creator[15] (its cause) or its parts (its atoms, its grounds). Schaffer and Trogdon worry that if its grounds have further grounds ad infinitum, we face a vicious infinite regress. However, I believe that the vicious infinite regress can be avoided if the statue's creator is the source of RCC. The creator made it the case that the statue is real and has a specific causal capacity. We do not need to look at further causes in the past to explain the RCC of the statue. This can be the case even if the statue's grounds have further grounds ad infinitum.

[15] More precisely, the person who was creating the statue. At some points, I will talk as if an object is the cause of an effect, but this can be considered a shorthand for the event that caused the effect (that can also be an event). Events can be the relata of causal relations. This does not influence my argument. Following Kim (1984), I conceive an event as the exemplification of a property by an object at a time. If we talk about an event being the source of RCC, both events and objects can be conceived as fundamental ontological categories. Then, events may be mentioned to show how a vicious infinite regress of grounded objects or events can be avoided. For example, an event e_1 at time t_1 can cause an event e_2 at time t_2. E_2 consists of a grounded object O having a property P at time t_2. O's source of RCC is e_1.

I argue that a statue gets its reality from its creator—the maker of the statue. However, someone could find this example unconvincing[16] since, in the statue case, what the creator does is rearrange various components that are already real (the material from which the statue is made) into a new order. If so, is it not more reasonable to say that the statue gets its reality from its parts or material, which are already real, rather than from its creator or rearranger?

I could say that without facing any problem. Let's say that the statue is created at time t_2. It has certain parts at that time: particular objects. At time t_1, those objects interact in specific ways that cause the statue's creation at t_2. Those objects can be the source of RCC of the statue. Alternatively, those objects and their causal interactions at t_1 can be the source of RCC of the statue. The statue's source of RCC could be the event that consists of those objects and their causal interactions. It could be asked which parts are relevant. Especially in a metaphysical infinitist ontology. The elementary particles? Something even smaller that we do not know that it exists? I respond: the elementary particles could be a source, and their parts could be another source, and the parts of these parts can also be a source, and so on. I believe that an object can have many sources of RCC. I find this kind of overdetermination unproblematic, and I will say more about it later in subsection 5.4.

Otherwise, we could say that the statue's source of RCC is the event of the creator arranging and causally interacting with the objects that become the parts of the statue. This event, which is a candidate for being a source, includes all the objects that contributed to the statue's creation. It includes both the creator and the objects that ground the statue. The event mentioned above could also be the source of RCC without facing any problems.[17]

---

[16] Thanks to an anonymous reviewer for raising this objection.

[17] Objection of an anonymous reviewer: I do not think that either the maker of the statue or any other physical cause can create something real out of nothing or give reality to something that is not already real. Namely, if the statue maker had no real parts at her disposal, how could she make a statue? The author does acknowledge that God might be a source of reality, and that makes sense. Hence, given the previous, I doubt that anything other than God can play the role of a cause that is a source of reality in the way the author requires. Response: I think the statue's

Still, someone may argue that to really explain why the statue is real and has causal capacity, we should search for the desires and beliefs of the creator that caused him to make the statue, and we should also find out how these desires and beliefs were formulated, and so on ad infinitum, and this leads to a vicious infinite regress of causes.

At this point, two responses are available: (a) there is not actually a vicious infinite regress, and (b) a God or a first physical cause can be alternative sources of RCC. I will argue for (a) in the next subsection. The main claim is that there is an infinite regress of causes, but there is no transference of the same status ad infinitum. (b) can be sketched here: I believe that worries about vicious infinite regress of causes can be avoided if God is the source of RCC of the statue.[18] It may be the case that a Prime Mover

maker can give reality to something not already real: the grounded object called 'statue'. Before the creation of the statue, there were only its parts. After the arrangement of the parts, a new object is created. A grounded object. The statue. As I will explain later, a first physical cause, something that was not created, may also be an ultimate source or explanation of why something has RCC. All the objects may originate from this first cause. If we believe that only something uncreated can explain why things have RCC, either a God or a first physical cause can play this role.

[18] It could be claimed that this is another form of foundationalism: God is the foundation. This could be the case, especially if 'causation' and 'grounding' refer to the same building relation (see Bennett (2011b) for this view, even though she changed her mind later (Bennett 2017)). If someone wants to call this view foundationalism, despite the existence of infinite priority chains, I do not see any problem with that. My main claim here is not to defend infinitism but to claim that the postulation of ungrounded objects is not the only way to avoid a vicious infinite regress. That is why I restrict the term 'foundationalism' only to the view that there is a fundamental level of objects. In other words, I restrict "metaphysical foundationalism" to the view that all grounded entities are grounded in ungrounded entities. There is another way to express the aim of this paper: I argue that considerations concerning avoiding vicious infinite regresses show that the world has a foundation. However, they do not help us to decide what this foundation is. I do not argue for metaphysical infinitism or a specific metaphysical view. I do not argue that a first cause is explanatorily advantageous compared to other foundationalists' accounts. Instead, I argue that considerations about infinite regress can lead to different metaphysical views

created the universe, and as a result, the chains of causal dependence stop at some point in the past. The Prime Mover is supposed to have the power to create things ex nihilo. Unlike physical entities, a Primer Mover does not need to be transferred RCC from somewhere else to have RCC. Otherwise, it may be the case that occasionalism is true: every moment, God creates the universe and brings about the RCC of each object. This way, the vicious infinite regress is avoided too. Creation, in both cases, is a causal process.

If somebody is sceptical of the existence of a God, an initial/first physical cause can be considered as the source of RCC. It may be the case that the universe has a beginning. At that time, there was a first object that directly or indirectly caused the existence of every other object, and so, this object is the source of RCC of every other object. This object has always existed and has not received its RCC from anything else. As a result, there is not a vicious infinite regress: the chains of causal dependence stop at the beginning of the universe. Sometimes, I will mention God just for ease of exposition. It could be replaced by 'first physical cause'.

What Schaffer and Trogdon need for their argument to be successful is the truth of the claim that only through ungrounded objects we can avoid a vicious infinite regress. However, as the above examples show and I will argue below, there are other possible scenarios that avoid this problem, and it is not obvious that we should prefer the foundationalist's suggestion.

5.1. Generative, Power Conferring Accounts of Causation

I will talk here about the alternative sources of RCC that rely on a generative and power-conferring account of causation, and below, I will consider whether non-generative and non-power-conferring accounts of physical causation can be used to develop the foundationalist argument and respond to my objections. Let's further develop the abovementioned claims by remembering the foundationalist's worry. A foundationalist worries that if metaphysical infinitism was true, a priority chain would not have a source of RCC, and it would be a mystery why anything exists. They have the intuition that there must be a source of RCC, a ground that metaphysically

and postulating the existence of grounded objects is just one solution among others. More needs to be said to decide which view is the correct one.

explains why anything else exists or has causal capacity. Cameron (2010), Schaffer (2010) and Trogdon (2017) have argued that the whole cosmos is the source, and Cameron (2008a) has considered a plurality of atoms as the source. All of them searched for the source inside the priority chains.

But maybe the source can be outside the priority chain. To see this, let's consider again the causal inheritance premise that Trogdon accepted[19]. "The causal inheritance premise: necessarily, if A is nonfundamental and has causal capacity then A inherits its causal capacity from whatever fully grounds it" (Trogdon 2018, 192).

I do not think A can inherit its causal capacity only from whatever fully grounds it. Instead, I believe that the following conditional is true: if A is nonfundamental and has causal capacity, then A inherits its causal capacity from its cause or its full ground[20]. This conditional reveals another way that a nonfundamental entity can have a source of RCC. The cause of A (B) can be its source, and the explanation of why A is real and has a certain causal capacity can end there. What really matters is that an explanation ends. Whether it stops in an ungrounded physical object or a cause does not matter.[21] It is not advantageous to have one over the other.

Assume that infinite priority chains exist. Each grounded entity may inherit RCC from its cause. The source of RCC of an infinite chain can be its cause. The cause causes the existence of each object in the infinite chain and makes these objects have the causal capacities they do[22]. It explains the nature and causal capacities of all these objects because of that. There is no extra need for something inside the chain that is also the source of its RCC.

---

[19] My following remarks would be similar, even if I discussed Schaffer's version of this argument.

[20] This conditional implies different possibilities depending on whether we understand "or" as inclusive (overdetermination) or exclusive (no grounding relations, uncaused priority chains). These possibilities will be discussed below.

[21] If what is needed is an ultimate explanation, either an ungrounded ground or an uncaused cause, God or a first physical cause could be the uncaused cause and the alternative source of RCC. I will develop this claim below.

[22] It could be objected that the RCC of any item in the chain will be overdetermined, as it is produced both vertically (within the chain) and horizontally (by God). I reply to this objection on page 27.

To see that the source of RCC of a grounded object can be its cause, let's digress and examine fully pedestalled chains presented by Dixon (2016). These chains were presented to argue for a specific definition of well-foundedness[23], but they are also relevant to the current topic. The structure of these chains is the following, where orange arrows symbolise grounding (e.g., Y_2 grounds Y_1):

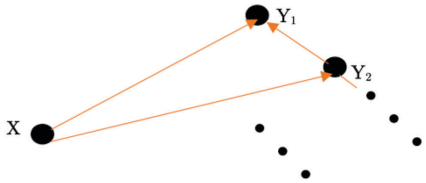

Figure 3: A Fully Pedestalled Chain

All the Ys and the X are facts. Even though this pedestalled chain contains a non-terminating grounding chain, each non-fundamental fact it includes is fully grounded by the fundamental fact X. Each Y_i is fully grounded by each Y_{i+1} and is also fully grounded by X. X is fundamental. Each fully pedestalled chain contains a non-terminating grounding chain. Still, they are well-founded because every non-fundamental fact is fully grounded by a fundamental fact. For Dixon (2016, 446), "the principle that best captures the claim that grounding is well-founded [is] [...] (FS) Every non-fundamental fact x is fully grounded by some fundamental facts Γ".

Now, instead of facts, imagine that X and all the Ys are objects[24] (X may be an ungrounded particle, and Y_1 may be an infinitely divisible

[23] They were used to argue against a set-theoretic notion of well-foundedness. This notion does not allow infinite priority chains to be well-founded because they do not terminate in fundamental entities. However, Dixon (2016) has argued that some infinite chains, such as fully pedestalled chains, are acceptable for the metaphysical foundationalist as every non-fundamental entity is fully grounded by fundamental entities, and there is still a source of reality for every non-fundamental entity.

[24] Remember that they can also be events. It does not influence my argument. Talking about events may be preferred because they are considered the relata of causal relations. I talk about objects just for ease of exposition.

particle). Y_1 is a composite object, Y_2 refers to the objects that ground Y_1, Y_3 refers to the objects that ground Y_2, and so on. Also, imagine that X is the cause of the non-terminating grounding chain. The arrows beginning from X represent a causal relation.[25] The other one still represents the grounding relation. I call what we imagine now a 'Fully Pedestalled Causal Chain 1' (FPCC1, see figure 4 where blue arrows symbolise causation[26]). X seems to be the source of RCC of all the Ys. Being is achieved in virtue of X. Every Y exists, is real, and has the causal capacity it does in virtue of X. This seems to be sufficient to metaphysically explain the reality and causal capacity of all the Ys. Their metaphysical explanation ends in the X.

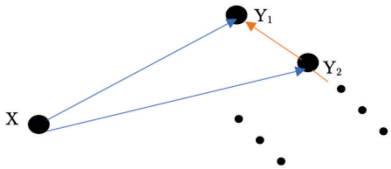

Figure 4: A Fully Pedestalled Causal Chain (FPCC1)

If we are sceptical of objects being the relata of causal relations, the X and the Ys can be understood as events. We can think of God, the creator of the universe, who caused everything else to exist either directly or indirectly. A particular event, God causing the existence of the first gunky object Y_1 (i.e., event x, God creating a gunky object at time t_1, causing event y_1, consisting of the physical object Y_1 being real at time t_2), is the

[25] An anonymous reviewer finds it unclear why X in Figure 4 is the cause of every individual item in the chain (one-to-infinity causation). It seems better to them if X is instead the cause of the chain as a whole (one-to-one causation).
I do not have any problem with this alternative account of causation. The source of RCC can still be X.

[26] In figure 4 and figure 5, I use singular variables to refer to grounds for convenience only. All grounds can be understood plurally. That is, each variable that refers to a ground can refer to multiple entities. Grounding may be a many-one relation: many entities ground one entity.

source of RCC of all physical objects and events. For example, Y_1 causing the existence of other gunky objects is happening because God is its source of RCC. Another example would be God directly causing the existence of several gunky objects, and the interaction between these objects causes the existence of other gunky objects.

These infinite chains seem well-founded in a sense close to what foundationalists are saying. Remember the definitions of well-foundedness we have seen already. For Schaffer, a priority chain is well-founded, iff it terminates in something fundamental. For Dixon (2016, 446), "the principle that best captures the claim that grounding is well-founded [is] [...] (FS) Every non-fundamental fact x is fully grounded by some fundamental facts Γ". Both Schaffer and Dixon want something that is the source of a priority chain. If we understand 'fundamental' as 'ungrounded', then not all kinds of fully pedestalled causal chains are well-founded as X may be a part of an infinite chain, and therefore, it will be grounded in something else (we will examine a chain like this one below). Still, these chains are well-founded in a close sense. They are well-founded* because they terminate in and are fully caused by their causes.[27] Their causes are their source. That is why they terminate there. The core motivation for endorsing foundationalism is having a source that provides an explanation that ends somewhere. However, we can have such a source even without foundationalism, and so it seems that we have lost the reason to advocate foundationalism.

It could be objected that the source of RCC of a grounded object cannot be its direct cause. Consider a grounded object, A, and its direct cause (B). It may be argued that B cannot be the source of RCC of A because B is real and has causal capacity in virtue of something else: its direct cause (C). C is also real

[27] Objection: The infinite chains are well-founded* in a sense different from the senses used by Schaffer and Trogdon. This makes all the discussion a verbal dispute! Reply: I do not think so. The dialectic does not go this way: we need well-foundedness; we do not have it, but at least we have well-foundedness*, and so, infinitism is good enough. Instead, the dialectic goes this way: Schaffer, Trogdon, and I believe that a grounded object must have a source of RCC, and I argue that foundationalism is not the only theory that provides such a source. As a result, considerations about vicious infinite regresses do not show that foundationalism is true.

Alternatively, we could say that infinite chains are not well-founded but are fully caused, and therefore, the lack of well-foundedness does not pose a problem.

and has causal capacity in virtue of something else, and so on ad infinitum. So, the explanation never stops, and there is a vicious infinite regress.

But I think that my view does not lead to a vicious regress. There is a difference between a grounding chain and a causal chain. In a grounding chain, there is transference of the same RCC all the time, but in a causal chain, there is a different transference every time (or most of the time).

For example, B can be the source of RCC of A, and the explanation ends there. If we ask why B has the causal capacity it does, then we ask for an explanation of a different causal capacity. While A may have causal powers P, B may have causal powers Q. So, we are not always asking for an explanation of the same causal capacity. Of course, we could ask why B has the causal capacity it does, but then we are not concerned with A anymore. We are not searching anymore for the source of causal capacity of A. We are searching for the source of causal capacity of B. For example, consider again the case of the statue and its creator: the statue and its creator have different causal powers.

We cannot say something similar about grounding. In the grounding case, the problem is that we constantly search for the source of the same causal capacity and never reach a point where this explanation stops. Consider the case of an object that has a certain mass in virtue of its parts that have the same mass, and these parts have this mass in virtue of their parts that have the same mass, and so on ad infinitum. The explanation of why the object has a certain mass always moves somewhere else.

Also, consider the case of determinable properties (e.g., being in pain) and determinate properties (e.g., being in a particular brain state). This is another case that shows the difference between a grounding chain and a causal chain. According to Shoemaker (2007) and Wilson (2011), the causal powers of each determinable are a non-empty proper subset of the causal powers of their determinates. This can be understood as the causal powers of the determinables being metaphysically explained by the causal powers of their determinates. If there is no fundamental level, the explanation of the same causal powers is always moved somewhere else.

For example, consider a property D that has causal powers P_1-P_5. Property D is grounded in property E that has causal powers P_1-P_{10}. Property E is grounded in property F that has causal powers P_1-P_{15}, and so on ad

infinitum. D's source of causal capacity always moves somewhere else: the metaphysical explanation of the causal powers of D always moves to another property that possesses, among others, these causal powers. The same causal powers are found in other properties ad infinitum. There is transference of the same causal powers ad infinitum.

Remember what Schaffer has considered as a vicious regress. For Schaffer (2016, 95), "a regress counts as vicious if and only if there is an endless chain of dependency with transference of the relevant status". We might be asked, 'Why does a grounded object A have this specific causal capacity?', and we would mention B, its cause, as the source. Then, if we were asked, 'But why does B have this specific causal capacity?', the causal capacity in question differs from A's causal capacity. Therefore, we are no longer talking about the transference of the same status. Therefore, an endless chain of causal dependencies is not vicious.

Someone may respond to me by saying that I was talking about the specific determinate causal powers (or the determinable property of having these specific causal powers) instead of the determinable property of having the capacity for causal activity (i.e., the determinable property of having causal powers), and the latter was the property that Trogdon was talking about. An object has this determinable property, iff it has determinate properties with specific, determinate causal powers. This determinable property reveals a way that there can be a transference of the same status from the cause to the effect. There can be an infinite chain of causal dependence in which there is always transference of the capacity for causal activity. A vicious infinite regress reappears.

However, I think we do not need to admit the existence of the determinable property of having the capacity for causal activity. It is true to say that an object has the capacity for causal activity iff it possesses properties with specific causal powers. An additional determinable property is not needed to explain why it is true or to make sense of our claims. There is no motivation for postulating the existence of an additional property. It is not the case that there are determinate properties with specific causal powers and a determinable property with the capacity for causal activity. There are only the former properties.

Even if we think that having causal capacity is a determinable property over and above the properties with specific causal powers and as a result, there is a transference of the same status from the cause to the effect, then it may be the case that the intuition behind the foundationalist argument can be used to argue for the existence of either a fundamental level or a Prime Mover. It may be argued that if A exists in virtue of B that exists in virtue of C and so on ad infinitum, then it is a wonder why A exists: the status of A is always transferred somewhere else. But if A is grounded in a fundamental entity, then it can be explained why A exists by avoiding a vicious infinite regress. Still, I wonder why we should not say instead that infinitism is true and a Prime Mover is the source of RCC of a grounded object. A vicious infinite regress is avoided this way too. It seems that the foundationalists' intuitions can be used to argue for this claim instead.

There is still a plausible objection to the foundationalist argument about infinite regress, even if my claims about transference of a different status in the case of causation are wrong. Vicious infinite regresses can be avoided by postulating grounded objects, a Prime Mover, or a first physical cause, and more needs to be said about which metaphysical view we should prefer. Considerations about infinite regress does not favour one of these metaphysical views.

5.2. Non-Generative, Non-Power Conferring Accounts of Physical Causation

So far, I assumed that a generative, power-conferring account of physical causation is true. However, metaphysical views about causation and existence, which Schaffer endorsed, may be used at this point to argue against my objection. Nonetheless, I will argue that these views fail to show that my objection does not work.

Schaffer (2016, 95-96) has allowed for limitlessly backwards causal and temporal sequences. Concerning causation, he has claimed that there is no transference of reality. The effect has intrinsic reality unto itself and is ontologically subsistent in its own right. No first cause is needed because of that.

> Within each distinct portion of reality, one must find an internal source of its reality (this is the required well-foundedness of grounding). But across distinct portions of reality, one is dealing

> with metaphysically independent tiles of the cosmic mosaic, and each tile is there from the start without needing a source (this is the permitted non-well-foundedness of causation) (Schaffer 2016, 96).

The quoted passage above claims that while limitlessly backwards causal sequences do not lead to vicious infinite regresses, infinite chains of grounding do so. It is not mentioned, though, why we should believe that this is the case. It is not explained why this contrast holds. Schaffer says that grounding must be well-founded to avoid vicious infinite regresses. But this does not seem to be the case.

It is unclear why all the objects of an infinite priority chain could not always exist without the need for a source. It is not obviously true that there must be a fundamental entity that generates the grounded entities. There is an alternative plausible story that I suggest in this section. Schaffer's story may seem more intuitive, but this does not make the alternative false (and this is what Schaffer needs for his argument to be sound). Schaffer did not give any argument against this alternative story.

Another way to express what has been said so far is that Schaffer has a specific foundational intuition in mind. At any moment, there is a fundamental level and entities that are generated by it. Without a fundamental level, how could all these entities be generated? The generator would move somewhere else ad infinitum. My answer: they could be generated by their (first) cause, or they could have always existed.[28] I do not see why Schaffer's generative model is the only possible solution.

Grounding can still be generative, in my view. It can still play the same role but in an overdetermining way. It can be the case that all entities of a priority chain always existed, and at any time, grounding overdetermines the causal capacity of entities. Alternatively, it can be the case that each

[28] It may be argued that if everything always existed, then foundationalism is true: everything is fundamental in a broader sense (X is fundamental iff X is ungrounded or always existed). As mentioned, I restrict 'foundationalism' to the view that all grounded entities are fully grounded in ungrounded entities. Even if other metaphysical views can be called 'foundationalist' in a broader sense, it does not influence my argument that vicious infinite regresses are not avoided only through ungrounded objects.

grounded object is real and has causal capacity because of its (first) cause and its ground. So, it does not need to have RCC in virtue of an ungrounded object. I will briefly discuss in subsection 5.4 whether this kind of overdetermination is problematic.

I think both caused and grounded entities need to have a source of RCC that does not receive its RCC from something else, or neither of them does. If both do, we may have two equally plausible solutions: a Prime Mover or a fundamental level. A Prime Mover may be the source of RCC of all objects. Otherwise, a fundamental level is this source: at any moment, certain fundamental entities are the source. They both can be the source, and foundationalists have to say more about why one option is better than the other.

A disanalogy between the grounding case and the causing case may be that while in the infinitist's account, a cause brings about the existence of grounded entities in a robust sense, in Schaffer's account, the cause does not bring about the existence of the effect in a similar robust sense (the effect was there all along). That is, causation is not generative (as I use the term). Infinitists may need a more robust account of causation than what Schaffer endorses.

Following Pearl (2000) and Spirtes et al. (2000), Schaffer (2016) has formalised causation through structural equation models that mention correlations to pick out the actual causal relations. Under certain conditions, one can infer causation from correlation.

> Structural equation models come with precise—and indeed freely downloadable—discovery algorithms that allow one, given certain plausible assumptions, to estimate causal structure from sufficiently rich correlational structure over three or more variables (Schaffer 2016, 60).

If you want to find out whether a specific type of causal relation holds, you can "input your data into TETRAD (or some other causal discovery algorithm), and receive a precise and empirically reliable estimate of direction and strength of causal influence" (Schaffer 2016, 60). If you want to find out whether a particular token causal relation occurs, you use the type-level picture you get from using a causal discovery algorithm, assign values to the relevant variables that refer to potential causes, and then look at what would happen to the variable that refers to an assumed effect if you re-

assign a particular variable while keeping the other variables fixed (distinct variables represent distinct features of the world). In other words, whether a token causal relation holds can be discovered through a test in terms of counterfactual covariation: wiggle the cause, and the effect wiggles.

A foundationalist may argue that for the cause to be the source of RCC of the effect, infinitists need a more robust account of causation than the structural equation models account[29] (e.g., a productive account of causation that identifies causation with the transference of energy from the cause to the effect[30]); they need a kind of causation that transfers RCC from the cause to the effect. This is the only way the cause can make the effect have RCC. But generative accounts of causation are false, and so, my objection is false too.

Nevertheless, even if we endorse the structural equation models account of physical causation, I still do not see why it is impossible that God produced everything from the beginning, and this is the source of RCC of everything. God causes events in a non-physical way. This alternative is still viable. Even if there are good reasons to conceive physical causation as non-generative and non-power conferring, they do not exclude the possibility of non-physical causation that is generative and power conferring. More generally, the proponent of this response would have to argue that the structural equation models account is an account of all causation, i.e., of the one and only causal relation. However, it is unclear how to argue for this.

---

[29] Objection by an anonymous reviewer: structural equation models are primarily *models*. As models, they are neutral on the underlying metaphysics of what is being modelled. Hence, they should be compatible with all kinds of views about causation, including generative ones.

Response: Here, I consider the possibility that a structural equation models account is presented as a deflationary account of causation that assumes there is nothing more to causation. If there is something more revealed by a power-conferring account of causation, this something more can be used to argue for the possibility of a cause as a source of RCC (as I have already done in the previous subsection).

[30] Schaffer (2016) has claimed that both causation and grounding are productive and generative, but he has meant this in a different sense from what the productive account of causation does. As we have seen above, causes and effects have always existed for Schaffer. A cause does not bring about the existence of an effect; there is no transference of reality from the cause to the effect.

Alternatively, if we have good reasons to think that the structural equation models account is an account of all causation, it could be claimed that all entities always existed, including the grounded entities (i.e., the universe is eternal). All times are equally real, and the universe has no beginning. There is no moment in which the universe was created either by the Big Bang or God. In this case, grounded entities do not have a source of RCC, and there is no reason to think they should. These entities always existed. There was no time that they were brought into existence, and causal capacity was transferred from something else to them. So, even in this case, worries about vicious infinite regresses do not motivate the truth of foundationalism.

Someone could object that a confusion between existential and temporal priority seems to be made. I adhere to the idea that if objects have always existed, they would not need a source for their reality or causal capacity. However, grounding theories assume that the groundee and the grounded can temporally coexist while the latter ontologically depends on the former.

I try to motivate a different metaphysical picture here. A grounded object can still do its metaphysical job. It can still transfer RCC to another grounded object even though this object already has that RCC. A source of RCC is not needed because there is no time that a grounded object did not exist, and somehow, it was brought into existence.

Otherwise, someone could claim they are unsure they see how eternal existence would remove the need for explanation. The grounding regress is synchronic-- it is (allegedly) mysterious how such an infinite priority chain could exist at any time, let alone every time.

However, I think that there is an available explanation. The eternal existence is the explanation. Grounded objects have the RCC they have because they always had it.

5.3. Causal Explanations, Metaphysical Explanations, and Sources of RCC

The foundationalist argument I want to undermine can be formulated as (a) or (b) below:

a) If there were infinite (backward) dependency chains, there would not be an explanation of why a given object exists and/or has

causal powers. But there must be such an explanation. Therefore, there are no infinite (backward) dependency chains.

b) If there were infinite (backward) dependency chains, there would not be a source of reality or causal powers for a given nonfundamental object. But there must be such a source. Therefore, there are no infinite (backward) dependency chains.

I deny the first premise of the argument (that is, the initial conditional statement in (a) and (b) respectively) by claiming that the explanation/source of reality and causal powers can be provided by the cause (proximal or distal) of the nonfundamental object.

A foundationalist may worry about this argument: it seems that my point against the first premise of the argument works just in case one understands 'explanation' as 'whatever explanation' and 'source of reality and causal powers' as 'whatever explains, in some sense of 'explaining', why a given object exists and has its causal powers'. If these liberal understandings of such terms are adopted, my argument against the first premise works: in fact, what causally explains why a given object exists and has its causal powers explains (in the causal sense of 'explaining') why the object exists and has its causal powers. But it may seem to a foundationalist that this is not the sense of 'explaining' foundationalists such as Schaffer, Cameron, and Trogdon have in mind. They are, in fact, thinking of metaphysical explanation or constitutive explanation: an explanation whose features are different from those of causal explanations (for instance, it is synchronic, it is such that the explanans necessitates the explanandum, it is true in virtue of the nature(s) of the things involved). Accordingly, what they have in mind when they speak about a 'source of reality and causal powers' is something that explains, in this metaphysical/constitutive/synchronic sense of 'explanation', why a given object exists and has its causal powers. But if one understands the first premise of the argument in these more exigent senses of 'explanation' and 'source of reality and causal powers', my argument does not work anymore: the vicious regress to which foundationalists are pointing is, in fact, precisely one concerning metaphysical/constitutive explanation, and it may seem to an objector that this regress cannot be stopped by pointing to another kind of explanation. In other words, the foundationalist may stress that by pointing to an external cause, one

explains how the object, with its causal powers, came into existence. However, this is not the request of explanation that the foundationalist wants to answer. What he wants to explain is, in fact, why the object exists and is what it is at any given time once it came into existence (synchronic/constitutive explanation).

I think that an infinitist can respond satisfactorily to this objection. Suppose that X, a grounded object, grounds Y, another grounded object. In this case, Y is metaphysically explained by X. This metaphysical explanation is satisfactory even though X is a grounded object because X already has RCC from its cause.

If we only focus on metaphysical explanations, there is a problem. The original problem of metaphysical infinitism was that infinitists were trying to metaphysically explain Y by mentioning only its ground. If we just mention X, it is a wonder why X can metaphysically explain Y satisfactorily. X has its RCC in virtue of another grounded object, and so on ad infinitum. The explanation never stops. But if we give a causal explanation of X, then it can be shown why a grounded object is sufficient for metaphysically explaining what it grounds.

The initial worry of the foundationalist was that X cannot metaphysically explain Y because X gets its RCC from something else that gets it from something else, and so on ad infinitum. But if X gets its RCC from its cause and the transference of RCC ends there (or the transference ends in a Prime Mover or a first physical cause), there is not any problem with a grounded object metaphysically explaining another grounded object.

The foundationalist worry may be that once X gets its RCC from its cause, something else must keep giving it its RCC or explain why X keeps having RCC. I do not see why this must be the case, and this worry brings us closer to the objection mentioned by Bliss (2013), which is that foundationalist arguments concerning vicious infinite regresses are circular. The search for a specific kind of metaphysical explanation is motivated by specific foundationalist intuitions, but we have these intuitions only if we are already persuaded by foundationalism.

The assumption that once X gets its RCC from its cause, it keeps having it seems very plausible. A reason must be given why we should not think this way.

I have decided to talk about arguments (a) and (b) because I think these are the best versions of foundationalist arguments concerning vicious infinite regresses because they avoid circularity. Still, these arguments also fail.

5.4. Objections and Replies

In what follows, I will mention some possible objections and my replies to them.

Objection: Earlier, when I imagined FPCC1, I imagined that X (e.g., an ungrounded particle) is not a part of an infinite chain. I do not consider a world in which some objects are parts of infinite chains while other objects are not problematic. If for any reason[31], someone does, the following case could be considered:

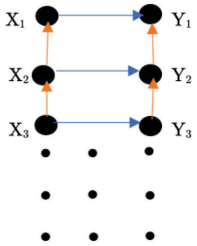

Figure 5: Fully Pedestalled Causal Chain 2 (FPCC2)

[31] Schaffer (2010) has claimed that compositional facts are not contingent. If priority monism or pluralism is true, it is true with metaphysical necessity. So, Schaffer believes that *"either it is metaphysically necessary for the cosmos to be a fundamental whole, or it is metaphysically necessary for the cosmos (if it has proper parts) to be derivative"* (Schaffer 2010, 56). A metaphysical infinitist could agree that compositional facts are not contingent and, contra Schaffer, argue that metaphysical infinitism is true with metaphysical necessity. So, it is metaphysically necessary for the cosmos to include only infinite chains.

Otherwise, she may think that compositional facts are contingent and still believe that there are only infinite chains in the actual world because of methodological reasons (a world in which there are only infinite chains is qualitatively simpler than a world in which there are both infinite and finite chains).

Reply: There are infinite Xs that form an infinite priority chain, and Ys do so too (e.g., X1 is the statue's creator, and Y_1 is the statue; both are infinitely divisible). As with the FPCC1, there is a source of RCC for every Y, even though the source is not inside the infinite chain of Ys. A difference is that not one X is the cause of all the Ys. Still, the metaphysical explanation for each Y ends somewhere (in its direct physical cause). If we worry about a vicious infinite regress of causes or seek an ultimate explanation, the source of each Y's RCC can be the universe's first physical cause.

Objection: The proposed view faces the problem of overdetermination. Xs determine Ys, and each Y, except Y_1, overdetermines another Y.

Reply: This is a big issue to discuss here, but if there is no problem with overdetermination (see Sider (2003) for reasons to believe so), these infinite chains seem unproblematic.

Trogdon (2018) has mentioned the denial of overdetermination as a part of his foundationalist argument. Schaffer has not mentioned having a similar problem with overdetermination. This shows that whether my reply is successful depends partially on whether the denial of overdetermination is reasonable and a part of the foundationalist argument. If it does, I may need to say more to persuade the foundationalist that overdetermination is unproblematic after all.

Objection: A revised foundationalist argument may still be possible. It can be stated by arguing that certain metaphysical positions I used here are false. The foundationalist has to deny the truth of occasionalism, the existence of a Prime Mover, that our universe is eternal, and that there was a first physical cause. If these entities exist or these theories are true, they can provide a non-foundationalist source of RCC. So, a foundationalist should show that these entities do not exist and that these theories are false to motivate their view.

Reply: In this way of thinking, this paper contributed to making explicit how a foundationalist should argue for their position. So, there is still a philosophical value to it. It revealed that the dialectic should change. It is not the case that foundationalism is the only way to avoid a vicious infinite regress. Foundationalism is the only way to avoid a vicious infinite regress, given that certain metaphysical views are false.

However, I am sceptical of the plausibility of this foundationalist reply. I am unaware of any argument for the definite falsity of these views. Usually, they are rejected because of methodological principles, bringing us closer to Cameron's (2008b) argument. It seems that what does the philosophical work in order to argue for metaphysical foundationalism is methodological principles (simplicity and theoretical utility: see Cameron, 2008b). There is no need to mention considerations about infinite vicious regresses. These considerations do not work. Whether Cameron's argument is sound is controversial and outside of the scope of this paper. I just mention it here to show how the dialectic may need to change and how foundationalists may have to argue for their view.

Objection: It is hard to see how God could produce a grounded entity except by producing its grounds. For example, no one, not even God, can create a house directly. One can create a house only by creating its parts in certain relations to each other. If this is right, then God can create any given member of an infinite priority chain only by creating its grounding member(s), and God creates them only by creating its grounding member(s), and so on. This looks similar to the original regress and is also vicious. The 'by' locution keeps passing God's task down to the next level so that if this were to continue infinitely, he would never have anything to create.

Reply: According to metaphysical foundationalism, the whole universe or some particles are created by God (literally or figuratively), and the rest follows. If that is the case, it is unclear to me why God could not create both directly. It is unclear why creating the one directly makes it impossible to also create the other directly.

I claim that God can, at the same time, directly create both the house and the parts, but someone may have doubts that this claim is coherent. Namely, it may be unclear to someone that God could directly create something that is an ontological free lunch that comes ontologically 'for free' once its ground is in place. It seems that on such a conception, not even God can create a house (assuming that houses are derivative) unless God first creates its parts.[32]

I do not see any independent reason why we should find impossible the creation of both the house and its parts by God. God could create an object

[32] Thanks to an anonymous reviewer for raising this objection.

that is grounded in another object, and it does not follow that the created object is not an ontological free lunch. Believing differently just reveals a different metaphysical picture that I argued against in this paper. I gave many examples to motivate this picture. More needs to be said about why this metaphysical picture is impossible.

Despite being created by God, the house is grounded in its parts. That is why it is an ontological free lunch. Its continuing existence is grounded in its parts. Still, its source is its cause (God).

This takes us back to footnote 25. It may be that God is the cause of every individual item in an infinite chain of grounding (one-to-infinity causation), or it may be that God is instead the cause of that chain as a whole (one-to-one causation). I can use both accounts to argue for God as the source of RCC of a grounded object belonging to that infinite chain. I do not see any reason to prefer one of these two accounts.

If what I have said so far is correct, our intuitions about infinite regresses cannot be used to argue for foundationalism because the source of RCC of a grounded object can be its (first) physical cause or a God. Fully pedestalled causal chains are possible and counterexamples to the claim that the source of RCC must be inside a priority chain. Priority chains can terminate, even if infinitism is true. They terminate in their direct or indirect cause. What really encourages our foundational intuitions about infinite regresses is the search for a source of RCC, and this source can be found even if infinitism is true. Therefore, a foundationalist must abandon their view if they do not have any other reason to endorse this theory.

6. Conclusion

Schaffer and Trogdon have argued that metaphysical foundationalism is the only way to avoid vicious infinite regresses that emerge as a consequence of infinite grounding chains, and therefore, we have a good reason to endorse it. However, I argued that alternative sources of reality or causal capacity that avoid such vicious infinite regresses are possible, even if metaphysical infinitism is true. These can be the direct physical cause of a grounded entity, the first physical cause, or God. Alternatively, it can be the case that there is no need for a source of RCC because our universe is eternal.

Therefore, foundationalism cannot be motivated by the argument concerning vicious infinite regresses.

Acknowledgements

I would like to thank Aaron Cotnoir, Derek Ball, Franz Berto, Tuomas Tahko, and an anonymous referee for Organon F for helpful comments on earlier versions of this paper. Thanks also to seminar and conference audiences at the University of Amsterdam, University of Salzburg, and University of St Andrews for helpful discussions and comments on previous versions of this paper.

References

Audi, Paul. 2012. "Grounding: Toward a Theory of the in-Virtue-of Relation." *The Journal of Philosophy*, 109(12): 685–711. https://doi.org/10.5840/jphil20121091232

Barnes, Elizabeth. 2018. "Symmetric Dependence." In *Reality and its Structure: Essays in Fundamentality*, edited by Ricki Bliss and Graham Priest, 50–69. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198755630.001.0001

Bennett, Karen. 2011a. "By Our Bootstraps." *Philosophical Perspectives*, 25(1): 27–41. https://doi.org/10.1111/j.1520-8583.2011.00207.x

Bennett, Karen. 2011b. "Construction Area (No Hard Hat Required)." *Philosophical Studies*, 154 (1): 79–104. https://doi.org/10.1007/s11098-011-9703-8

Bennett, Karen. 2017. *Making Things Up*. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780199682683.001.0001

Bird, Alexander. 2007. *Nature's Metaphysics: Laws and Properties*. Oxford: Clarendon Press. https://doi.org/10.1093/acprof:oso/9780199227013.001.0001

Bliss, Ricki. 2013. "Viciousness and the Structure of Reality." *Philosophical Studies*, 166 (2): 399–418. https://doi.org/10.1007/s11098-012-0043-0

Bliss, Ricki. 2014. "Viciousness and Circles of Ground." *Metaphilosophy*, 45(2): 245–56. https://doi.org/10.1111/meta.12072

Bohn, Einar Dunguer. 2018. Indefinitely Descending Ground. In *Reality and Its Structure: Essays in Fundamentality*, edited by Ricki Bliss and Graham Priest, 167–81. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198755630.001.0001

Cameron, Ross P. 2008a. "Truthmakers and Ontological Commitment: Or How to Deal with Complex Objects and Mathematical Ontology without Getting into

Trouble." *Philosophical Studies*, 140(1): 1–18. https://doi.org/10.1007/s11098-008-9223-3

Cameron, Ross P. 2008b. "Turtles All the Way Down: Regress, Priority and Fundamentality." *The Philosophical Quarterly*, 58(230): 1–14. https://doi.org/10.1111/j.1467-9213.2007.509.x

Cameron, Ross P. 2010. "From Humean Truthmaker Theory to Priority Monism." *Noûs*, 44(1): 178–98. https://doi.org/10.1111/j.1468-0068.2009.00736.x

Castaneda, Hector Neri. 1984. "Causes, Causity, and Energy." *Midwest Studies in Philosophy*, 9(1): 17–27. https://doi.org/10.1111/j.1475-4975.1984.tb00050.x

Correia, Fabrice. 2010. "Grounding and Truth-Functions." *Logique et Analyse*, 53(211): 251–79.

Dasgupta, Shamik. 2014. "On the Plurality of Grounds." *Philosopher's Imprint*, 14(20).

Dixon, T. Scott. 2016. "What is the Well-Foundedness of Grounding?." *Mind*, 125(498): 439–68. https://doi.org/10.1093/mind/fzv112

Fair, David. 1979. "Causation and the Flow of Energy." *Erkenntnis*, 14(3): 219–50. https://doi.org/10.1007/BF00174894

Fine, Kit. 2001. "The Question of Realism." *Philosophers' Imprint*, 1(1): 1–30.

Fine, Kit. 2012. "Guide to Ground." In *Metaphysical Grounding: Understanding the Structure of Reality*, edited by Fabrice Correia and Benjamin Schnieder, 37–80. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139149136

Heil, John. 2003. *From an Ontological Point of View*. Oxford: Clarendon Press. https://doi.org/10.1093/0199259747.001.0001

Heil, John. 2012. *The Universe as We Find It*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof%3Aoso/9780199596201.001.0001

Jenkins, Carrie S. 2011. "Is Metaphysical Dependence Irreflexive?" *The Monist*, 94(2): 267–76. https://doi.org/10.5840/monist201194213

Kim, Jaegwon. 1984. "Epiphenomenal and Supervenient Causation." *Midwest Studies in Philosophy*, 9: 257–70. https://doi.org/10.1111/j.1475-4975.1984.tb00063.x

Kim, Jaegwon. 2005. *Physicalism or Something Near Enough*. Princeton, New Jersey: Princeton University Press. https://doi.org/10.1515/9781400840847

Morganti, Matteo. 2009. "Ontological Priority, Fundamentality and Monism." *Dialectica*, 63(3): 271–88. https://doi.org/10.1111/j.1746-8361.2009.01197.x

Morganti, Matteo. 2014. "Metaphysical Infinitism and the Regress of Being." *Metaphilosophy*, 45(2): 232–44. https://doi.org/10.1111/meta.12080

Morganti, Matteo. 2015. "Dependence, Justification and Explanation: Must Reality be Well-Founded?" *Erkenntnis*, 80(3): 555-72. https://doi.org/10.1007/s10670-014-9655-4

Pearl, Judea. 2000. *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press.

Rabin, Gabriel Oak, and Rabern, Brian. 2016. "Well Founding Grounding Grounding." *Journal of Philosophical Logic*, 45(4): 349–79. https://doi.org/10.1007/s10992-015-9376-4

Rosen, Gideon. 2010. "Metaphysical Dependence: Grounding and Reduction." In *Modality: Metaphysics, Logic, and Epistemology*, edited by Bob Hale and Aviv Hoffmann, 109–36. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199565818.003.0007

Schaffer, Jonathan. 2009. "On What Grounds What." In *Metametaphysics: New Essays on the Foundations of Ontology*, edited by David Chalmers, David Manley, and Ryan Wasserman, 347–83. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780199546046.003.0012

Schaffer, Jonathan. 2010. "Monism: The Priority of the Whole." *Philosophical Review*, 119(1): 31–76. https://doi.org/10.1215/00318108-2009-025

Schaffer, Jonathan. 2012. Grounding, Transitivity and Contrastivity. In *Metaphysical Grounding: Understanding the Structure of Reality*, edited by Fabrice Correia and Benjamin Schnieder, 122–39. Cambridge: Cambridge University Press. http://doi.org/10.1017/CBO9781139149136.005

Schaffer, Jonathan. 2016. "Grounding in the Image of Causation." *Philosophical Studies*, 173(1): 49–100. https://doi.org/10.1007/s11098-014-0438-1

Shoemaker, Sydney. 1980. "Causality and Properties." In *Time and Cause*, edited by Peter Van Inwagen, 109–35. Dordrecht: Reidel Publishing Co. https://doi.org/10.1007/978-94-017-3528-5_7

Shoemaker, Sydney. 1998. "Causal and Metaphysical Necessity." *Pacific Philosophical Quarterly*, 79: 59–77. https://doi.org/10.1111/1468-0114.00050

Shoemaker, Sydney. 2007. *Physical Realization*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199214396.001.0001

Sider, Theodore. 2003. "What's so Bad about Overdetermination?" *Philosophy and Phenomenological Research*, 67(3): 719–26. https://doi.org/10.1111/j.1933-1592.2003.tb00321.x

Spirtes, Peter, and Scheines, Richard. 2004. "Causal Inference of Ambiguous Manipulations." *Philosophy of Science*, 71: 833–845. https://doi.org/10.1086/425058

Tahko, T. E. (Fall 2018 Edition). Fundamentality. In *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta. Retrieved from https://plato.stanford.edu/archives/fall2018/entries/fundamentality/

Trogdon, Kelly. 2017. "Priority Monism." *Philosophy Compass*, 12(11): e12458. https://doi.org/10.1111/phc3.12458

Trogdon, Kelly. 2018. "Inheritance Arguments for Fundamentality. In *Reality and its Structure: Essays in Fundamentality*, edited by Ricki Bliss and Graham

Priest, 182–98. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198755630.003.0010

Wilson, Jessica. 2011. "Non-Reductive Realization and the Powers-Based Subset Strategy." *The Monist*, 94(1): 121–54. https://doi.org/10.5840/monist20119417

Organon F 32 (3) 2025: 300–322 ISSN 2585-7150 (online)
https://doi.org/10.31577/orgf.2025.32302

RESEARCH ARTICLE

Scientific Realism/Anti-Realism Debate: Roy Bhaskar's Position

Maryam Poostforush* – Mostafa Taqavi**

Received: 11 June 2024 / Revised: 31 March 2025 / Accepted: 6 April 2025

Abstract: The debate between scientific realism and anti-realism has long revolved around scientific theories. Realists contend that scientific theories are true or approximately true. In contrast, anti-realists posit that scientific theories are shaped as mere instruments that help to predict and categorize observable phenomena. As such, anti-realists find no truth value in scientific theories and only accept their empirical adequacy. Roy Bhaskar, as a critical realist, believes in the existence of the unobservable entities and considers them knowable. Furthermore, because he considers knowledge fallible, he maintains that theories arising from social activities may or may not be true. The purpose of this article is to clarify the position that Bhaskar takes in the debate between scientific realism and anti-realism. Before addressing this central purpose, the article first tries to find out whether Bhaskar is indeed a realist in three metaphysical, epistemic

* Sharif University of Technology
https://orcid.org/0000-0002-1085-9957
Department of Philosophy of Science, Sharif University of Technology, Azadi Stree, Tehran, P.C.: 1458889694, Iran
mpoostforush@gmail.com

** Sharif University of Technology (corresponding author)
https://orcid.org/0000-0002-7755-1375
Department of Philosophy of Science, Sharif University of Technology, Azadi Stree, Tehran, P.C.: 1458889694, Iran
m_taqavi@sharif.edu

and semantic dimensions. The study argues that his alternative position on the truth/falsity of theories would detach him from the anti-realist camp, and he is not an anti-realist in terms of any of the three dimensions mentioned above. Of course, Bhaskar draws what could be called a "delicate border" between his realist approach and that of realists who attribute the approximate truth of a theory to its success.

Keywords: Scientific realism; anti-realism; Roy Bhaskar; theory; truth; experiment.

1. Introduction

Belief in unobservable entities and a mind independent world have long been the central issues addressed in debates between scientific realism and anti-realism. Such debates have focused on various topics. Yet, specifically speaking, realism refers to any philosophical stance that believes in the reality of things (Manee 2018, 32; Pölzler 2018). Scientific realism takes a realist position in relation to what best human theories can describe. Realism suggests that scientific theories are true or approximately true and their theoretical terms have putative factual reference (Bueno 2015, 153). On the contrary, anti-realists (e.g., instrumentalists, constructive empiricists) do not believe in the association between scientific theories and truth and view theories as mere tools that can predict or categorize observable phenomena (Mizrahi 2020a, 38).

Although this conflict began many centuries ago, it remains one of the topical issues in philosophy. Defending their specific schools of thought, philosophers have speculated on the debate and in some cases have offered new readings (Rowbottom 2019). As such, some thinkers have concentrated on the dichotomy, trying to provide understandable and clear explications of both realist and anti-realist assumptions as a way of exploring the roots of the opposition (Rouse 2018; Mulder 2012). Proposing the idea of relative realism, some philosophers have sought to resolve the conflict by expressing some premises that could be acknowledged by both realists and anti-realists (Mizrahi 2020b).

The present study seeks to clarify the position of Roy Bhaskar, who identifies himself as a critical realist, in this debate. On the one hand, he believes in a mind independent world, finding it possible to gain knowledge about the unobservable aspects of the world. On the other hand, he posits that cognition is fallible and that truth cannot be known with certainty. This study primarily elaborates on the debate between realists and anti-realists. Next, to answer the central question raised, the study investigates Bhaskar's realism in terms of its metaphysical, epistemic and semantic. We argue that his approach, in all three dimensions, stands in opposition to anti-realism. Of course, in the debate, he draws a delicate border between his approach and that of realists. Paragraph: use this for the first paragraph in a section, or to continue after an extract.

2. The Debate between Realism and Anti-Realism

Are theories true or approximately true descriptions of the world and its theoretical entities have putative factual reference, or are they mere instruments that facilitate scientific goals (e.g., prediction)? The way a philosopher approaches this question would decide his/her position on the debate between realism and anti-realism (Bonilla 2019, 3962).

As Stathis Psillos point out scientific realism has three stances that each of the anti-realist schools has reacted to one of these three stances (Psillos 1999, xix–xx):

(1) The metaphysical or ontological stance: the world has a definite and mind independent natural-kind structure. It is in contrast to the idealist and phenomenalist approaches.
(2) The semantic stance takes scientific theories at face-value, seeing them as truth-conditioned descriptions of their intended domain, both observable and unobservable. This stance differentiates scientific realism from eliminative instrumentalist and reductive empiricist accounts.
(3) The epistemic stance asserts that unobservable entities are knowable and regards mature and predictively successful scientific theories as well-confirmed and approximately true of the world. In

contrast, epistemic antirealist, like constructivist empiricists, reject the possibility of recognizing the unobservable entities and suffice with the empirical adequacy of theories.

To better understand the nature of the debate, one should first realize how realists and anti-realists defend their claims. For this reason, their most central arguments are explored below.

2.1. The "No-Miracle Argument" (NMA)

Science is successful and the scientific community has always relied on successful theories to pursue its goals. Realists link this success with truth, and in response to the central question in the debate, they attribute scientific success to its truthfulness (Lewis 2001). The NMA, advanced by Hilary Putnam, represents the most important argument in sustaining realism (Putnam 1975). In his argument, Putnam uses the inference to the best explanation (IBE). Among a set of existing hypotheses, the IBE tries to select the hypothesis that can provide the best explanation based on available evidence (Harman 1965, 89).

Putnam argues that realist hypotheses "that terms in mature scientific theories typically true that the theories accepted in a mature science are typically approximately true, [and] that the same term can refer to the same thing even when it occurs in different theories" are the only scientific explanation of success of science (Putnam 1975, 73). On this account, he maintains that realism is the only philosophical approach that does not view scientific success as a "miracle" (ibid.). The reason for this is that believing in any explanation other than the truth of scientific theories would attribute scientific success to miraculous happenings. Putnam's argument as be formulated as follows (Mizrahi 2020c, 52–53):

- Premise 1: Our best theories can make successful predictions and explanations;
- Premise 2: If the theories are not true or approximately true, their success would be like a miracle;
- Premise 3: The best explanation for this success is realism, which finds scientific theories true or approximately true;
- Conclusion: Our best theories are true.

Apart from anti-realists, some realists have criticized Putnam's idea, finding faults with his NMA (Macarthur 2020). Some even called it fallacious (Hoyningen-Huene 2018) or did not generally view the IBE as a relevant contribution to the debate (Frost-Arnold 2010). Some other critics did not it prioritize success as an adequate condition for the truth of a theory, suggesting that even if success could serve as a factor of truth, it would not be possible to conclude that most successful theories are true (Wray 2013, 1720). Furthermore, epistemic success would not guarantee the ontological necessity of the unobservable entities assumed to exist in a theory (Wray 2018). Despite all these criticisms, the NMA remains the most important argument that realists rely on.

2.2. Pessimistic Meta-Induction (PMI)

In response to advocates of the NMA, Laudan, as an instrumentalist, provides a history of the philosophy of science and successful theories that were ruled out over time. He mentions many previously proposed theories that were successful and central terms in scientific theories genuinely refer, according to the claims of theorists. However, such theories were proven to be wrong, and today no one could find them "true" or claim that a term such as "aether" is a referring term (Laudan 1981, 35). To further support his argument, Laudan offers a list of theories although he claims the list could still include more instances (ibid., 33):

- The effluvial theory of static electricity
- The phlogiston theory of chemistry
- The caloric theory of heat
- Theory of the electromagnetic aether
- The optical aether
- The theory of circular inertia.

Advancing an inductive argument, he concludes that some theories are successful and are claimed to refer to real entities by the terms they use and are thus called "true", but they may be ruled out in the future. As such, their terms would be non-referring. Considering this argument, the success of a theory does not necessarily establish its truth (Laudan 1981, 32). As Laudan further explains, if there were no such a thing as atom,

atomic theories would not be (approximately) true. If genes did not exist, genetic theories would not be (approximately) true, no matter how valid such theories may appeared (ibid., 33). Questioning the NMA, Laudan demonstrates that the empirical success of a theory could not substantiate its truth.

There are various versions of the PMI, although they have been criticized as well (Park 2014). Criticizing the PMI, realists emphasize that the false parts of a theory would not have any role in its success, and that the success of a theory would not justify its entirety (Kitcher 1993, 142). Meanwhile, the aspects of a theory that have a central function in its success will be preserved in scientific image (Psillos 1999, 108, 139).

2.3. Arguments of Constructive Empiricism

As an epistemic anti-realist, van Fraassen takes an agnostic stance toward the existence of unobservable entities, viewing experience as the only source of human knowledge. As such, he contends that we cannot talk about what moves beyond empirical evidence. He advances some arguments against realism, which are briefly reviewed below.

The Positive Argument of Constructive Empiricism

Introducing constructive empiricism, van Fraassen explains that the purpose of science, contrary to the claim of realists, is not to achieve truth but empirical adequacy. He does not regard scientific activity as an attempt to discover the truth of unobservable entities, but rather he suggests that science seeks to construct models that are empirically adequate. On this account, accepting a theory means that the theory in question provides an accurate description of observable phenomena (van Fraassen 1980, 5, 12). Otherwise put, van Fraassen does not strongly associate the success of science with its truth. Meanwhile, he does not believe that success arises from miracles, suggesting that the success of science stems from its empirical adequacy (van Fraassen 1980, 39–40). Relying on such stances, van Fraassen calls his stance on science and scientific activity a positive argument for constructive empiricism because "it makes better sense of science, and of scientific activity, than realism does and does so without inflationary metaphysics" (van Fraassen 1980, 73).

Underdetermination of Theory by Evidence (UTE)

The UTE denotes that in most cases several theories may be compatible with the available evidence, although the evidence is not substantial enough to determine a theory in the face of its competing theories (Manee 2018, 36). There is debate over whether van Fraassen relies on the UTE. However, many philosophers, such as Psillos, believe that van Fraassen uses the UTE to repudiate truth in scientific theories (Psillos 1999, 156). Regardless of these debates, the UTE is indeed one of the arguments that anti-realists resort to, criticizing the NMA.

The NMA claims that only one theory can provide the best explanation, and only realism is capable of explaining the success of theories by finding them true. The UTE, however, claims that evidence cannot determine the truth of a theory. The reason for this is that in most cases there may be at least two theories that explain the same phenomenon, and existing empirical evidence equally supports the two theories. As such, it would not be possible to select a theory among the competing ones, and there would be no reason to find a theory to be true. Relying on the UTE, anti-realists show that there is no reason to believe in the truth of one specific theory (Manee 2018, 36; Psillos 1999, 156–58).

2.4. The Anti-Induction (AI) Argument

As mentioned earlier, the most important argument of realists is the NMA. Recently, however, a new argument called the "anti-induction" has been advanced, which claims to be more effective than the NMA because it remains valid in the face of the problem of theory rejection due to scientific development over time. The anti-induction argument rests on the history of science and postulates that because previous theories were wrong, most successful theories in the present are approximately true. The reason for this idea is that scientists hope to reach correct and successful theories by constantly ruling out false theories. Anti-realism uses the uniformity principle in the PMI and emphasizes previously wrong theories to repudiate the truth of presently existing theories. In contrast, the anti-induction argument relies on the disuniformity principle, underscoring the approximate truth of current theories (Park 2018, 330–32).

Defending his argument, Park explains to critics that most of them are anti-inductionists when they resort to trial and error in science or everyday life or when they assert that failure is the mother of success. Such anti-inductionist tendencies rest on the disuniformity principle (Park 2018, 340). Therefore, Park contends that scientific theories could provide approximately true descriptions of the world. In his argument, Park refers to the disuniformity principle and a number of famous proverbs. Yet, what can be criticized about his argument is that it centrally relies on the disuniformity principle, without providing any justification for using the principle. Meanwhile, he does not explain whether the disuniformity principle is a necessary or contingent issue? Another question left answered is when it is justifiable to use induction in science and when it is justifiable to draw on anti-induction.

3. Aspects of Roy Bhaskar's Realism

With more clarity about the various dimensions of the realism / anti-realism debate, one can now observe that realists try to advance arguments that scientific theories are true or approximately true. In contrast, anti-realists' counter-argument tries to pursue an instrumentalist approach to scientific theories. From the perspective of Bhaskar's realism, science deals with ontological questions about what things exist and how they behave. In the transcendental arguments Bhaskar advances, he depicts a world that involves causal structures and enduring mechanisms; he believes the purpose of science is to gain knowledge about these mechanisms. This section of the present study seeks to answer the question in what sense Bhaskar's critical realism is a realist approach. Otherwise put, is Bhaskar's philosophy of science a realist approach from metaphysical, epistemic, and sematic stances?

Before engaging with Bhaskar's position, it is essential to note that in the philosophy of science literature, standard realism is typically defined by Paul Psillos' tripartite division into ontological, epistemic, and semantic dimensions. The longstanding debate between realism and antirealism also primarily unfolds within this framework, with antirealists generally dissenting from realists on one or more of these dimensions. Bhaskar, who establishes a depth ontology and diverges from standard realism, naturally

arrives at a distinct epistemology that distinguishes him from conventional realists. A key point of contention between Bhaskar and standard realism lies in his rejection of the correspondence theory of truth. Since antirealists don't subscribe to the truth of theory, the question arises: Does Bhaskar's denial place him among antirealists?

To address this, it is necessary to analyze Bhaskar's stance across the aforementioned three dimensions, delineate his boundaries with standard realism and antirealism, and clarify the realist aspects of his position as he claims. In other words, while a precise mapping of Bhaskar's realist discourse–due to the complexity of his ontological foundations and their tight interconnection with concepts such as causal mechanisms, natural necessity, and normic laws–requires independent research, the first step involves situating him within the traditional debate. This allows demonstrating how he attempts to transcend the limitations of standard realism by proposing a dynamic, dialectical version of the school without veering into antirealism's camp.

3.1. The Metaphysical Stance

Inspired by Kant, Bhaskar bases his critical realism on transcendental questions. Considering the fact that scientific activity cannot be denied, he tries to figure out the specifications of the world in which science is possible. As such, he posits that experience in science is as much dependent on sense-perception as it is on experimental activity. Following this premise, he suggests that experience in science is conducted by humans as both perceivers and causal agents, and thus he transcendentally analyzes sense-perception and experimentation (Bhaskar 2008, 13, 21).

Analyzing sense-perception

Contrary to empiricism, which postulates that humans are only certain of the things they can experience through sensation, Bhaskar tries to show that "There could be a world of events without experiences" (Bhaskar 2008, 22). Exploring human perception of objects, he posits that the intelligibility of sense-perception presupposes the intransitivity of the object perceived. For instance, the fall of an apple, as an event experienced by a human being, does not exclusively depend on his/her perception, because the event exists independently from experience.

This intransitivity and the independence of objects are also presupposed in scientific change and scientific education. This understanding assumes that there is something distinct and independent of the teacher and students that is described differently by different people (Bhaskar 2008, 21). In other words, objects are independent of human knowledge and as humans change their approach to them, objects will not change. Therefore, whether humans gain knowledge about objects or not, objects simply remain what they are. For instance, current flow in metal conductors is an event that does not depend on human knowledge and metals were conductors even before the law of conductivity was discovered by human agents (Bhaskar 2008, 11–12).

Given these premises, Bhaskar concludes that events are categorically independent of experiences, and that there are events in the world that have not been experienced. The history of science witnesses that at any given moment in time there could be unimaginable events that may be theoretically or empirically acquired as knowledge (Bhaskar 2008, 22). Furthermore, Bhaskar contends that if a scientist believes in an ontological realm, it would be possible for him/her to rationally criticize epistemic claims. That is to say, scientists, in their scientific explorations, seek to gain knowledge of intransitive objects that are independent of human knowledge, and this pursuit makes it possible for them to criticize reservoirs of knowledge gained (Bhaskar 2008, 13, 33).

Analyzing experimental activity

Bhaskar criticizes philosophers who would view a constant conjunction of events as the necessary condition for the law of causality. As such, he clarifies that experiment is necessary because causal laws are different from sequences of events (Benton & Craib 2010, 125). Proposing the concept of open and closed systems, he recognizes the world as an open system in which there are diverse mechanisms simultaneously operating, and this situation could give rise to disturbances in mechanisms and finally prevent them from being actualized (Bhaskar 2008, 114).

An open system, then, rarely involves sequences of events. However, humans as causal agents working in scientific laboratories, as closed systems, try to dissect a given mechanism and produce constant conjunctions. Bhaskar focused on this feature of experimental activity as a basis for his investigation. He posits that casual laws are not the same as the sequences

of events produced by human agents in experiment. He explains that only if an "ontological distinction" exists between causal laws and the sequences of events beyond the closed experimental system, the scientist can discover causal structures by artificially generating regularities (Bhaskar 2008, 23). As such, Bhaskar's transcendental understanding of sense-perception and experimental activity can be recapitulated as follows:

(1) Causal laws are independent of the sequences of events, and events are independent of experiences;
(2) There is a world existing independent of the human mind, while human beings make efforts to gain knowledge about the mechanisms in the world;
(3) Knowledge is a mode of achievement that is acquired by human involvement in nature.

As a result, Bhaskar emphasizes the above transcendental stance and distinguishes transitive objects (that are variable and human-dependent) from intransitive objects (that are independent of human knowledge and agency). As such, he underscores the existence of a real world beyond the scope of human mentality. This position is opposed to that of idealists who deny the existence of the world independent of the human mind.

Additionally, Bhaskar's analysis of sense-perception and experimental activity suggests some ideas other than the distinction between causal laws and the sequences of events (which are themselves independent of human experience). More specifically, he uses the distinction as evidence to propose that reality is composed of three levels: empirical, actual, and real (Bhaskar 2008, 47, 2009, 19). The empirical level refers to the level that is observed and experienced by us. The actual level of reality is occurring regardless of being observed or experienced. Finally, the real level involves causal structures and fundamental mechanisms that generate events. According to Bhaskar, the world is composed of complex things that have causal powers and tendencies, and operating a mechanism in its characteristic way is nothing but a thing goes on acting in a certain way. Bhaskar finds such mechanisms real and regards them as intransitive objects that are embedded in reality and generate events. As such, he expresses his opposition to anti-realism, which denies the existence of the unobservable level and finds the

world confined to the observable level. Therefore, the ontological aspect of Bhaskar's philosophy is evidently in line with realism.

3.2. The Epistemic Stance

It was mentioned earlier that epistemic realists believe it is possible to gain knowledge of the unobservable dimensions of the world, while suggesting that highly successful scientific theories are approximately true. The question here is whether Bhaskar, who is a metaphysical realist, subscribes to epistemic realism or not. Following a transcendental approach, Bhaskar describes the world as being inherently structured, differentiated, and stratified. On this basis, he views scientific activity as an ongoing motion that makes every effort at any given moment in time to permeate into deeper strata of reality, providing insights into generative mechanisms. Clarifying that the purpose of science is to discover generative mechanisms of nature (Bhaskar 2008, 4), Bhaskar both emphasizes the possibility of knowing the world and describes science as an activity for unearthing reality.

He explains that scientists do not exclusively try to accumulate constant conjunctions, but they also try to discover causal powers of things (Bhaskar 2008, 205), and such causal powers or generative mechanisms are what explain the behavior of the things. For instance, thing A shows a particular behavior because it has the power B. As such, science seeks to figure out what kinds[1] of thing exist, what powers such things have, and how they behave (ibid., 165-169). Contrary to transcendental idealism, which posits mechanisms are human subjective constructs, Bhaskar underscores the "real" nature of mechanisms and contends that they are not artificial constructs. He further states that science tries to explain phenomena by referring to such mechanisms (ibid., 37, 157).

Because in Bhaskar's view reality is composed of empirical, actual, and real levels, he does not find science confined to the empirical level of observation and clarifies that scientific activity is governed by experimental intervention that tries to discover mechanisms at the real level. Given this premise, he contends that observation is often accompanied with action. Pointing out human intervention in nature and experiments, he asserts that if there were no real mechanisms, they would not reveal their effects in

experiments and science would be meaningless and merely limited to observation (Benton & Craib 2010, 125).

Bhaskar further acknowledges that notion of causal powers help us confirm the existence of entities that all we knows is their powers (Bhaskar 2008, 171). The entities that can only be known, through the detection of their effects, not shown to exist (as in the detection of radio-active materials by a Geiger counter, of electricity by an electroscope, of a magnetic field by a compass needle (ibid., 171, 177).

These explanations clarify that Bhaskar's realism depicts a world independent of the human mind and finds it possible to gain knowledge about the observable and unobservable dimensions of the world. This stance taken by Bhaskar is opposite to the views of epistemic anti-realists (e.g., van Fraassen) who advocate an agnostic approach to the existence of unobservable entities and believe knowledge could be only acquired from observable dimensions. Yet, does Bhaskar, like other realists, maintain that the best scientific theories are approximately true? The answer to this question is provided in the next section, which addresses the semantic stance of Bhaskar's philosophy of science.

3.3. The Semantic Stance

A semantic realist would claim that the terms used in scientific propositions should be interpreted literally. Such a person would also believe that scientific propositions are structured in a way that they can be found true or false depending on their correspondence or lack of correspondence with the world. In order to find out whether Bhaskar advocated semantic realism and answer the question remaining from the epistemic stance, one should elucidate Bhaskar's process of theorizing in his philosophy of science.

"Theory" in Bhaskar's Philosophy

The logic of scientific discovery in Bhaskar's philosophy primarily rests on regularities and semi-regularities that are explained by referring to generative mechanisms. As such, first plausible ideas about hypothetical mechanisms are constructed and then the ideas are subjected to theoretical criticism and empirical examination (Bhaskar 2008, 157). Bhaskar places mechanisms at the core of theory, emphasizing that the rationality of a scientific theory is only supported when its objects exist in reality (Bhaskar 2008, 36-

37). He suggests that gaining knowledge about mechanisms depends on a combination of perceptual, technical and intellectual skills (ibid., 37). He believes that scientists, in producing plausible ideas, use creative models, metaphors, and analogies. In fact, he sees a theory as the product of a creative process in which an unfamiliar thing is understood through analogy with the familiar (Baert 2005, 94). For instance, the movement of atoms may be similitude with the motion of billiard balls, the blood circulation system may be likened to a hydraulic model, or the biochemical structure of genes may be metaphorically described as linguistic codes (Bhaskar 2008, 12-13).

Although Bhaskar regards theories as way of referring to hypothesized inner structure of the world (Bhaskar 2008, 149), he believes scientific theorizing concerned with mechanisms and the world's causal structures is fallible. As such, he states that the claims of scientific theories about reality are not necessarily true, and they may be wrong (Baert 2005, 93). The reason for this situation is that scientists rely on their antecedent knowledge, including pre-established theories, models, tools, facts and methods to produce new knowledge. Because any of these items may undergo change and correction, new knowledge itself may be transformed (Bhaskar 2008, 11–12). Bhaskar, then, is opposed to classical empiricism, which suggested that a theory could be reducible to experience and could remain invariant. Bhaskar, however, maintains that a theory is neither reducible to experience nor can it remain unchanged. A theory is a product generated through social activity and is prone to change like any other social product (Bhaskar 2008, 178). As a result, he contends that it is always possible for a description of mechanisms to be wrong; for instance, our explanation of "why water must boil when heated" may be resting on a false assumption of this mechanism (ibid., 198).

Because Bhaskar views the world as an "open" system, he does not believe a theory's power of prediction could serve as an accurate test of its validity. More specifically, in an open system, the operation of conflicting mechanisms may prevent a given mechanism from being actualized or may lead to the generating of counter-instances. As a result, he does not find it a necessary or a sufficient condition for a theory to be compatible with facts,

emphasizing that only in a closed system a theory can be fairly examined (ibid., 128, 151).

The above observations clarify why Bhaskar not proving any criterion for the truth of a theory. However, it must be noted that the lack of a criterion for the truth of a theory does not mean that Bhaskar denied the possibility of reaching true theories. As he states, science seeks to provide a true explanation of reality. Although a theory arising from scientific work beyond their observational propositions always remains uncertain, there is also the possibility of a true account of reality. Bhaskar's stratified description of the world and his conviction that scientific theories are social products reveal why he always took a critical approach to human knowledge of reality and saw scientists' explorations potentially fallible attempts. In Bhaskar's opinion, it is always possible that there are more fundamental mechanisms in deeper strata of reality for a given phenomenon we are describing and explaining that remain undiscovered to us.

Yet, despite all these issues, the answer to the question whether Bhaskar is a semantic anti-realist is negative. More specifically, he is neither a reductionist who would reduce the meaning of unobservable entities in scientific theories to that of observable entities, nor is he an eliminative instrumentalist who would deny any meaning of the conceptual parts of a theory and consider them merely useful instrument (Psillos 1999, 10–11). Similarly, Bhaskar is not among thinkers who would consider only the observable part of the theory to have truth-value. As a result, he cannot be regarded as a semantic anti-realist, although he draws a delicate border between his philosophy and that of semantic realists. He is silent about the truth and falsity of scientific theories, providing no criterion for deciding truth/falsity. Of course, he does not find it impossible for a scientific theory to be true, thus distinguishing his approach from that of thinkers such as Laudan.

Bhaskar believes that because humans are capable of manipulating nature, they can produce a closed experimental system through intervention and inquiry how mechanisms hypothesized in theories work (Bhaskar 2008, 232). As he explains, "theory without experiment is empty. Experiment without theory is blind" (ibid., 182). Therefore, discovering whether the theoretical entities in a theory have a real referent or not would demand experimentation and action with reality. As such, mere philosophical

arguments cannot provide valid judgments. Of course, Bhaskar's emphasis on experimental activities does not denote that such activities can certainly determine the truth or falsity of a theory. Furthermore, he points out that experiment requires functional equipment, and in some cases a century may pass by until the instruments for testing a theory can be manufactured in practice (ibid., 183). Even then, experiments may reveal that some terms used in a theory have a real referent and some fail to do so. Otherwise put, Bhaskar leaves the truth and falsity claims of the theory to the valley of scientific activity and experimental scholars, although he does not provide a philosophical criterion for truth and falsity.

Given the discussions above, the answer to the remaining question about the epistemic aspect of Bhaskar's philosophy seems to be obvious. Does he believe that the most successful scientific theories are (approximately) true? As explained earlier, contrary to epistemic realists, he does not attribute the approximate truth of a theory to its accurate predictions and success. However, this stance should not imply that Bhaskar repudiates the possibility of truth in a theory. In contrast to van Fraassen, Bhaskar does not maintain that acceptance of the theory is the only guarantee of the empirical adequacy of the theory. He clearly defends the possibility of gaining knowledge about the unobservable dimensions of the world, suggesting that such knowledge could be even true. Given this stance, he obviously distinguishes his approach from that of epistemic anti-realists.

4. Bhaskar's Position in the Debate between Realism and Anti-Realism

In this section, the central question of the study is raised once again: Are scientific theories true or approximately true descriptions of the world or are merely useful instruments? In response to this question, anti-realists have taken different stances. Some basically rule out the existence of theoretically assumed unobservable entities, and some pursue an agnostic approach or deny the possibility of acquiring any knowledge of such entities.

Meanwhile, anti-realists generally agree that theories can be exclusively evaluated in terms of their (non)usefulness, and not in terms of their truth

or falsity. Even anti-realists (e.g., van Fraassen who is recognized as a realist in terms of ontology and semantic), contend that only the observable parts of a theory can be subject to truth-value. Realists, on the other side of the debate, suggest that a theory's (approximate) truth is decided by its success.

Although Bhaskar does not propose any ideas regarding the truth of theories, the present study, relying on the discussions mentioned above, emphasizes the argument that Bhaskar's position in the debate does not advocate anti-realism. In the metaphysical aspect, Bhaskar uses his transcendental analysis to describe a stratified world involving causal powers and generative mechanisms. Real structures and mechanisms are embedded in reality and they can be discovered through active engagement with nature, creative procedures, and experimentation.

This ontology of a stratified world is also reflected in the epistemic aspect of Bhaskar's philosophy, which posits that the purpose of science is to explain events and phenomena by referring to fundamental mechanisms. Although from Bhaskar's perspective exploring deeper strata to discover causal structures depends on the scientific theorizing and speculation and is influenced by social factors, his specific emphasis on intransitive dimension and the independent effects of causal powers salvage his approach from the abyss of relativism. Accordingly, he makes it possible to capture causal powers and generative mechanisms under experimental activities. In other words, Bhaskar both underscores the possibility of gaining knowledge about the unobservable dimensions of the world and finds it possible to achieve a true account of reality. Therefore, one can argue that, in his approach, a theory can be true or false and its assumed entities may be referring. As such, he distances himself from semantic anti-realists, although he does not provide any criterion for the truth/falsity of scientific theories.

Finally, Bhaskar's position in the debate seems to be evidently distinct from anti-realist positions. He defends a realistic approach to science in various ways and considers a true account of reality is possible. However, he draws a delicate border between his approach and that of other realist approaches. Contrary to the conventions of both sides of the debate, which try to support their camps by relying on philosophical and historical arguments, Bhaskar leaves unanswered the central question in the debate

concerning the truth/falsity of theories. Meanwhile, he is cognizant of the importance of experimental activity in uncovering the accuracy of theoretical entities and in correcting, improving, modifying, and advancing scientific theories.

5. Conclusion: The Way that Bhaskar Puts in Front of Us

Standard realists believe that the aim of scientific activity is to achieve the true or approximately true theories. In this approach, the theoretical entities at the heart of scientific theory refer to the real, and they imply statements that are true or approximately true. When talking about truth, standard realists normally consider the correspondence theory of truth. To compare this approach to that of Bhaskar, one has to first answer this question: What is the aim of scientific activity from the perspective of Bhaskar, and how does he respond to idea of "truth" or "approximate truth" in scientific theories?

According to Bhaskar, scientific activity are meant to discover generative mechanisms existing in reality. At first glance, Bhaskar's approach to the aim of science may not seem different from that of realists, and there may be some mere terminological differences. However, if put his understanding of the aim of science in the broader body of his philosophy, we could observe an important gap between his approach and that of realists. Bhaskar states that scientific activity seek to discover generative mechanisms and causal structures, although he does not express any specific ideas regarding the truth/falsity of theories. More specifically, he does not seem to be concerned with this issue. Bhaskar does not offer any criteria for revealing the truth/falsity of a scientific theory, while his philosophy does not address the problem of truth/falsity. Yet, why does Bhaskar not specify the criteria for evaluating the truth/falsity of theories? The reason for this is that he views a scientific theory as a transitive product dealing with intransitive objects. As Bhaskar observes, the material cause or the content that a scientific theory is constructed to address shapes the context in which scientific activity is conducted. As such, the content may be affected by metaphors, presumptions, values, and other factors that may have a contextual role in scientific activity. Therefore, a scientific theory is composed

of transitive objects and is a variable construct, and thus finding the correspondence between a theory and intransitive reality does not represent a strategic concern to Bhaskar.

If the correspondence between a theory and intransitive reality is not a strategic concern, then what is it Bhaskar finds important in scientific theories and theorizing? In Bhaskar's approach, the significance of a theory lies in the guidelines that it suggests for creating artificial conditions shaping experiments. Yet again, what is the purpose of creating artificial conditions in scientific activity? As the earlier review of Bhaskar's arguments revealed, he arrived at the conclusion that reality involves various causal mechanisms that work simultaneously. To discover such mechanisms through scientific activity, they must be analyzed under artificial conditions so that the effect of one specific mechanism can be observed as a regularity. Of course, it would be necessary to create such artificial conditions, because if mechanisms work at the same time they could disturb each other and their effects as regularities cannot be discovered.

Under such circumstances, a scientist would require theories to be able to create artificial conditions. A theory and its presumably existing entities direct the scientist toward implementing a particular experimental layout. In fact, the idea of scientific activity from Bhaskar's viewpoint suggests that the scientist relies on existing empirical evidence and materials obtained through scientific activity to innovatively construct a theory. Such a theory, like any other theory in science, involves some theoretical entities. In the next step, the scientist tries to draw inspirations from the theoretical content and the entities contained in it, designing experiments to find some mechanisms. Hoping to find potentially causal mechanisms, the scientist tries to explain previously observed phenomena. In Bhaskar's philosophy of science, contrary to what is said by the standard realists, a theory is not supposed to be true or probably true in order to be a theory, rather the theory's essence lies in its capability of being a source of inspiration for new experiments for probable finding of new mechanisms.

Bhaskar's philosophy of science takes into account a theory and its presumably existing entities, but it does not concern itself with whether the theory in question is true or not. The major issue is whether existing facilities and technologies, following theoretical inspirations, can construct

experimental conditions that help scientists discover causal mechanisms. Theory is transitive while causal mechanisms are intransitive. Bhaskar emphasizes the possibility that a theory may properly lead us to a causal mechanism.

There is a contrast that can better distinguish the views of standard realists from those of Bhaskar. Realists have advanced such philosophical arguments as the NMA and AI to demonstrate why scientific theories are true or approximately true. Yet, Bhaskar's philosophy and his realism does not basically require such philosophical attempts. Why would Bhaskar seek to defend philosophical arguments (e.g., the NMA and AI), while considering theories to be social and human products transformed by contextual factors and historical determinants? For instance, the falling of an apple from a tree is an event caused by a causal mechanism (or some mechanisms) that can be explained by Newton's classical mechanics or Einstein's general relativity, which are two remarkably different theories. This is why Bhaskar does not find it practical to focus on the truth of a theory or on the reality of the entities it assumes to exist. Instead, he tries to explain how a theory, through the inspirations it provides, could direct the scientist toward discovering a causal mechanism, which is the main aim of scientific activity.

Now, it would be important to mention a feature of Bhaskar's philosophy of science: He does not offer any guarantee that the regularity observed under experimental conditions would necessarily implies a causal mechanism. He points out that such observations may implies a causal mechanism. More specifically, he only mentions the possibility of discovering a causal mechanism. If from Bhaskar's perspective one of the aspects of progress in scientific activity lies in discovering an increasing number of causal mechanisms, then such progress would be a contingent matter. As such, following Bhaskar's explication, scientific activity may not necessarily lead to progress in science, it may only lead to progress. Bhaskar, however, does not determine the criteria for measuring such progress. Therefore, Bhaskar, with the methodological code he proposes (by invitation to experimentation), only promises the possibility of progress in discovery of more causal mechanisms, but not anything beyond that. Moreover, he does not, or better said cannot, suggest any metrics for measuring this contingent progress.

It was mentioned earlier that Bhaskar believes the aim of scientific activity is to discover causal mechanism embedded in things. Given the discussions about Bhaskar's views in this study, one could re-state his aim of scientific activity. He suggests that the aim of science is "practical achievement" that can be realized through experimental conditions. In such a case, however, some causal mechanisms are eliminated, which is a condition that makes it possible to observe a regularity that may be generated by a given mechanism. This practical achievement, however, differs from what realists prioritize as theoretical aim in scientific activity: the truth or approximate truth of theories. Of course, this contrast should not imply that realists do not seek any mode of practical achievement arising from scientific activity, but rather it is meant to highlight the difference stances taken in Bhaskar's view and that of realists.

The PMI argument could challenge the views of advocates of standard realism. Yet, given the explication of Bhaskar's philosophy provided in this study, his approach cannot be challenged by the PMI argument. The reason for this is that Bhaskar sees a theory as a human/social product that, affected by transitive objects, undergoes transformations throughout the history of science. As such, finding a correspondence between a theory and reality, or defining criteria for evaluating such a correspondence is not part of Bhaskar's philosophical agenda. This should not imply, however, that theory does not have any character in his approach. In fact, theory, with its experimental suggestions, serves as channel that could lead a scientist to causal mechanisms.

In Bhaskar's philosophy of science, theories always remain uncertain. Similarly, causal mechanisms discovered under experimental conditions are uncertain. What is definitely certain in Bhaskar's philosophy of science is the possibility of discovering causal mechanisms, and even if some experiments may go wrong, future ones can hopefully accomplish scientists' goals. In other words, what is not uncertain in Bhaskar's philosophy of science is possibility of achieving progress in the aims decided for science. Of course, he does not define any criteria that could help to measure scientific progress, but he suggests methodological code that can make progress possible in practice.

Acknowledgements

We would like to express our gratitude to Dr. Yassaman Hoshyar who studied the paper and shared their constructive ideas with us.

References

Baert, Patrick. 2005. *Philosophy of the Social Sciences: Towards Pragmatism*. Cambridge: Polity Press. https://doi.org/10.1177/0001699305056562

Benton, Ted, and Ian Craib. 2010. *Philosophy of Social Science: The Philosophical Foundations of Social Thought*. New York: Palgrave Macmillan.

Bhaskar, Roy. 2008. *A Realist Theory of Science*. New York: Routledge. https://doi.org/10.4324/9780203090732

Bhaskar, Roy. 2009. *Scientific Realism and Human Emancipation*. New York: Routledge. https://doi.org/10.4324/9780203879849

Bonilla, Jesús P. Zamora. 2019. "Realism versus Anti-Realism: Philosophical Problem or Scientific Concern?" *Synthese*, 196(10): 3961–77. https://doi.org/10.1007/s11229-015-0988-6

Bueno, Otávio. 2015. "Realism and anti-realism about science: a pyrrhonian stance." *International Journal for the Study of Skepticism*, 5(2): 145–67. https://doi.org/10.1163/22105700-04031176

Frost-Arnold, Greg. 2010. "The No-Miracles Argument for Realism: Inference to an Unacceptable Explanation." *Philosophy of Science*, 77(1): 35–58. https://doi.org/10.1086/650207

Harman, Gilbert H. 1965. "The Inference to the Best Explanation." *The philosophical review*, 74(1): 88–95. https://doi.org/10.2307/2183532

Hoyningen-Huene, Paul. 2018. "Are There Good Arguments against Scientific Realism?" In *Philosophy of Science*, 3–22. Cham: Springer. https://doi.org/10.1007/978-3-319-72577-2_1

Kitcher, Philip. 1993. *The Advancement of Science: Science without Legend, Objectivity without Illusions*. Oxford: Oxford University Press.

Laudan, Larry. 1981. "A Confutation of Convergent Realism." *Philosophy of Science*, 48(1): 19–49. https://doi.org/10.1086/288975

Lewis, Peter J. 2001. "Why the Pessimistic Induction is a Fallacy." *Synthese*, 129(3): 371–80. https://doi.org/10.1023/A:1013139410613

Macarthur, David. 2020. "Exploding the Realism-Antirealism Debate: Putnam contra Putnam." *The Monist*, 103(4): 370–80. https://doi.org/10.1093/monist/onaa010

Manee, Samal HR. 2018. "The 'Truth'Between Realism and Anti-Realism." *International Journal of Philosophy*, 6(2): 32–39. https://doi.org/10.11648/j.ijp.20180602.13

Mizrahi, Moti. 2020a. "Key Positions in the Contemporary Scientific Realism/Antirealism Debate." In *The Relativity of Theory*, 35–49. Cham: Springer. https://doi.org/10.1007/978-3-030-58047-6_3

Mizrahi, Moti. 2020b. "Relative Realism: The Best of Both Worlds." In *The Relativity of Theory*, 109–62. Cham: Springer. https://doi.org/10.1007/978-3-030-58047-6_6

Mizrahi, Moti. 2020c. "Key Arguments for Scientific Realism." In *The Relativity of Theory*, 51–77. Cham: Springer. https://doi.org/10.1007/978-3-030-58047-6_5

Mulder, Jesse M. 2012. "What Generates the Realism/Anti-Realism Dichotomy?" *Philosophica*, 84: 53–84. https://doi.org/10.21825/philosophica.82156

Park, Seungbae. 2014. "A Pessimistic Induction against Scientific Antirealism." *Organon F*, 21(1): 3–21.

Park, Seungbae. 2018. "The Anti-Induction for Scientific Realism." *Grazer Philosophische Studien* 95(3): 329–42. https://doi.org/10.1163/18756735-000044

Pölzler, Thomas. 2018. "Can the Empirical Sciences Contribute to the Moral Realism/Anti-Realism Debate?" *Synthese* 195(11): 4907–30. https://doi.org/10.1007/s11229-017-1434-8

Psillos, Stathis. 1999. *Scientific Realism: How Science Tracks Truth*. New York: Routledge. https://doi.org/10.4324/9780203979648

Putnam, Hilary. 1975. *Philosophical Papers/1 Mathematics, Matter and Method*. Ca,bridge: Cambridge University Press. https://doi.org/10.1017/cbo9780511625268

Rouse, Joseph. 2018. "Beyond Realism and Antirealism–At Last?" *Spontaneous Generations: A Journal for the History and Philosophy of Science*, 9(1): 46–51. https://doi.org/10.4245/sponge.v9i1.26979

Rowbottom, Darrell P. 2019. *The Instrument of Science: Scientific Anti-Realism Revitalised*. New York: Routledge. https://doi.org/10.4324/9780429022517

Van Fraassen, Bas C. 1980. *The Scientific Image*. Oxford: Oxford University Press.

Wray, K. Brad. 2013. "Success and Truth in the Realism/Anti-Realism Debate." *Synthese* 190(9): 1719–29. https://doi.org/10.1007/s11229-011-9931-7

Wray, K. Brad. 2018. *Resisting Scientific Realism*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108231633

Organon F 32 (3) 2025: 323–347 ISSN 2585-7150 (online)
https://doi.org/10.31577/orgf.2025.32303

RESEARCH ARTICLE

Brunero's Non-Normative Disjunctivism and Means-Ends Reasoning

Byeong D. Lee*

Received: 26 June 2024 / Accepted: 6 April 2025

Abstract: On Brunero's non-normative disjunctivism, agents exhibiting means-ends incoherence are irrational because they are guaranteed to have at least one attitude that fails to achieve its constitutive aim, and such an aim failure is not a failure of reason. This paper primarily aims to show that this account fails to adequately explain the irrationality of means-ends incoherence. More specifically, I argue that agents exhibiting this incoherence can be justifiably criticized for irrationality, and they cannot ward off such criticism by claiming that there is no reason to be means-ends coherent. Furthermore, I argue that Brunero's three objections to the strong normativity of means-ends coherence are not successful.

Keywords: means-ends incoherence; means-ends reasoning; constitutivism; non-normative disjunctivism; Brunero.

1. Introductive Remarks

Means-ends incoherence is widely regarded as a form of irrationality. Consider, for example, a case in which an agent, S, intends to visit Venice,

* Sungkyunkwan University
https://orcid.org/0000-0001-7080-6702
Department of Philosophy, Sungkyunkwan University, 25-2, Sungkyunkwan-ro, Jongno-gu, Seoul, Republic of Korea
bydlee@skku.edu

believes that buying a ticket is necessary for doing so, but fails to intend to buy one. If S holds such a combination of attitudes, he is means-ends incoherent. Why then does rationality require us to avoid such incoherence?

In his 2020 book, *Instrumental Rationality: The Normativity of Means-Ends Coherence*, John Brunero defends a *non-normative disjunctivism* about means-ends incoherence. According to this view, agents exhibiting means-ends incoherence are guaranteed to have at least one attitude that fails to achieve its constitutive aim: either they have an intention that does not result in the intended action, or they hold a belief that is false. This view has three origins. To begin with, Brunero borrows the idea from David Velleman (2000) and others that belief constitutively aims at truth. He also borrows (and modifies) the idea from Michael Bratman (2009a; 2009b) that intention constitutively aims at controlled action. Moreover, he borrows the 'disjunctivism' part of his approach from Niko Kolodny (2008, esp. 368). One important thing to note about Brunero's non-normative disjunctivism is that the failure involved in means-ends incoherence is to be explained in terms of a constitutive aim failure, and such an aim failure is not a failure of reason. According to Brunero, the aims of belief and intention do not generate normative reasons for achieving them. Instead, these constitutive aims are to be understood in terms of their roles within our psychological apparatus. A normative reason for (or against) a response, such as having an attitude or performing an action, is roughly a consideration that counts in favor of (or against) the response.[1] Hereafter, by 'reasons' I mean normative reasons.

This paper primarily aims to show that Brunero's non-normativist approach fails to adequately explain the irrationality of means-ends incoherence. Additionally, it seeks to refute his three objections to the strong

[1] Normative reasons should be distinguished from motivating and explanatory reasons. Motivating reasons are considerations for which, or in light of which, one makes a response, such as having an attitude or performing an action. Explanatory reasons are considerations that explain such a response without necessarily justifying it or being the reasons that motivated it. But motivating reasons for a response can be understood as a subset of explanatory reasons because a reason that motivates the response can always explain it. These distinctions are standard. See Alvarez (2016) for an overview.

normativity of means-ends coherence – the view that one should always be means-ends coherent. This paper is structured as follows. Section 2 argues that Brunero's account fails to adequately explain the irrationality of means-ends incoherence. More specifically, it argues that agents exhibiting this incoherence can be justifiably criticized for irrationality, and they cannot ward off such criticism by claiming that there is no reason to be means-ends coherent. Section 3 argues against Brunero's claim that rational requirements are not normative. Finally, Section 4 addresses Brunero's additional arguments for his non-normativist view. He presents three significant objections to the strong normativity of means-ends coherence, but I argue that these objections are unsuccessful.

2. Brunero's Non-Normativist Approach

On Brunero's non-normativist account, agents exhibiting means-ends incoherence are irrational because they are guaranteed to have at least one attitude that fails to achieve its constitutive aim. However, Brunero does not regard such an aim failure as a failure of reason. He interprets the constitutive aims of belief and intention in terms of their 'job descriptions' – their roles within our psychological apparatus. If this is correct, these aims do not generate reasons for achieving them. Let me explain this view further.

As mentioned in section 1, Brunero borrows the idea from Velleman that belief constitutively aims at truth. According to Velleman (2000, esp. 246), what distinguishes belief from other cognitive attitudes is that belief aims at truth. And he provides a functional interpretation of this aim. On this interpretation, to say that belief constitutively aims at truth is to say that truth is the constitutive function of belief. The proper function of the liver is to filter impurities and toxins from the bloodstream, and having this function is part of what it is for something to be a liver. In a similar vein, on this interpretation, the proper function of belief is to regard something as true only if it really is, and having this function is part of what it is for something to be a belief.[2] Brunero accepts this functional interpretation of

[2] For a more detailed discussion of the functional interpretation about the aim of belief, see Côté-Bouchard (2016, 3187–89).

the aim of belief. So, on his view, the constitutive aim of belief has nothing to do with *reasons*. Instead, the constitutive aim of belief is its 'job description' – the role of belief within one's psychological apparatus (see Brunero 2020, 177).

Brunero also borrows (and modifies) the idea from Bratman that intention constitutively aims at controlled action. According to Bratman (2009a, 2009b), just as truth is a constitutive aim of belief, coordinated and effective control of action is a constitutive aim of intention. Brunero agrees with Bratman's general strategy of appealing to the constitutive aim of intention as part of an explanation of intention-belief consistency and means-ends coherence. But he does not agree with a particular way in which Bratman carries out this strategy. Instead, he defends a weaker view that the constitutive aim of intention is effective control of action by jettisoning 'coordination'.[3] He also holds that the constitutive aim of intention is its 'job description' – the role of intention within one's psychological apparatus (see Brunero 2020, 185). Accordingly, the constitutive aim of intention also has nothing to do with reasons.[4]

Brunero does not deny that drawing attention to someone's means-ends incoherence is likely to create pressure for them to adjust their attitudes in order to resolve this incoherence. However, he attributes this pressure to the roles of our beliefs and intentions in our mental lives. He writes:

> We can hold that an intention, unlike a wish or a desire, *just is* the kind of mental state that generates pressure in the direction

[3] On Bratman's view, there is an important disanalogy between the aims of belief and intention. When we assert that belief aims at truth, we imply that a specific belief 'p' aims at getting things right with respect to 'p'. Notably, the aim of a specific belief does not make reference to one's other beliefs. In contrast, a specific intention to do X aims at effective control of doing X in coordination with one's other intentions. So, unlike belief, the aim of a specific intention makes reference to one's other intentions. Brunero argues that his weaker view has the advantage of preserving the parallel with belief, as the constitutive aim of a belief does not make reference to one's other beliefs.

[4] I accept neither Brunero's view that belief constitutively aims at truth, nor his view that intention aims at controlled action. But it is beyond the scope of this paper to refute these views.

> of being means-ends coherent. Part of what it is for you to *intend* an end, as opposed to merely desire it or hope that it obtains, is for you to feel pressure to intend believed necessary means when it's pointed out to you that you fail to do so.…We don't need to rely upon an agent's beliefs about normative reasons to explain pressure toward compliance. It can instead be explained by looking at the very attitudes governed by means-ends coherence, intentions and beliefs, and the roles they play within our mental lives (Brunero 2020, 208–209; original italics).

So, according to Brunero, we would feel pressure to revise our attitudes to resolve means-ends incoherence, mainly because an intention is the kind of mental state that generates pressure in the direction of being means-ends coherent; nevertheless, there is no reason to be means-ends coherent. Therefore, on his view, the rational requirement of means-ends coherence is only apparently normative. Against this view, however, I will argue in the remainder of this section that agents exhibiting means-ends incoherence can be justifiably criticized for irrationality, and they cannot ward off such criticism by claiming that there is no reason to be means-ends coherent.

Before presenting detailed arguments, five key points need to be clarified.

First, our concept of rationality is closely tied to rational criticism. For instance, if it is rational for you to hold a belief, you should not be criticized for holding that belief. Conversely, if it is irrational for you to hold a belief, you can be criticized for it. In this context, it is important to recognize that attributing irrationality is a form of criticism. On this point, Benjamin Kiesewetter and Errol Lord write, respectively, as follows:

> [O]rdinary attributions of irrationality are commonly understood as criticism. Moreover, the criticism involved seems to be personal criticism: when agents get called irrational, they do not merely understand this to mean that they fall short of some evaluative standard; they feel personally criticized for their responses (Kiesewetter 2017, 39).

> When one is rational in the relevant way, one is worthy of a certain kind of praise. And when one is irrational in the relevant

> way, one is open to a particular kind of criticism. When one is incoherent one is open to this sort of criticism (Lord 2018, 4).

Second, following Kiesewetter (2017, 2), we can distinguish between two senses of 'rationality'. The first is the *capacity sense* of rationality, according to which we are rational beings because we possess the capacity for rationality. If a creature can believe or act in compliance with norms of rationality, such as modus ponens and means-ends reasoning, we may say that the creature is rational in this capacity sense. In this context, 'rational' contrasts with 'arational', not with 'irrational'. The second sense is the standard-related sense of rationality, where 'rational' contrasts with 'irrational'. There are certain standards or norms that are authoritative or binding for any rational being. For example, suppose X is an arbitrary rational being. If X believes that $\sim q$, despite believing both that if p then q, and that p, then X can be rationally criticized for being irrational on the grounds that X fails to adhere to the rule of modus ponens, which is a norm of rationality. As this example illustrates, irrationality involves the violation of a standard or norm of rationality that applies universally to all rational beings.

Third, rational criticism is closely tied to justification. For instance, if one's attitude or action is justified, it should not be subject to rational criticism. Conversely, if a response is unjustified, it may be subject to rational criticism. Furthermore, if an individual is criticized for their response, they have the right to demand justification. If this demand cannot be met, they can rightly dismiss the criticism.

Fourth, it is inevitable to address any rational question on the basis of our social practice of demanding justification and responding to such demands. Let me explain. As Ludwig Wittgenstein (1969, §§341–43) points out, genuine doubt is possible only against a background of beliefs that are not simultaneously doubted. Additionally, according to Immanuel Kant (1996b, 5:16), reason has its own sovereignty, meaning that reason must answer any question about itself and its principles from its own resources upon due reflection. To put it differently, our conceptual framework provides the norms, criteria, or rules for defending or criticizing any claim. Therefore, it is inevitable to address any demand for justification on the basis of our conceptual framework. Consequently, it can be argued that

meeting such a demand requires engaging in rational discourse grounded in that framework. In other words, any demand for justification is not intelligible outside of our conceptual framework. Here 'a conceptual framework' can be understood roughly as a framework on the basis of which we can settle genuine doubt or meet demands for justification.[5]

In addition, according to Sellars (1963), our concept of justification has been developed on the basis of our social practices of demanding justification and responding to such demands (or giving and asking for reasons). So our concept of justification should be understood in accordance with this social practice model of justification. And on this social practice model, we have no other rational way to address any demand for justification except on the basis of our social practice of justification.[6]

Moreover, as Robert Brandom (1994, esp. 204–206) argues, our social practice of justification requires the default-and-challenge structure of justification. In the first place, the infinite regress of justification is impossible in our social practice. Suppose that we defend a claim by offering a ground, *p*. A challenger can call this ground into question by saying, 'Why *p*?' To meet this challenge, we might provide another ground, *q*. The challenger can, in turn, call this ground into question by saying, 'Why *q*?' Here it should be noted that if the challenger were allowed to keep raising a question, 'Why is that?' to any of our replies, there would be no claim that we can ultimately justify. In the second place, we are justified in accepting a claim just in case we can answer all objections raised against it within our social practice of justification. It is important to note that such objections can be addressed only on the basis of a background of beliefs and norms

[5] Building upon this perspective, Sellars (1963, §36, 169) distinguishes the logical space of reasons from the realm of law. Objects within the realm of law are those which can, in principle, be explained by natural sciences. But reasons are not the kind of thing that can be explained by natural sciences. To put the point another way, causal relations are one thing, and justificatory relations are quite another. Consequently, the logical space of reasons are *sui generis* in that it cannot be conceptually reduced to the realm of law. Therefore, we have no other rational way to address any demand for justification except on the basis of reasons within this logical space.

[6] For a detailed defense of a Sellarsian social practice theory of justification, see my previous work (Lee 2017; Lee 2021).

that hold positive justificatory status within that practice. Therefore, the possibility that we can defend something requires that some claims (or norms) be treated as having default positive justificatory status in our social practice of justification, unless they are successfully challenged with good positive reasons. And as for claims that have default positive justificatory status, we can shift the burden of proof to a challenger, if any.

Fifth, and finally, there are valid rules of reasoning which have positive justificatory status in our social practice of justification. To illustrate, consider the following two arguments:

(1) If it is raining, then the ground is wet. It is raining. Therefore, the ground is wet.
(2) I shall visit Venice. Buying a ticket is a necessary means of visiting Venice. So I shall buy a ticket.[7]

(1) is an instance of modus ponens, which has the following form: "If *p*, then *q*. *p*. Therefore, *q*." This rule of theoretical reasoning can be considered valid because the joint truth of its premises is incompatible with the falsity of its conclusion. In other words, the joint truth of its premises implies the truth of its conclusion.

Let us now turn to (2), which is an instance of means-ends reasoning, which has the following form: "I shall do *E*. Doing *M* is a necessary means of doing *E*. Therefore, I shall do *M*." And we can defend the validity of this rule in the following way. Suppose that S intends to do *E*. Here it might be worth distinguishing between expressing an intention and describing an intention. S can express his intention of doing *E* by saying 'I shall do *E*'. And we can describe this intention by saying 'S intends to do *E*'. Admittedly, expressions of intention, unlike descriptions of intentions, are neither true nor false; in other words, they do not have truth conditions. Nonetheless, we can think of them as having the following success conditions: S's intention to do *A* is realized if he does *A*, otherwise it is unrealized. Now observe that insofar as S's doing *M* is a necessary means of doing *E*, S's intention to do *E* is realized only if his intention to do *M* is realized. In other words,

[7] In this paper, following Sellars (1967, 179), I will use 'shall' and 'will' in such a way that 'shall' expresses an intention, whereas 'will' expresses the standard indicative future.

when the second premise of means-ends reasoning is true, the realization of its first premise is incompatible with the non-realization of its conclusion. Therefore, we can regard means-ends reasoning as a valid rule of practical reasoning.

Now, with the above five key points on the table, consider a scenario where S believes the two premises of (1), but fails to believe its conclusion, despite caring about whether the conclusion is true. In this scenario, S can be criticized for being modus-ponens incoherent. If S is criticized in this manner, he has the right to demand justification. But this demand can be met on the grounds that he violates a valid rule of theoretical reasoning. Let me elaborate on this point a bit further.

To begin, we need to distinguish between the following two questions:

(3) Why are we all bound by authoritative norms of rationality?
(4) Which norms are authoritative norms of rationality?

We can address (3) based on a Kantian view of rational beings. According to this view, we differ from mere animals because we are normatively bound by authoritative norms of rationality, which apply to all rational beings. Thus, the reason why we are all bound by those norms is that we are rational beings.[8] We can address (4) based on the aforementioned Sellarsian social practice theory of justification. According to this theory, we are normatively bound by a norm of rationality insofar as it holds positive justificatory status within our social practice of justification. As mentioned earlier, modus ponens can be regarded as a valid rule of theoretical reasoning because the joint truth of its premises is incompatible with the falsity of its conclusion. Thus, this rule holds positive justificatory status within our social practice of justification. In other words, we may regard modus ponens as a valid rule of theoretical reasoning unless we are given reasons to override its positive justificatory status within that practice. Hence, we can explain why S can be justifiably criticized for violating modus ponens: the rule holds positive justificatory status as a valid rule of reasoning within our social practice of justification.

A parallel argument can be made about means-ends incoherence. To illustrate, consider again the scenario where S intends to visit Venice,

[8] For a detailed defense of this Kantian view, see Lee (2023).

believes that buying a ticket is necessary, but fails to intend to buy one. In this scenario, S can be criticized for means-ends incoherence. In this context, two things are worth pointing out. First, Brunero acknowledges that means-ends coherence is a genuine rational requirement. This stance leads him to reject the myth theory, which holds that the rational requirement of means-ends coherence is a myth.[9] Second, he admits that agents exhibiting this incoherence are irrational. His claim is that this irrationality is best explained by his non-normative disjunctivism. However, when criticized for means-ends incoherence, an individual has the right to demand justification. This demand can be met on the grounds that the individual violates a valid rule of practical reasoning, which holds positive justificatory status as a norm of rationality within our social practice of justification. Let me elaborate on this point further.

To begin, as previously mentioned, we are all bound by authoritative norms of rationality because we are rational beings. Additionally, we are normatively bound by a norm of rationality if it holds positive justificatory status within our social practice of justification. As previously argued, means-ends reasoning can be defended as a valid rule of practical reasoning because, when the second premise of means-ends reasoning is true, the realization of its first premise is incompatible with the non-realization of its conclusion. Therefore, we may regard it as a valid rule of practical reasoning unless we are given reasons to override its positive justificatory status. Therefore, we can explain why S can be justifiably criticized for violating means-ends reasoning: this rule holds positive justificatory status as a valid rule of practical reasoning within our social practice of justification.

In addition, being subject to rational criticism can be best understood as being normatively bound by norms of rationality. As noted, when an individual is criticized for irrationality, they have the right to demand justification. This demand can be best met by appealing to our norms of rationality, which hold positive justificatory status within our social practice of justification. For example, consider a scenario in which S believes the two premises of (1) but fails to accept its conclusion, despite caring about the truth of that conclusion. In such cases, S can be criticized for modus

[9] See Brunero (2020, 212). The myth view is defended notably by Kolodny (2005) and Raz (2005).

ponens incoherence. His demand for justification can be met by pointing out that modus ponens is a valid rule of theoretical reasoning. Consequently, if S fails to correct this violation and does not offer an adequate justification, he can be criticized for being irrational. An analogous point can be made about norms of practical rationality. Therefore, if one believes or acts in a way that violates a norm of rationality, one is open to rational criticism. This shows that our social practice of rational criticism involves the assumption that we are bound by norms of rationality.

Moreover, it can also be argued that there is a reason to be means-ends coherent. Since the argument (2) is valid, if its premises hold, then its conclusion must also hold. Consequently, if S is justified in accepting these premises, then S is also justified in accepting the conclusion. Therefore, if S believes that the premises of (2) are justified (or correct), these premises provide him with a reason to form the intention expressed in the conclusion – particularly in light of his practical goal of visiting Venice. On this basis, it can be argued that the premises of a valid practical argument provide S with a reason to form the intention expressed in the conclusion, provided that those premises are justified. This reason also serves as a consideration against not intending to buy a ticket. Thus, S has a reason against not intending to buy a ticket, at least insofar as he cares about visiting Venice. If this is correct, we can also explain why S can be justifiably criticized for means-ends incoherence: the premises of (2) provide a reason against not intending to buy a ticket. Note that if S does not intend to buy a ticket, he is unlikely to achieve his goal of visiting Venice. Therefore, S's failure to form this intention is not justified, insofar as he genuinely cares about visiting Venice. I will say more on this point in the next section.

3. The Normativity of Rational Requirements

Brunero argues that rational requirements are not normative. He writes:

> Niko Kolodny, in his influential paper "Why be Rational?," presents several challenges to the idea that rationality is normative. According to one challenge, if one always has a reason to be

> rational, we should be able to state *what* that reason is. But it is not clear what the reason is (Brunero 2020, 17; original italics).
>
> For prudence, we can say that if prudence requires you to *X*, then there's a reason for you to *X*. And for morality, we can say that if morality requires you to *X*, then there's a reason for you to *X*. … But for rationality, we *can't* say that if rationality requires you to *X*, there's a reason for you to *X*. In short, the requirements of prudence and morality are normative, whereas the requirements of rationality are not. (Brunero 2020, 210; original italics)

As the above quotations indicate, Brunero claims that rational requirements are not normative (in the reasons-providing sense of 'normative'). On this point, he concurs with Kolodny (2005, §2). In this section, however, I argue against this claim.

As argued in the previous section, if you violate a rational requirement such as means-ends coherence, you can be subject to rational criticism. In this context, it is worth noting the distinction between the deontological 'ought' and the propriety 'ought'. If one's failure to act in accordance with an 'ought' implies one's being culpable or blameworthy for it, then the 'ought' is the deontological 'ought'. The representative example is the moral 'ought'. But there is another kind of 'ought' which does not imply such culpability or blameworthiness. For example, if a botanist says about a trillium that it ought to have three petals, then they do not mean that the trillium is culpable or blameworthy for not having three petals; instead, what they mean is just that it is appropriate or proper for the trillium to have three petals. In such cases, the 'ought' is the propriety 'ought'.[10]

Given the above distinction between deontological and propriety normativity, it can be argued that rational requirements such as means-ends coherence are normative in the deontological sense. For example, as argued in section 2, individuals exhibiting means-ends incoherence can be justifiably criticized on the grounds that they violate a valid rule of practical reasoning or that the premises of means-ends reasoning provide a

[10] For the distinction between the deontological 'ought' and the propriety 'ought', see Wolterstorff (2005, 330).

subject who accept these premises with a reason to form the intention expressed in its conclusion. And such criticism involves demanding that they revise their attitudes to comply with the relevant norms of rationality. Along these lines, it can be argued that the rational requirement of means-ends coherence is normative in the deontological sense. In addition, the fact that a means-ends incoherent agent can be justifiably criticized for breaching a norm of rationality is clearly a consideration that counts against this incoherence. In other words, one ought to comply with means-ends coherence because failing to do so renders one liable to rational criticism for violating a valid rule of reasoning. Therefore, we can also argue that means-ends coherence is normative in the reasons-providing sense of 'normative'.

At this point, it is worth considering a possible objection. Attributing impoliteness is a form of criticism, but it does not necessarily imply that one is under an obligation to be polite. Analogously, attributing irrationality may likewise be a form of criticism without thereby implying an obligation to comply with rational requirements. However, as Brunero (2020, 209–214) points out, rational requirements differ fundamentally from those of chess, law, grammar, or etiquette, as they are not grounded in non-universal social rules or conventions. This distinction is crucial: whereas criticism for impoliteness is grounded in socially contingent norms, criticism for irrationality is not. Nevertheless, individuals who wish to avoid being criticized for impoliteness have a reason to adhere to the norms of etiquette. By parity of reasoning, all rational agents have a reason to comply with rational requirements, since doing so is necessary to avoid rational criticism. Accordingly, the fact that attributing impoliteness as a form of criticism does not provide everyone with a reason to be polite does not undermine the claim that attributing irrationality provides all rational agents with a reason to adhere to rational requirements. The key difference between etiquette norms and rational requirements is that, unlike the former, the latter apply universally to all rational beings. This difference does not weaken, but rather supports, the claim that there are reasons to comply with rational requirements. In this light, the absence of an obligation to be polite does not undermine the claim that rational requirements are normative – in the reasons-providing sense of 'normative'.

4. Brunero's Objections against Strong Normativity

As additional grounds for his non-normativist view, Brunero raises three important objections against the strong normativity of means-ends coherence (Strong Normativity, henceforth), the view that one always ought to be means-ends coherent. This section argues that those objections are not successful.

His first objection is that being means-ends incoherent can be advantageous (2020, 131–37). To illustrate this point, Brunero provides the following example: Professor Henri Liable, despite his sincere intentions, struggles to fulfill his departmental service responsibilities due to his incompetence. Recognizing Henri's consistent lack of follow-through, his department chair, adept at understanding people, reassigns his service responsibilities to other members of the department. Under this situation, consider the following instance of means-ends incoherence. Although Henri plans to certify his new course for general education credit and understands the need to contact relevant administrators, he has no intention of doing so. His plan to get his new course certified will secure the good will of the chair, who reliably detects such intentions, while his not intending to talk with the administrators could reduce the likelihood of being burdened with future service assignments. According to Brunero, this case poses a serious problem for Strong Normativity, because this view implies that Henri ought not to have the incoherent combination of attitudes. But the incoherent combination of attitudes is advantageous to him, and hence he has a good reason to remain means-ends incoherent, rather than abandoning one of the attitudes in this incoherent combination.

However, Strong Normativity is not undermined by such a case. We can concede that Henri's means-ends incoherence might be practically beneficial to him, at least for a while. But it is important to recognize that if an agent's attitude is really rational, the agent should not be rationally criticized for holding that attitude. But Henri can be criticized for violating means-ends reasoning. As previously emphasized, means-ends reasoning is a valid rule of practical reasoning. And its validity is not affected by any potential practical benefits that might be obtained from its violation. Therefore, it is one thing for an agent to violate such a valid rule of reasoning; it

is quite another thing for that violation to be practically beneficial to the agent. Most importantly, if the department chair demands certification for Henri's new course, Henri cannot dismiss this demand by arguing that avoiding contact with the administrators is practically beneficial to him. If he were to argue in that manner, he would lose the chair's goodwill. Furthermore, as long as Henri cannot dismiss the chair's demand for certification, he cannot justify violating means-ends reasoning on the grounds that such a violation is practically beneficial to him. Therefore, Brunero's first objection, illustrated by the case of Henri Liable, doesn't undermine Strong Normativity. For Henri's violation of means-ends reasoning is open to rational criticism, and he can hardly ward off such criticism by claiming that it is practically beneficial to him.

Let us now turn to Brunero's second objection against Strong Normativity (2020, 137–41). This objection pertains to a scenario where an agent, despite being means-ends incoherent, is unable to alter either their goal or their instrumental belief. To illustrate, Brunero presents the following example from Kieran Setiya (2007, 672). A certain agent, say S, intends to smoke, but fails to intend to buy cigarettes, despite believing that buying cigarettes is necessary to smoke. In this example, it is not the case that S ought to intend to smoke. Unfortunately, however, both his intention and instrumental belief are psychologically unalterable. Thus, the only way he can comply with means-ends coherence in this case is to intend the means. Then Strong Normativity seems to imply that S ought to intend to buy cigarettes. But this consequence is implausible.

The above objection does not pose a serious problem for Strong Normativity, either. Consider the following normative means-ends reasoning:

> S ought to smoke. His buying cigarettes is necessary to smoke. So he ought to buy cigarettes.

This reasoning is valid. So, if S accepts the two premises, then he must also accept the conclusion. However, despite his psychological inability to abandon his intention to smoke, it can still be the case that S ought not to smoke. For instance, if he has lung cancer, smoking could be life-threatening. In this scenario, S could deny that he ought to smoke, while admitting that he cannot control his intention to smoke due to a psychological compulsion**.** In other words, by rejecting the first premise in this manner, he

may also dismiss the conclusion. The point is that S can reasonably deny the first premise, while admitting his irresistible urge to smoke.[11] Besides, one may have an obligation not to do something, even if one is unable to alter one's intention to do it.[12]

A related point is that the assumption that S's intention to smoke is irrational is compatible with his psychological inability to alter his intention. Consider an analogy from belief. If it is irrational for S to believe that p (due to the high likelihood of its falsehood), then he should not believe it.

[11] It is important to note that rationality is not confined to maintaining consistency and coherence among attitudes. In this case, S is justified in buying cigarettes only if his intention to smoke is justified. But this intention is arguably indefensible if smoking poses a life-threatening risk to him.

[12] To illustrate this point, consider a psychopath who can hardly restrain his impulse to harm others because of his antisocial personality disorder. Here it should be noted that, unlike mere animals, the psychopath is a rational agent, although we may admit that he is not fully rational in some sense. For example, he is capable of understanding that he is not morally allowed to harm others. As a consequence, he can resist his impulse to harm others under certain conditions, such as when a police officer is keeping an eye on him, and so he thinks he can hardly get away with his wrongdoing. Thus, given that he is a rational agent, it follows by Kant's Formula of Humanity that he should not be treated merely as a means to some other end (see Kant 1996a, 4:429). This is tantamount to saying that he should be regarded as a member of the moral community. If so, moral norms apply to him, because moral norms are universal norms. Unfortunately, however, he is morally defective. But this moral defect is not a sufficient reason for him to be excluded from the moral community. And insofar as the psychopath is regarded as a member of the moral community, he is responsible for his wrongdoing, such as killing an innocent child, even if he can hardly refrain from doing so due to his antisocial personality disorder. In this case, the psychopath could make an excuse for his wrongdoing by claiming that his antisocial personality disorder made it impossible for him to resist his impulses. And he could be less blameworthy to the extent that resisting his impulses was beyond his capacity in a specific situation. Insofar as he is a member of the moral community, however, he cannot be completely exempt from the relevant moral responsibility. So, he could be put into a mental institution or a prison, if necessary to protect innocent people. And we are allowed to do so precisely because even the psychopath is no exception to our moral norms. Along these lines, it can be argued that even if the psychopath cannot alter his intention to harm an innocent child, the universal moral norm that prohibits such wrongdoing still applies to him, as it does to every member of the moral community.

Yet, assume that he cannot help but believe that *p*. In this case, his psychological inability to believe otherwise is consistent with the irrationality of his belief. If there is a way to remedy his psychological inability, we can justifiably demand that he receive medical or psychological treatment. If it is incurable, he might be exempt from rational criticism, but this does not change the fact that his belief in '*p*' is irrational due to the high likelihood of its falsehood. The same point applies to intention. If it is irrational for S to intend to *X*, he should not intend to do so. Yet, assume that he cannot help but intend to do so. In this case, S's psychological inability is also compatible with the irrationality of his intention. If this irrationality is treatable, we can demand that he seek treatment. If it is incurable, although S might be exempt from rational criticism, the irrationality of his intention remains unchanged.

Here I do not mean to deny that if S fails to intend to buy cigarettes, he violates the rational requirement of means-ends coherence. However, it is important to recognize that his only alternative in this case is to violate his rational requirement not to smoke. Due to his psychological inability, he faces a dilemma: either violate the rational requirement of means-ends coherence or violate his rational requirement not to smoke. Given this dilemma, it might be more reasonable for S to be means-ends incoherent than to violate his rational requirement not to smoke. He could excuse his inevitable violation of the rational requirement of means-ends coherence by arguing that the life-shortening effects of smoking are far worse than the negative consequences of breaching this rational requirement. Nevertheless, this does not change the fact that he is still rationally required to be means-ends coherent. In this case, S's point is that, despite the lack of justification for his smoking, he cannot resist it due to a psychological compulsion. He can acknowledge this compulsion precisely because his smoking lacks justification. If his smoking is unjustified, then the first premise – that he ought to smoke – is also unjustified. Consequently, the conclusion that he ought to buy cigarettes is unjustified as well. Therefore, S's case above does not show that Strong Normativity forces us to accept an unacceptable consequence that S ought to intend to buy cigarettes.[13]

[13] Bratman (2009c) also provides a significant reply to Setiya's argument against Strong Normativity. On Bratman's view, we have a reason to be self-governing, and

Finally, let us consider Brunero's third objection against Strong Normativity (2020, 141–45). According to the rational requirement of means-ends coherence, rationality requires that: if one intends to X, and believes that one will X only if one intends to Y, then one intends to Y. On the wide-scope interpretation of this requirement, the rationality requirement takes wide scope over the entire conditional. Brunero endorses this wide-cope view. Consequently, on his account, an agent can satisfy the requirement of means-ends coherence in any of three ways: by intending the means, by abandoning the end, or by rejecting the instrumental belief. With this point in mind, consider the following principle, which has some plausibility.

> *Transmission to Sufficient Means*: If you ought to do E, and doing M is a sufficient means of doing E, then there is a reason to do M.

Suppose that Bob ought to kill his colleagues. Additionally, assume that he believes poisoning them is a sufficient means of doing so, and that there is no epistemic reason for him to abandon this instrumental belief. In this

---

when self-governance is possible, there is a reason to be means-ends coherent. But when such self-governance is not possible, as when one's ends and instrumental beliefs are unalterable, there is no such reason to be means-ends coherent. In other words, on his view, the rational requirement of means-ends coherence is normative only if self-governance is possible. Against this claim, however, Brunero (2010) argues that self-governance is possible even when one's ends and instrumental beliefs are unalterable. To illustrate, he provides an example of a Frankfurtian counterfactual intervener: A certain intervener wants S to smoke. But the intervener will not interfere as long as S intends to smoke. However, if S decides not to smoke, the intervener will intervene and make S's intention ineffective. In this scenario, S cannot change his intention to smoke, but as long as he does not decide otherwise, he can carry out his intention without interference. Therefore, S is still capable of self-governance. But my response to Setiya's example is not vulnerable to this kind of problem with Bratman's view. On my view, Strong Normativity does not necessarily lead us to accept the unacceptable conclusion that S ought to intend to buy cigarettes. As argued in the main text, S can justifiably deny the first premise that he ought to smoke. It is crucial to note that his uncontrollable intention to smoke, stemming from a pathological mental condition, may not be justified. In such a scenario, the premise that S ought to smoke lacks justification. Consequently, the conclusion that he ought to buy cigarettes is also unjustified.

case, if Transmission to Sufficient Means is correct, it follows that Bob has a reason to intend to poison his colleagues. On the wide-scope view of means-ends coherence, Bob could comply with this requirement in any of three ways: by intending the means, abandoning the end, or rejecting the instrumental belief. In other words, each of these three responses is sufficient for satisfying means-ends coherence. On this basis, Brunero argues that if we accept Strong Normativity and Transmission to Sufficient Means, we must also accept that there is a reason for Bob to abandon his instrumental belief – even in cases where there is no epistemic reason to do so. According to Brunero, this is a very implausible consequence, because Bob's reason to abandon his instrumental belief is neither evidence-based nor pragmatically beneficial, and so we are forced to admit a new and problematic category of theoretical reason. However, this objection also does not pose a serious problem for Strong Normativity.

To begin, even if Bob in the above case does not intend to poison his colleagues, this does not necessarily imply that he is means-ends incoherent. He could acknowledge a reason to poison them, yet refrain from intending to do so on the grounds that he lacks a sufficient reason, and thus needs to explore a better means of achieving his goal. And insofar as he is not means-ends incoherent, he has no reason to abandon his instrumental belief, especially when there is no epistemic reason to abandon it. In addition, contrary to Brunero's claim, the acceptance of Strong Normativity and Transmission to Sufficient Means does not necessarily lead to the introduction of a new and problematic category of theoretical reason. Brunero's argument depends on the assumption that Bob has a reason to abandon his instrumental belief due to his means-ends incoherence. However, this assumption is very problematic. As we have previously argued, Bob may not be means-ends incoherent in the first place. Therefore, we may deny that Bob has a reason to abandon his instrumental belief. If so, embracing Strong Normativity does not force us to admit a new and problematic category of theoretical reason. For these reasons, Brunero's third objection also fails to successfully challenge Strong Normativity.

5. Concluding Remarks

On Brunero's non-normative disjunctivism, agents exhibiting means-ends incoherence are irrational because they are guaranteed to have at least one attitude that fails to achieve its constitutive aim, and such an aim failure is not a failure of reason. Against this view, I have argued that it fails to adequately explain the irrationality of means-ends incoherence. More specifically, I have argued that agents exhibiting this incoherence can be justifiably criticized for irrationality, and they cannot ward off such criticism by claiming that the rational requirement of means-ends coherence is only apparently normative. Furthermore, I have argued that Brunero's three objections against the strong normativity of means-end coherence are not successful.

If the arguments presented in this paper are correct, the rational requirement of means-ends coherence is strongly normative. For example, consider again the aforementioned scenario where S intends to visit Venice, believes that buying a ticket is necessary, but fails to intend to buy one. In this scenario, S can be criticized for means-ends incoherence. When criticized, S has the right to demand justification. As previously argued, this demand can be met on the grounds that S violates means-ends reasoning, which holds positive justificatory status as a valid rule of practical reasoning within our social practice of justification.

If this paper is correct in claiming that agents not adhering to rational requirements can be justifiably criticized, its implications extend beyond the critique of Brunero's non-normative disjunctivism. Let me mention two notable consequences.

As argued in section 3, rational requirements such as means-ends coherence are normative in the deontological sense. Consequently, the arguments presented in this paper provide grounds for refuting any view that denies the deontological normativity of rational requirements. For example, according to Alex Worship (2021), there are requirements of structural rationality, which include means-ends coherence. He acknowledges that these requirements are normative – in particular, he regards means-ends incoherence as a kind of defect that can be criticized as irrational. Nevertheless, he contends that such a criticism is evaluative in nature, concerned with whether an agent falls short of an evaluative standard. As such, it is akin

to criticizing a person for being unintelligent in some respect, a performance for lacking skill, or an agent for failing to empathize with others. Therefore, the arguments advanced in this paper provide reason to reject Worship's account of rational requirements.

In addition, the arguments presented in this paper enable us to reject any view that denies the normativity of rational requirements. For example, Samuel Fullhart and Camilo Martinez (2024) agree with Brunero that means-ends coherence fundamentally concerns whether it is logically possible for one's attitudes to satisfy a certain success condition. They extend this view by arguing that the possibility of joint attitudinal success can explain coherence requirements in general. According to them, a set of attitudes is coherent if, roughly, it is logically possible for the attitudes to be jointly satisfied. Satisfaction here is understood as a kind of fit between each attitude and the world. For instance, a belief is satisfied if and only if it fits the world – that is, if it represents the world as it actually is. Therefore, their view of coherence as joint satisfiability does not construe failures of rational requirements, such as means-ends coherence, as failures of reason. As a result, it also fails to adequately explain why agents who do not comply with a rational requirement can be justifiably criticized.

Finally, I agree with Kiesewetter (2017) and Lord (2018) that we always ought to be means-ends coherent. However, there are important differences between my view and theirs regarding rationality and reasons. Most significantly, they endorse the reasons-first approach to rationality. According to this approach, the concept of a reason is more fundamental than the concept of rationality, and rationality is a matter of appropriately responding to reasons.

More specifically, Kiesewetter proposes the Evidence-Relative Account of Reasons, arguing that rationality consists in correctly responding to reasons available to us – where a reason is available if it is part of our evidence. On the other hand, Lord advocates the Reasons Responsiveness View, according to which rationality consists in correctly responding to the objective reasons one possesses. Consequently, on their view, means-ends incoherence is irrational because it involves a failure to correctly respond to some available reason or possessed objective reason. By contrast, I do not accept the reasons-first approach. Instead, I endorse a Sellarsian coherence theory of

justification, according to which any rational question can be addressed only within our social practice of demanding justification and responding to such demands. This is a coherence theory that rejects the idea that reasons are primitive or foundational in either epistemic or practical justification. Determining which of these views is ultimately correct lies beyond the scope of this paper. Nevertheless, it may be helpful to briefly indicate why I reject the reasons-first approach.

According to the reasons-first view, the concept of a reason is more fundamental than that of reasoning, and thus correct reasoning must be explained in terms of reasons rather than vice versa. However, this consequence is very implausible. To begin, the concept of a reason is inseparable from that of justification: x is a good reason for y only if x enhances the justification of y. Justification, in turn, depends partly on reasoning. For instance, suppose you assert that '*p*' is true. We can then ask why you assert this, effectively demanding that you justify the assertion. To meet this demand, you must offer a reason, say '*r*', for this claim. Your justification for '*p*' can thus be expressed as an argument: "*r*. Therefore, *p*." But a correct argument depends on two factors: the premises must be true (or correct), and the argument must be valid or good, meaning that if the premises are true, the conclusion is also true or likely to be true. Therefore, '*r*' can serve as a good reason for '*p*' only if the reasoning from the premise '*r*' to the conclusion '*p*' is valid or good. This strongly suggests that reasons do not take precedence over reasoning.

What then is a reason? On the Sellarsian social practice theory of justification, we must begin with a social practice of justification to address rational questions. The question 'What is a reason?' is no exception. As a coherence theory, this approach rejects foundationalist approaches to normativity, including any reductive account of reasons. On this coherence theory, although we cannot provide a reductive account of reasons, we can offer a conceptual explication of the concept of a reason. Roughly, to say that *X* is a good reason for *Y* is to say that *X* is a consideration that justifies, or contributes to justifying, *Y*. Justification, in turn, can be understood in terms of answering all objections or beating all competitors within our social practice of justification.[14] Two points merit emphasis. First, this explication

[14] For a detailed defense of this view of justification, see Lee (2022).

is based on a Sellarsian social practice theory of justification, which can appeal to the default-and-challenge structure of justification. Second, this explication also embraces concept holism, specifically in the form of an inferentialist theory of meaning. Accordingly, the conceptual interconnection between justification, reasons, and reasoning is not a problem but rather a natural consequence of concept holism. Therefore, despite this mutual dependence among these concepts, 'answering all objections' can still be regarded as a good explication, allowing us to understand the concept of justification more clearly. As noted earlier, a proper defense of this alternative account of reasons lies beyond the scope of the present paper. I intend to take up that task in a separate, independent paper.

What I want to emphasize, for the purpose of this paper, is that my account offers a better defense of the normativity of means-ends coherence than competing views, such as those advanced by Kiesewetter or Lord. On my account, means-ends reasoning has normative force because it currently holds positive justificatory status as a valid rule of practical reasoning within our social practice of justification. This contrasts with the view that an agent who fails to exhibit means-ends coherence is merely failing to respond correctly to some reason. If means-ends incoherence is conceived simply as a certain combination of attitudes, its normative status may appear obscure. However, when means-ends reasoning is understood as a norm of rationality that has normative force within our social practice of justification, it becomes clear why means-ends coherence is normative.

References

Alvarez, Mara. 2016. "Reasons for Action: Justification, Motivation, Explanation." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Winter 2016 edition.

Brandom, Robert. 1994. *Making It Explicit*. Cambridge: Harvard University Press

Bratman, Michael. 2009a. "Intention, Belief, and Instrumental Rationality." In *Reasons for Action*, edited by David Sobel and Steven Wall, 13–36. Cambridge: Cambridge University Press. https://doi.org/10.1017/cbo9780511720185.002

Bratman, Michael. 2009b. "Intention, Belief, Practical, Theoretical." In *Spheres of Reason*, edited by Simon Robertson, 29–62. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199572939.003.0002

Bratman, Michael. 2009c. "Intention, Practical Rationality, and Self-Governance." *Ethics*, 119(3): 411–43. https://doi.org/10.1086/599985

Brunero, John. 2010. "Self-Governance, Means-Ends Coherence, and Unalterable Ends." *Ethics*, 120(3): 579–91. https://doi.org/10.1086/652448

Brunero, John. 2020. *Instrumental Rationality: The Normativity of Means-Ends Coherence*. Oxford: Oxford University Press.

Côté-Bouchard, Charles. 2016. "Can the Aim of Belief Ground Epistemic Normativity?" *Philosophical Studies*, 173(12): 3181–98. https://doi.org/10.1007/s11098-016-0657-8

Fullhart, Samuel, and Camilo Martinez. 2024. "Coherence as Joint Satisfiability." *Australasian Journal of Philosophy*, 102(2): 312–32. https://doi.org/10.1080/00048402.2023.2269963

Kant, Immanuel. 1996a. "Groundwork of the Metaphysics of Morals." In *Practical Philosophy*, translated and edited by Mary J. Gregor, 37–108. Cambridge: Cambridge University Press.

Kant, Immanuel. 1996b. "Critique of Practical Reason." In *Practical Philosophy*, translated and edited by Mary J. Gregor, 135–271. Cambridge: Cambridge University Press.

Kiesewetter, Benjamin. 2017. *The Normativity of Rationality*. Oxford: Oxford University Press.

Kolodny, Niko. 2005. "Why Be Rational?" *Mind*, 114(455): 509–63. https://doi.org/10.1093/mind/fzi509

Kolodny, Niko. 2008. "The Myth of Practical Consistency. *European Journal of Philosophy*, 16(3): 366–402. https://doi.org/10.1111/j.1468-0378.2008.00325.x

Lee, Byeong D. 2017. The Truth-Conduciveness Problem of Coherentism and a Sellarsian Explanatory Coherence Theory." *International Journal of Philosophical Studies*, 25(1): 63–79. https://doi.org/10.1080/09672559.2016.1236140

Lee, Byeong D. 2021. "A Coherentist Justification of Epistemic Principles and Its Merits." *Acta Analytica*, 36(4): 533–51. https://doi.org/10.1007/s12136-021-00463-3

Lee, Byeong D. 2022. "Knowledge as Objectively Justified Belief." *Acta Analytica*, 37(3): 397–414. https://doi.org/10.1007/s12136-021-00495-9

Lee, Byeong D. 2023. "The Authority and Content of Morality: A Dilemma for Constitutivism and a Coherentist Approach to Normativity." *The Journal of Value Inquiry*, published online 20 October, 2023. https://doi.org/10.1007/s10790-023-09962-5

Lord, Errol. 2018. *The Importance of Being Rational*. Oxford: Oxford University Press.

Raz, Joseph. 2005. "The Myth of Instrumental Rationality." *Journal of Ethics and Social Philosophy*, 1(1): 2–28. https://doi.org/10.26556/jesp.v1i1.1

Sellars, Wilfrid. 1963. *Science, Perception and Reality*. Atascadero: Ridgeview.

Sellars, Wilfrid. 1967. *Science and Metaphysics: Variations on Kantian Themes*. Atascadero: Ridgeview.

Setiya, Kieran. 2007. "Cognitivism about Instrumental Reason." *Ethics*, 117(4): 649–73. https://doi.org/10.1086/518954

Velleman, David. 2000. "On the Aim of Belief." In *The Possibility of Practical Reason*, edited by David Velleman, 244–81. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198238256.003.0011

Wittgenstein, Ludwig. 1969. *On Certainty*, edited by G.E.M. Anscombe and G.H. von Wright, and translated by D. Paul and G.E.M. Anscombe. New York: Harper & Row.

Wolterstorff, Nicholas. 2005. "Obligation, Entitlement, and Rationality." In *Contemporary Debates in Epistemology*, edited by Matthias Steup and Ernest Sosa, 326–38. Oxford: Blackwell Publishing.

Worship, Alex. 2021. *Fitting Things Together: Coherence and the Demands of Structural Rationality*. New York: Oxford University Press. https://doi.org/10.1093/oso/9780197608142.001.0001

Organon F 32 (3) 2025: 348–354 ISSN 2585-7150 (online)
https://doi.org/10.31577/orgf.2025.32304

RESEARCH ARTICLE

A Quandary for the Naturalist

Howard Sankey*

Received: 11 July 2024 / Revised: 1 December 2024 / Accepted: 6 April 2025

Abstract: The paper raises a quandary for the naturalist friend of truth who rejects the *a priori* outright. The quandary is that instances of the T-scheme are analytic, hence knowable *a priori*. The naturalist must either renounce their friendship with truth or soften their stance on the *a priori*. The paper recommends the latter option.

Keywords: analytic; *a priori*; naturalism; truth.

1. Naturalism and the *a priori*

The naturalistic friend of truth faces an unremarked quandary with the T-scheme. The T-scheme is analytic, hence knowable *a priori*. That does not sit well with the naturalist who rejects the *a priori*.

Some naturalists (e.g., Devitt 2005) reject the *a priori* outright. It is obscure and we do not need it. Other naturalists (e.g. Papineau 2011) allow that there may be some *a priori* knowledge but downplay its role. It is philosophically insignificant.

* University of Melbourne
https://orcid.org/0000-0002-5168-150X
School of Historical and Philosophical Studies, University of Melbourne, Grattan Street, Parkville, Victoria 3010 Australien
chs@unimelb.edu.au

I am concerned here with naturalists of the former variety. The quandary I shall present is one that confronts the naturalist who rejects the *a priori* outright. The naturalist who downplays the *a priori* avoids the quandary.

I shall assume that the naturalist is a friend of truth. The friend of truth is one who thinks that truth exists, that it is valuable, that it is something that we aim for and sometimes achieve, that it is required for knowledge, and that it may play a role in explaining the success of our endeavours. But perhaps the naturalist should reconsider this friendship. I shall explore that option after presenting the quandary and asking whether it may be resolved by holistic considerations.

2. The T-Scheme

To begin with, let us remind ourselves of the T-scheme:

'P' is true iff P.

The T-scheme is widely thought to provide important insight into the nature of truth. Some philosophers take the T-scheme to tell us most of what there is to know about truth. Others hold that the T-scheme is to be supplemented in some way to arrive at a full-blown account of truth.

What is the status of the T-scheme? Is it true? Strictly speaking, the T-scheme is not itself a sentence that asserts a specific proposition. It is a scheme or schema that can be used to produce a sentence that asserts a proposition. We must therefore focus on particular instances of the T-scheme.

Let us take Tarski's own preferred example as paradigm (Tarski 1972):

(S) 'Snow is white' is true iff snow is white.

When first exposed to a sentence like (S), bewilderment is a common reaction. It seems obviously true, a truism, even uninformative. It is not uninformative. But it is indeed an obvious truism. For (S) is an analytic truth, something that can be known *a priori* to be true.

To see this, consider how you work out that (S) is true. For someone who does not know what 'true' means, 'S' informs them that the predicate 'is true' is only to be applied to a sentence when the state of affairs specified

by the sentence in fact obtains. For someone who does know what 'true' means, (S) is a statement of the obvious. For to say of the sentence mentioned on the left-hand side that it is true is just to say that the state of affairs specified on the right-hand side does indeed obtain. It is almost entirely what is meant by 'true' that what it is to say that a sentence is true is precisely that the state of affairs reported by the sentence does indeed obtain.[1]

Given that the attribution of truth to a sentence means that what the sentence reports to be the case obtains, we may conclude that (S) is analytic. But the notion of analyticity is subject to a well-known ambiguity (Boghossian 1996). A sentence is metaphysically analytic if it is true by virtue of its meaning alone. A sentence is epistemically analytic if one who understands the sentence is thereby justified in believing that the sentence is true.

I think (S) is analytic in both senses. First, for a sentence to be true the state of affairs that it reports must obtain. For (S) to be true, it must be the case that 'Snow is white' is true just in case snow is white. But it is precisely because (S) tells us that snow must be white for 'Snow is white' to be true that (S) is true. It is true in virtue of what 'true' means. So (S) is analytic in the metaphysical sense. Second, if one understands what (S) means, then one sees immediately that it must be true. For if one understands what 'true' means, one thereby understands that for 'Snow is white' to be true snow must in fact be white. Equally, one who understands what 'true' means will see immediately that, if snow is in fact white, then the sentence 'Snow is white' must be true. One who grasps the meaning of (S) is thereby justified in believing it to be true. So (S) is epistemically analytic as well.[2]

---

[1] The reason I say "almost entirely" is that I do not wish to say that it is entirely what is meant. To say that 'Snow is white' is true is not to say exactly what one says when one says, "Snow is white". The former attributes a property to the sentence whereas the latter does not. I do not regard the truth-predicate as devoid of meaning in the way that the redundancy theory of truth does.

[2] In this section, I have assumed that it is possible for (S) to be true. This raises the question whether the truth predicate applied to (S) is the same as the one employed in (S). In English, it is possible to employ a metalanguage at a range of

3. The Quandary

From here, it is a short step to the point that knowledge of (S) is *a priori*. For, if (S) is analytic, then no evidence is required in order for one to know that (S) is true. It is just a matter of what (S) means. Once it is recognized what (S) means, one thereby arrives at knowledge that (S) is true. No empirical evidence is required. It is known on an *a priori* basis.[3]

Here the naturalist friend of truth may need to reconsider that friendship. For if the naturalist is one who denies the *a priori* outright, they face a quandary. The quandary is what to do with (S) (or any other instance of the T-scheme more generally). (S) is known *a priori*. But the naturalist under consideration here is one who rejects the *a priori* outright. They cannot both be a friend of truth who accepts the T-scheme and reject the *a priori* outright. To remain friends, they may no longer reject the *a priori* outright. To persist with outright rejection, they must renounce the friendship.

4. The Holist Option

But before deciding whether to remain friends or persist with rejection, there is an option that the naturalist may explore. Maybe there is no need to agree that (S) is known *a priori* in the first place.

If one adopts a holistic epistemology along the lines of Quine (1953), one might argue that there is no *a priori* knowledge at all. Our beliefs form complex systems. Some of our beliefs are directly exposed to experience while others are protected from direct exposure to experience. This idea is captured in the well-known image of our belief systems as having a core and

levels as required. The truth predicate is shared across levels. It retains the same functional role across levels in a manner that ensures that it occurs at all levels with constant meaning.

[3] Of course, one may object that, surely, some empirical experience is required for a subject to even have a grasp of what (S) means. But that is the so-called "enabling" role of experience rather than the evidential role. Experience is required in order to possess the conceptual content of (S), not to recognize that it is true.

a periphery. Beliefs situated near the periphery, typically ones immediately connected with perception, are the ones directly exposed to experience. Beliefs situated closer to the core of the belief system, such as the truths of logic and mathematics, are not subject to immediate exposure to experience.

On this model, beliefs that lie at the centre of the belief system are protected from direct refutation in light of experience. But they remain open to possible revision in light of experience. As such, they are not strictly speaking *a priori* at all. Rather, like all beliefs, they are *a posteriori*. It is just that their connection with experience is indirect. They may indeed be among the least likely beliefs to be revised or replaced. And yet experience might in principle require revisions of the belief system that stretch all the way to the interior of the system. As such, beliefs that lie close to the centre of the belief system may be revised ultimately in response to occurrences at the periphery where the belief system comes into direct contact with reality by way of experience.

The naturalist might very well adopt this holistic approach (cf. Devitt 2005, 107). They might then say that the T-scheme (or instances of the T-scheme) are like the truths of logic and mathematics in that they lie close to the centre of the belief system. As such, they only appear to be *a priori*. But really, they are not. Like all the rest of our beliefs, they are open to possible revision in light of experience. Appearances to the contrary, (S) (and other instances of the T-scheme) are *a posteriori* truths after all.

While I have considerable sympathy with this picture, I find it difficult to reconcile with the case at hand. It is one thing to say that there are beliefs that are situated more or less close to the edge of the belief system. It is quite another to adopt the holistic approach to justification that goes along with this picture. I have no doubt that there are some beliefs whose justification depends on systemic considerations, such as whether they fit with closely associated beliefs within the belief system, or whether they provide a simpler account of some phenomena than other beliefs.

But the recognition that (S) is true is not obviously a matter of weighing up how the belief that (S) fits with surrounding beliefs in the belief system. Once one grasps the meaning of (S), it is a simple matter of recognizing that (S) must be true. The epistemic credentials of belief in (S) do not readily admit of construal in holistic terms. Rather, in this kind of case the

justification of the belief is immediately tied up with the understanding of the content of the belief. For this reason, I do not think that the holistic approach provides an out for the naturalist who outright rejects the *a priori*.

5. The Need for Truth

At this point, the naturalist may consider abandoning the friendship altogether. In the same way that some realists take realism to be an ontological position with no semantic component, the naturalist might declare that they have no need for truth, or the T-scheme for that matter.

How would this go? The naturalist might say that they are interested in substantive matters, such as the way the world is, rather than the question whether claims about the world are true. Or perhaps rather than seeking true beliefs about the world, the naturalist might hold that we wish to believe that the world is a given way only when the world is that way. The naturalist might, in other words, wish to reframe the various ways in which they previously trafficked in truth without using the word 'true'.

But while I have no doubt that such laborious circumlocution may be possible it would appear to be to no avail. For what is it to say that one is only interested in the substantive question of the way the world is, if it is not to ask what the truth about the world is? And what is it to say that we wish only to believe that the world is a given way when it is that way, if not to say that we wish only to believe the truth? The concept of truth is at play even if the word is not.

I do not think that this is the way to go for the naturalist. The friendship is not one that is so easily given up. Instead, I think the naturalist who outright rejects the *a priori* should reconsider their outright rejection.

6. The T-Scheme as *a priori*

What I would like to suggest is that the naturalist who outright rejects the *a priori* must remain a friend of truth but soften their stance on the *a priori*. Rather than reject the *a priori* outright, the naturalist should allow that there are some *a priori* truths. In the present context, they should

allow that the T-scheme (or, to speak strictly, its instances) may be known *a priori* to be true. Whether they should extend this attitude more broadly to other claims that might have a claim to being analytic and *a priori* is not something that needs to be decided here. And indeed, it is something that I would suggest is to be decided on a case-by-case basis. We may let the chips fall where they may. What can be granted now is that the naturalist friend of truth should maintain that friendship and forego outright rejection of the *a priori*.

Acknowledgements

I am grateful to an anonymous referee for this journal whose comments led to an improvement in the paper.

References

Boghossian, Paul (1996), "Analyticity Reconsidered." *Nous*, 30(3): 360–91 https://doi.org/10.2307/2216275

Devitt, Michael (2005), "There is No *A Priori*." In *Contemporary Debates in Epistemology*, edited by M. Steup and E. Sosa, 105–14. Oxford: Blackwell. https://doi.org/10.1002/9781394260744.ch8

Papineau, David (2011), "The Philosophical Insignificance of *A Priori* Knowledge." In *What Place for the A Priori?*, edited by M. J. Shaffer and M. Veber, 61–84. Chicago: Open Court.

Quine, Willard, V. O. (1953), "Two Dogmas of Empiricism." In *From a Logical Point of View*, 20–46. New York: Harper & Row.

Tarski, Alfred (1972), "The Semantic Conception of Truth." In *Semantics and the Philosophy of Language*, edited by L. Linsky, 13–49. Urbana: University of Illinois Press.

Organon F 32 (3) 2025: 355–377 ISSN 2585-7150 (online)
https://doi.org/10.31577/orgf.2025.32305

RESEARCH ARTICLE

The Criteria for the Emergence of Collective Epistemic Traits

Strahinja Đorđević* – Andrea Berber**

Received: 3 August 2024 / Accepted: 6 April 2025

Abstract: The aim of this paper is to offer an analysis of the emergence that group epistemic traits may exhibit. Our research implements a methodological shift in the study of collective traits from defensive argumentation against summativism (the position that reduces the collective traits to the sum of members' traits), to offering a positive thesis on the nature of irreducible collective traits. We start from the assumption that there are cases of emergent epistemic traits. By analyzing and comparing examples of emergent traits, we establish three criteria that a trait must meet in order to be considered emergent. Additionally, based on the established criteria we introduce a distinction between the two types of emergences – strong and weak. In the end, we discuss whether there is a correlation between the types of emergence and types of epistemic traits (viz. character and faculty).

* University of Belgrade
https://orcid.org/0000-0003-0076-7861
Institute for Philosophy, Faculty of Philosophy, University of Belgrade, Čika-Ljubina 18-20, Beograd 11000, Srbija
strahinja.djordjevic@f.bg.ac.rs

** University of Belgrade
https://orcid.org/0000-0002-2463-410X
Institute for Philosophy, Faculty of Philosophy, University of Belgrade, Čika Ljubina 18-20, Beograd 11000, Srbija
berberandrea@gmail.com

Keywords: collective virtue epistemology; emergence; summativism; non-summativism; weak and strong emergence.

Introduction

A significant part of intellectual work and epistemic endeavors in science, business, and everyday life takes place within larger or smaller groups. For that reason, interaction within groups should be an important field for the study of epistemic virtues and vices. However, groups are not just an environment for the manifestation of individual virtues and vices, but groups as such can be proper agents for attributing epistemic traits.[1] The main idea of this research is that there are cases of group traits that are irreducibly collective or emergent. These particular traits cannot be explained summatively – by being reduced to the sum of individual characteristics of group members. With this in mind, we want to analyze examples of group traits to determine which conditions a trait needs to meet in order to be considered emergent.

The current thematization of collective traits in the literature is mostly focused on the arguments against the position that tends to analyze all group traits as a mere sum of individual traits. So the strategy in dealing with this topic is mostly negative and defensive – it is argued that not all group traits should be analyzed in a summative way (e.g., Gilbert 1989; Schmitt 1994; Lahroodi 2007). The strategy we employ in this paper is different: we will try to get to the essence of emergence, that is, to answer the question of what makes emergent traits irreducibly collective by offering criteria for emergence. This kind of understanding can have further importance for research relating to the attribution of different collective states, such as knowledge, belief, phenomenal states, desires, and intention to the groups.

The paper is structured as follows. In the first section, we explicate the basic assumptions used in this research and explain the summativist position concerning collective epistemic traits. Section two introduces examples

[1] In this paper, the term epistemic trait is used as a term that encompasses both epistemic virtues and epistemic vices.

of collective traits that are not subjectable to summative analysis. In the next section, section three, we analyze and compare the cases of collective epistemic traits which are not explainable in a summative manner. Section four proposes the general criteria for the emergence of collective traits and elaborates on them. In the final, fifth section we compare faculty and character collective epistemic traits with regards to the types of emergences they can manifest.

1. Setting the Stage for Research

At the outset, we want to explicate basic assumptions on which our investigation rests. Firstly, we want to stipulate that the types of groups relevant to our research are the established social groups since they can be seen as "paradigmatic cases of collectives" (Lahroodi 2019, 408). Established social groups, e.g. families, teams of coworkers, or juries, are characterized by a high degree of social cohesion and interaction between their members (cf. Ritchie 2020, 402–3), unlike loosely associated groups (people who have the same profession, or people older than 30).

The main assumption from which our research starts is that there are collective epistemic virtues and vices, that is, that groups can be a legitimate subject for attributing epistemic traits (Fricker 2010; Bird 2014; Tollefsen 2015; Lahroodi 2007, 2019). The collective epistemic traits hypothesis implies that groups can be genuine epistemic agents and as such, in an analogous way as individual epistemic agents, susceptible to the ascription of epistemic virtues and vices. Ascriptions of epistemic traits to groups can be encountered in different aspects of human interaction. Recall numerous cases when we have heard that a group of people has displayed intellectual virtue, for example, that the jury at a trial was fair-minded or that a team of scientists displayed intellectual courage in defense of their ideas. It seems like these examples imply that when some group of people comes together they are capable of manifesting examples of epistemic virtuousness as a group.

However, according to the position called summativism (Quinton 1976; Cohen 1989), statements about group virtues should not be interpreted literally as attributing these traits to the groups themselves. Namely,

according to this view, attribution of a particular trait to a group is, in fact, an attribution of that trait to the individuals belonging to the group.[2] This means that when we say that the scientific team is intellectually courageous, we are in fact saying that every member, or at least most of the members, has the virtue of intellectual courage. Thus, when one speaks of a group's virtue, this is merely a disguised speech about the virtues of the individuals that make up the group. Summativism is essentially a reductionist position with regard to group epistemic virtues: it boils down group virtues to individual ones.[3] We believe that summativism cannot be a universally accepted position when it comes to analyzing group epistemic traits, that is, certain cases cannot be analyzed in a summative way, and reduced to the individual level. It is important to point out that there are traits that are exclusively related to the collectives and which, even in principle, cannot be reduced to individual traits because they do not have individual counterparts. Examples of such traits are the virtue of self-regulation and solidarity (Byerly & Byerly 2016). Yet, most of the traits that can be attributed to a group are such that they have individual counterparts, and the question of whether these traits should be understood summatively remains open. Denying summativism as a universally acceptable position does not imply that summativism is inadequate in all cases (Jones 2007; Fricker 2010). Understandably, there are some cases where summativism is exactly the right approach. For example, if we were to say that a family is honest, we may really want to claim that all, or most of its members, are honest.

We will not argue in detail against summativism as a universal position, since the compelling reasons against it have already been offered in the literature (for example Gilbert 1989; Fricker 2010; Lahroodi 2019). Our strategy is to cite examples that do not seem to be subjectable to summative analysis and thereby provide a *prima facie* case for non-summativism. By

[2] Summativism is formulated as a general view of attributing mental states to groups. In epistemology it has been extensively considered when it comes to attributing beliefs to groups (e.g. Wray 2007; Lackey 2014, 2020; Faria 2021). it can also be interpreted as a position on attributing epistemic virtues to groups (cf. Lahroodi 2019).

[3] In the literature, summativism is often referred to as individualism, and the term anti-individualism is used for a position that opposes it.

non-summativism, we mean the anti-reductionist view which claims that certain traits ascribed to groups are irreducibly collective and cannot be analysed as attributions of individual traits. On the background of presented examples, we will examine the notion of emergence and offer the conditions that group traits need to satisfy to be considered as irreducibly collective.

At the end of this section, it is important to introduce a distinction between faculty and character epistemic traits since both types of traits will be the subject of our consideration. This distinction is usually cited as significant in the context of understanding the differences between the two main approaches to the epistemology of virtue – reliabilism and responsibilism. Faculty virtues concern cognitive abilities or faculties such as sight, memory, and introspection. Character virtues are more sophisticated cognitive traits like open-mindedness, fair-mindedness, intellectual courage, and tenacity. Faculty and character virtues manifest differences in many significant aspects. Faculty virtues are innate or acquired in a non-reflective way, while intellectual virtues are intrinsically acquired, and their acquisition, maintenance, and exercise require reflection. Also, faculty virtues can be passive in their deliverance (e.g. seeing and hearing), unlike character virtues which are necessarily active and essentially related to the notion of good research (e.g. being observant or curious). Duncan Pritchard also emphasizes the axiological difference between the two types of virtues. Character virtues are constitutive of a productive and fulfilling intellectual life, while faculty virtues are at best just a necessary condition for it (Pritchard 2017: 7–8). It should be noted that almost all virtue epistemologists maintain that the very existence of both faculty and character epistemic virtues is not controversial, but the role of these kinds of virtues in epistemology is a matter of disagreement. Those who insist on the importance of faculty virtues for epistemology are called reliabilists (e.g. Sosa 1980, 1991; Greco 1993, 2002), while those who give central place to the importance of character virtues are considered to be responsibilists (e.g. Code 1984; Zagzebski 1996; King 2014). We assume that both types of traits can be ascribed to groups and as such are important for collective epistemology.

2. Examples of Collective Virtue Emergence

In this section, we list three hypothetical scenarios in which groups appear to exhibit irreducible collective traits. The scenarios will present situations in which it seems that some groups are able to manifest new or modified traits due to the dynamics of complex in-group interactions. Based on the analysis and comparison of these examples, we will try to figure out what would mean for a trait to be emergent. To put it differently, we will try to extract the "essence" of emergence from the examples of traits that are *prima facie* not summatively explainable. The idea behind our methodological strategy is to take a step further in the debate between summativism and non-summativism by moving away from considering individual cases and to offer general criteria to assess when a trait has an irreducibly collective character. Let us now consider the examples of emergent traits.

EXAMPLE 1 (*E*$_1$) Consider a case of two students who have a deadline to submit a joint research paper. Both manifest different epistemic vices, one of them is excessively meticulous, while the other is superficial. Initially, their work is not productive, due to the differences in their approaches which stem from the discrepancy in their epistemic characters. Since there is no initial intellectual agreement between them, compromises may occur over time. The student who is excessively meticulous will not be able to go into all the irrelevant details, because her superficial colleague is constantly flooding her with new ideas, which she cannot analyze in unnecessary depth and cannot overly devote to due to the lack of time. On the other hand, the superficial student will have to dedicate more time to concepts and problems that she would otherwise take for granted because her meticulous colleague is constantly demanding a more elaborated form of research. Only in this way, the two of them will be able to cooperate and perform the assigned task together. As a consequence of these intellectual compromises, it appears that they are both epistemically altered, and as a result of their joint work, studiousness emerges. In this situation, the interaction between the two epistemic vices, excessive meticulousness and superficiality, produced an epistemic virtue: studiousness. This is a trait that neither of the students individually had possessed before joining the group and will not

possess once separated. They manifest this epistemic virtue only when they are together and work as a team.[4]

EXAMPLE 2 (*E*₂) Let us imagine a team of 10 scientists doing research. All the researchers are of the same academic rank, are equally respected in the academic community, and have no external pressure that would affect their intellectual actions related to their research. Once the first phase of the research was completed, they published a paper that was praised and recognized by the academic community. All 10 scientists were satisfied with these results. However, after a while, they found that they have made some mistakes, which could be corrected, but that it would require some segments of their research to start over. Since the academic public did not notice these errors, the scientists could continue research and present further results as if no error had occurred. Most of the group members are in favor of continuing the research, but three of ten are against it. These three scientists want to adhere to the rules of good academic and scientific practice even at the cost of losing their current reputation and slowing down their career advancement. In that respect, they exhibit the virtue of intellectual courage because they put the truth above their personal interests. It is necessary to note that these three scientists outside this research group also manifest the virtue of intellectual courage, while on the other hand, the seven scientists who oppose admitting the mistakes are not intellectually courageous in their actions outside the group. Knowing that they have no influence on the other members in terms of seniority or any other means of coercion, the three courageous scientists set out to fix the mistakes which they all made at the beginning of the research and want to acknowledge them as their own, so the whole team wouldn't bear the blame. As time went by, other members of the group, prompted by the intellectual courage of these three members, realized that everyone should acknowledge to the public that they have made some mistakes and return to the beginning of

[4] It should be noted that the idea underlying this example – that individual epistemic vices through in-group interaction can produce a virtue at a group level – can be found in Hookway: "A research team may benefit from having some dogmatic members, and unwilling to take on board new possibilities, while others are much more ready to take seriously seemingly wild speculations. What would be vices in individual inquirers may be virtues when possessed by members of a team" (Hookway 2003, 189).

the research. After deciding to admit their mistakes at the cost of potentially losing their reputation or slowing down their career advancement, this whole team of scientists was perceived in the eyes of the academic public as an intellectually courageous team. The described episode of the collectively displayed courage by this research group did not, however, affect any of them in their actions outside the group in terms of manifesting the virtue of epistemic courage – those who were courageous continue to be so, those who were not, do not begin to show signs of this virtue.

EXAMPLE 3 (E_3) Let's consider a case where an old married couple consisting of person A and person B is instructed to remember as many details of the jointly experienced event as they can. There is a certain set of details p that person A remembers and a certain set of details q that person B remembers. By their mutual interaction, the couple comes up with a set of remembered details r, which is larger than a sum of sets of remembered details p and q together. As a result of the effort of the couple to remember as many details as they can, the mutual interaction between the two people caused their individual memories to increase, and this, in turn, affected their collective memory in the sense that it has expanded. Each one of them individually leaves this interaction with a larger set of remembered details than they had when entering the interaction.[5] A concrete example of such an interaction can be seen in the following case:

> In trying to remember the name of a film, for instance, one person might volunteer that "It begins with a B." The other might say, "Ooh, ooh, wait, wait," and then later mention that the film was a comedy with a Faustian theme. This image might help the first to recall that Dudley Moore's costar wore a red satin "devil" suit in part of the movie. Eventually, one or the other partner might finally hit on the name (Wegner et al. 1985, 262).

[5] Empirical research shows that when members of small-scale groups are remembering events they had experienced together, so-called cognitive facilitation may occur. Cognitive facilitation means that they are able to remember more details of an event or to recall more individual events when remembering together than they could recall in isolation (Meade et al. 2009; Harris et al. 2017).

The mechanism that could enable this type of increase in the individual memory within a group is the so-called transactive memory system.[6] Transactive memory systems function within smaller groups whose members are closely related, like the couple from our example (Wegner et al. 1985, 256; Wegner 1987, 192), as well as families and teams, where each member of the particular transactive memory system has knowledge about other members' memories. Their mutual interaction results in the enhancement of their own memory stores (Lahroodi 2019, 412). It could be said that a transactive memory system is a kind of mechanism that can enable the collective memory of the group members to be greater than their individual memories summatively put together (Wegner et al. 1991, 923).

Before proceeding to the analysis of examples, we need to address a potential concern that may relate to the scope of our methodological strategy. There are two potential goals our methodology can serve. The first goal is dialectical – to get a comprehensive way to deal with the summativist position in the form of clear criteria that tell us which traits are not subjectable to summative analysis. This goal is still in the domain of discussion with the summativist; it only tries to offer a strategy that is more advanced with respect to mere consideration of individual cases. The second goal is metaphysical – the criteria offered can serve as a basis for understanding the metaphysical nature of the phenomenon of collective traits even outside the context of the debate with summativism.

Of course, one can question the extent to which these two goals coincide – will we automatically reach the essence of collectivity by isolating the criteria that tell us which group characteristics are not explainable summatively? These two goals may coincide, thereby if we find which properties make a trait resistant to summative analysis, we will also find which properties make it emergent. However, the assumption of coincidence of these goals carries certain dangers. By choosing examples that are dialectically

[6] Transactive memory systems could also be considered important because some authors treat such epistemic collaborations (as well as scientific research teams) as capable of producing group knowledge that is "resistant to aggregative analyses" (Palermos 2022). This is somewhat in favor of our main hypotheses in this paper, although our main point is focused primarily on the emergent nature of epistemic virtues and vices of groups.

most potent in the debate against summativism and building criteria based on them we are in danger of raising the bar too high – to make criteria too restrictive and thereby exclude examples of traits that are emergent. In a nutshell, this problem boils down to the question of whether there are traits for which a plausible summative analysis can be offered, and they are, nevertheless, truly emergent. This problem is not easy to answer because it seems as if there may exist group traits for which it is possible to give a plausible summative analysis without those traits actually being summative. In these cases, the summative analysis, although plausible, would not be correct. Thus, by excluding all traits that can be plausibly summatively analysed, we would exclude certain truly emergent properties by our criteria. However, we believe that if we have enough information about the individual epistemic profiles of group members, the dynamics within the group, and the epistemic profile of the group itself, it is not possible to provide a *plausible* summative explanation of a trait that is truly emergent. Conversely, if we have all the needed information about the trait and we are able to provide a summative analysis of it, it is an indicator that this trait is not emergent. With this in mind, we believe that the dialectical and metaphysical goals of this paper coincide and that the properties that make a trait resist summative analysis are precisely the properties that make it truly emergent.

3. Analysis of Non-Summative Cases

In this section, we will analyze the above examples to determine the similarities and differences between them. Based on that analysis, we will explicate the criteria that a property must satisfy in order to be irreducibly collective.

In E_1 the group manifests the property that is not present in any member of the group separately. This property is genuinely new with respect to properties possessed by the individuals composing the team. If we accept that the situation conceived in E_1 is realistic and that similar situations can occur in reality, then E_1 is a strong case in favor of non-summativism. The virtue possessed by a group cannot, even in principle, be analyzed as an assertion that concerns the members of the group individually because no

member of the group individually possesses that virtue. E_1 suggests that in some cases, group-level virtue may be produced due to the interaction of individual-level vices.[7]

A significant aspect of E_1 is that the individuals who made up the team when split up no longer possess the virtue that the team manifests. This means that the team members as individuals have not been changed due to their participation in the team; in their individual activities, they continue to manifest epistemic vices as before. The ability to manifest a particular epistemic virtue depends essentially on being part of the team.

In E_2, unlike E_1, the trait exhibited by the group is not completely new with respect to the individual-level traits. At least some members possess the trait that manifests itself at the group level. Yet the virtue manifested by the group is the virtue that only a minority of individuals in the group possess. If we were to accept summativism, claims about the virtue of a group should imply that most members of that group possess that virtue, which is not the case in our example. So this example is not subjectable to a summative explanation and for this reason, we consider it an instance of non-summativism.[8] In the same way, as in E_1, the individuals that make up the team, in their individual activity outside the team, do not experience a change in their epistemic character: those who were intellectually courageous are still like that; those who were not do not become courageous. Again, the manifestation of virtue in those individuals who do not

[7] Also, it is possible to imagine a scenario in which individual-level virtues give rise to group-level vices.

[8] However, while the situation in E_2 cannot be explained in a summative way, it seems that it can still be treated in a way that reduces the specific group virtue to individual level virtues, unlike E_1. In the E_1 example, a trait that emerges at the group level is not possessed by any team member and in that sense cannot be reduced to possessing that virtue at the individual level in any way. In E_2, a trait possessed by a group can be explained in a way that involves invoking the possession of that trait by some members of the group and further explaining how those members influenced the whole group and conveyed their behavioral model to the others. Notwithstanding this difference, we believe that the example E_2, as well as E_1, can be taken into consideration as an example of non-summativism which will be further discussed in section 4.

possess that virtue outside the team depends essentially on belonging to the team.

Unlike E_1 and E_2 which deal with character virtues, E_3 is concerned with the faculty virtue. In this example, the group manifests the virtue each of its individual members possesses. This means that, as in E_2, at the group level, a completely new trait with respect to those present at the individual level does not appear. Although the group exhibits exactly the same faculty virtue (in our example, it is memory) as each of its members individually, the scenario proposed in E_3 cannot be analyzed in a summative way. Specifically, the capacity of group-level memory is much greater than the sum of individual memories of group members. The memory capacity in this example is quantified through the amount of detail one can think of. In our case, the set of details that a group can think of is much larger than the sum of the sets of details that each person can remember individually. The particular type of group interaction seems to increase individual memory capacity. However, as in E_1 and E_2, this effect is intrinsically linked to membership in the team. As soon as the person leaves the group, this effect disappears.[9]

What all three examples have in common is that the trait at the group level arises from the specific interaction of individuals who are members of the group. This situation should imply that if these individuals split up and after some time reunite and interact in a similar way the same collective trait will reappear at the group level. Of course, it is possible that persons who have been in a group and whose interaction has produced the collective trait meanwhile change in their epistemic character. This change could prevent the specific interaction that led to the group trait from recurring. However, it is important to emphasize that the appearance of the same group

[9] Of course, a person, even when she leaves the group, still has the information that she obtained during the group interaction, which means that it is subjected to some form of change. However, we think that these changes individuals are subjected to do not make the significant difference between E_3 example and E_1 and E_2 because the ability to remember better disappears when persons leave the group which means that their individual ability to remember is not permanently changed, although the opus of details that can be remembered about the event has increased during and due to the group interaction.

trait should be expected when the same people with more or less unmodified epistemic characteristics find themselves in sufficiently similar circumstances in which they should interact.

4. Criteria for Emergence

The various criteria for determining emergence have already been proposed (e.g. Wimsatt 1986; Szanto 2014; Huebner 2016). However, as far as we know, these criteria were not specifically designed to analyze the potential emergence of epistemic virtues and vices.[10] Therefore, it is not clear whether they would be applicable, and if so, how they should be interpreted to be adequately applied to collective epistemic traits. For this reason, we believe that it is useful to introduce criteria customized for collective epistemology.[11] We propose the following three criteria for determining whether a property is emergent. Each of these criteria is individually a necessary condition for emergence and together they represent a sufficient one.

(c_1) ***Novelty condition*** requires that a group must acquire a new type of property in comparison to the individual properties of its members for that property to be considered emergent. Therefore, group interaction must result in indisputable showcasing of a new property. This formulation of novelty condition can be considered as a novelty in the strong sense.

However, we can also distinguish the weak sense of novelty. The novelty in the weak sense is present when the collective property's level, intensity,

[10] The literature dealing with the formulation and application of these criteria is extensive, and there have been attempts to apply them to collective memory, collective behavior, collective intentionality, and distributed mind. Since the interpretation and application of the criteria is a very complex task with many challenges, this topic is fit to be a subject of a whole new research, and we will not deal with it in this paper. For more details see: Theiner (2013), Huebner (2014) and Szanto (2014).

[11] It is our hope that these criteria will have a broader significance and could be used to determine not only whether a collective epistemic virtue is emergent but also be applied to other kinds of group properties and processes (belief, intention, decision-making or problem-solving). But this question remains open for some future research.

or some other parameter exceeds the total sum of individual property parameters possessed by group members separately. The crucial difference between the strong and the weak senses of novelty is that the first sense implies that a completely new property emerges, while that is not the case for the second sense. Weak emergence occurs in those instances in which there is an increase in the scope of an existing property that cannot be explained in a summative way, but there is, strictly speaking, still no new property arising from the group interaction.

In a nutshell, strong emergence is a qualitative innovation with respect to the individual level, while weak emergence is a quantitative one. For the novelty condition to be satisfied, it is necessary for a trait to represent novelty either in the strong or in the weak sense. Clearly, it is not possible for one trait to satisfy both conditions at the same time. Traits that satisfy the novelty condition in the strong sense will be considered to have strong emergence, while those that satisfy the weaker form of this condition are said to be emergent in the weak sense. This distinction between weak and strong emergence on the basis of novelty conditions stems from the fact that this is the only one of the three conditions that have two forms and can thus distinguish between emergent traits.

(***c₂***) The second condition for emergence we termed the ***non-influence condition***. This condition prohibits the members of the group be influenced by the group interaction in such a way that they acquired the trait ascribed to the collective. This means that the group shows virtue only when it is together, and as soon as separated the virtue disappears. It is important to note that this is not a persistence across time condition, rather it takes into account a hypothetical situation in which we imagine what it would be like if the group was separated. Therefore, we do not consider the time in which it was separated, but a hypothetical scenario from the perspective of the time in which the members of the group are together. Meaning that this is a necessary feature of the synchronic identity of emergent traits.

If individuals, due to the in-group interaction, change in such a way that they develop the group (up to that point emergent) trait at full group-level intensity, the novelty condition would no longer be satisfied. Although the group trait would precede both ontologically and temporally the trait at the individual level, it could no longer be said to be emergent. This is

because nothing at the group level would be new with respect to individual traits nor would transcend the sum of individual traits.[12] In that way, a trait that is emergent at one point in time does not have to be emergent at some later point in time. The non-influence condition can be seen as the reverse of the novelty condition, as it specifies when the novelty condition ceases to apply.

(***c*₃**) The third and final condition for emergence is the ***non-coincidence condition*** that requires that members who have separated from a group that displays emergent (or emergently magnified) property and have lost the newly acquired property while separated will gain this property once again if they reunite with the same group. Interaction of the same group of people in similar circumstances will lead all of the group's members to once again jointly display the property that neither of them individually possesses. This condition is necessary to show that a collective trait was not a product of mere coincidence, that is, that certain properties did not develop (or that the already existing ones didn't magnify) randomly at the same time when certain individuals became part of the group. Therefore, the emergent effect must be reproducible in similar circumstances. By circumstances, in this context, we understand the epistemic character of group members and the tasks that a group should perform. Thus, the epistemic character of group members should not be significantly altered, at least not in terms of those traits that affect the formation of a collective trait that is considered potentially emergent. Likewise, when a group interacts again, it should perform a similar task in an epistemic sense, because if the task changes radically (for example, instead of writing a scientific article, they should devise a script for a short movie), it can prevent a specific

[12] It is possible that the members of the group develop a trait at the same time as, or after, the group does. The new group experiences can show them the value of this trait. Just like being exposed to individual courage can lead one to develop courage, being exposed to group courage (as a trait of a group one is part of) could lead one to develop individual courage. In such cases, the group trait would indeed be emergent but only until a member or members develop it in full intensity. Thus, our criterion does not exclude such cases completely but indicates that they can be treated as emergence only until a point in time. Thanks to the anonymous reviewer for pointing us to this question.

interaction that led to collective trait manifested in the previous episode of interaction.

We will now briefly consider whether our E_1-E_3 scenarios meet the emergence criteria. Scenario E_1 meets criterion c_1 in the strong sense. The trait displayed by the team of students is truly new in comparison to their individual traits. Conditions c_2 and c_3 are also fulfilled, the team shows virtue only when it is together, and as soon as separated the virtue disappears, which shows that they did not develop this virtue individually (c_2). Also, the same virtue should be exhibited each time the team is put together, provided that the members do not change in an epistemic sense and gather for the sake of a similar task. Under these conditions, it is to be expected that they will interact in a similar fashion as before and that the virtue of studiousness will again emerge from that interaction (c_3). In light of this, scenario E_1 can be considered as an example in which the group shows a strongly emergent trait.

When it comes to E_2, the c_1 in the strong sense is not satisfied. As for this condition in the weak sense, the situation is unclear. If the trait, in this case – the virtue of intellectual courage does not consist merely in the courage of the three members, the novelty condition in the weak sense would be satisfied. However, since we have no way of summing the intensities of intellectual courage, we cannot determine whether this condition is satisfied or not.

It should be borne in mind that we do not claim that the example E_2 does not meet our criteria for emergence, which would be quite strange given the fact that these criteria were created on the basis of three examples, including E_2 itself. However, although we intuitively think that E_2 is an example of emergence, we refrain from making that claim due to methodological difficulties. Namely, although empirical psychology offers scales that measure the intensity of courage (e.g. Woodard and Pury 2007), it is still not clear how it is possible to add the intensities of courage of different individuals. To establish that E_2 is an example of emergence we need to claim that the sum of the intensities of the individual courage of the members is less than the intensity of the courage shown by the group. Although this claim may seem intuitively acceptable, it is theoretically difficult to argue this in the absence of any idea of a plausible way to add intensities

of courage in different individuals. We hope that this methodological difficulty is only temporary and that the growing cross-fertilization of the fields of empirical psychology and epistemology of virtue will enable the development of both theoretical and empirical tools to tell us how the intensities of courage of different individuals interact.

When it comes to condition c_2, it is satisfied because both the people who were courageous and those who were not, remain the same as prior to group interaction. Additionally, it seems that the same group of scientists, should they find themselves in a similar situation, would show intellectual courage again, which would imply the fulfillment of c_3. For, as described in the E_2 scenario, the intellectual courage of the group had resulted from the interaction of members, and if the same interaction were to be repeated in a similar situation, it is to be expected that it would produce the same collective virtue. To sum up, since c_1 condition is necessary and we are not sure whether it is satisfied, despite the fact that c_2 and c_3 are satisfied, the situation as a whole seems to be undecidable with respect to the question of emergence.

In E_3 the c_1 condition is satisfied in the weak sense. The group memory capacity is greater than the sum of individual memory capacities. As in E_1, conditions, c_2 and c_3 are satisfied. The property in question in this scenario, the increased memory capacity, disappears when the transactive memory system is not together (c_2) and in the case that the members of the transactive memory system reunite, the property reappears (c_3). Since all three conditions are met, the memory capacity manifested by the transactive memory system qualifies as an example of weak emergence.

We have seen that two of our three examples meet the specified criteria. When it comes to the second example, we face a currently unsolvable methodological difficulty, due to which the status of this example is unclear. Nevertheless, we believe that our criteria, even in cases where we encounter similar methodological problems, can provide guidelines that are intuitively plausible for distinguishing cases of true emergence from those that are not.

5. Differences between Faculty and Character Virtue Concerning Emergence

After introducing the criteria for emergence, we believe that it is useful to briefly examine whether there are differences between faculty virtues and character virtues emergence. This consideration can contribute to the understanding of the differences between faculty and character epistemic virtues in their collective manifestation. The additional motivation is to indicate that a full understanding of epistemic traits and the differences between them cannot be achieved without considering their group manifestations.

We have found that there is a difference between the strong and the weak types of emergences based on a difference in the sense in which the novelty condition is satisfied. The examples we have constructed and analyzed suggest that both faculty and character virtues can be emergent. However, based on our examples, there seems to be a difference in the type of emergence (weak or strong) that faculty and character virtues may possess. The collective virtue in E_1 satisfies the conditions for emergence in the strong sense, while the collective virtue in E_3 satisfies the conditions for emergence in the weak sense. The question arises of whether there is a necessary connection between faculty virtues and emergence in the weak sense, on the one hand, and character virtues and emergence in the strong sense, on the other hand. In other words, we need to consider: 1. Is weak emergence the only form of emergence that faculty virtues can manifest or is it possible for them to be emergent in the strong sense?; 2. Can character virtues be emergent solely in the strong sense or is there a possibility that they manifest weak emergence too?

When it comes to the first question, we can claim that faculty virtues can be emergent only in the weak sense. This belief is based on the very nature of the faculty virtues. These are abilities that are mostly innate or acquired without willing control. Our ability to see or remember may of course change throughout life, become better or worse, but our possession of that ability is not something that varies in different contexts in such a way that we have this ability in one context and do not have it in another. Virtues such as memory and vision cannot be all of a sudden manifested in a group unless we as individuals possess them. As we have seen, the strong

novelty condition requires that a group manifest a whole new virtue with respect to the virtues of its members. This means that if we were to claim that some collective faculty trait is emergent in the strong sense we would have to accept that it is possible that individuals in the group do not possess some faculty at all and that this faculty can emerge through their group interaction. This would be the case if two people who do not possess the capacity of memory or vision would at once acquire these abilities through interaction within a group. Such acquisition and the manifestation of faculty virtues are not possible. However, from the E_3 scenario, which is based on insights gained through the psychological research of collective memory, we can conclude that an individual can still surpass her individual faculties during and due to group interaction.

When it comes to the second question, it seems much more difficult to give a conclusive answer. There are no theoretical obstacles for collective character virtues to be an example of emergence in the weak sense. If we recall example E_2, as we already mentioned, the team may exhibit collective courage in intensity that is greater than the sum of the individual courage intensities. However, there is one currently insurmountable methodological difficulty here. The question is how we can sum the degrees of intellectual virtue. Can the intensity of a trait in one person simply additively be built on the intensity of that trait in another person or are there some other rules by which trait intensities are combined? If we had an answer to this methodological difficulty, we could determine if E_2 is a case of emergence in the weak sense. Namely, if the intensity of intellectual courage displayed by the team is greater than the sum of the courage intensities of the three team members who possess this trait, then we could speak of emergence in the weak sense. Since we have no way of quantifying the degree or intensity of a virtue, the treatment of the E_2 scenario, and in general the question of whether collective character epistemic virtues can be emergent in the weak sense, must be left for some future research.

It is necessary to briefly explain why in the case of faculty virtues there is no parallel methodological difficulty when it comes to their quantification. Memory capacity can be quantified based on the amount of information available through it. If a person individually possesses one set of information obtained through memory, but when she is in the group has access to

another, larger set, we can say that that the person's memory capacity increased in the group. Therefore, we can compare memory capabilities by comparing the sets of information that are available through their implementation. The parallel methodology can be used for any faculty that is aimed at gathering information of any kind.

Finally, one may wonder whether the insights that we gained through examining emergence have some further significance for our intellectual self-understanding. We have already pointed out that a new faculty virtue cannot appear in a group unless it is possessed by individual members. However, certain types of group interaction, such as memberships in transactive memory systems, may make accessible the information that would not be available separately to members of such systems. In this way, certain types of interactions can be said to enhance our faculties at least as long as we belong to a particular group or system. Bearing in mind that the product of enhanced memory capacity is a larger set of information that is available to us even individually, the epistemic benefit of such interaction is evident.

However, the fact that a group can manifest virtues not possessed by its members is far more intriguing. In the first section, we agreed with Duncan Pritchard's diagnosis that character epistemic virtues are constitutive of our intellectual flourishing, that is, of a truly productive and fulfilling intellectual life. However, if we consider the possibility that we can manifest some new (in addition to the virtues we individually possess) character epistemic virtues only as part of a group, then in-group interaction may become a necessary condition for our intellectual flourishing. Additionally, there is the possibility that some of our individual epistemic virtues, which we might otherwise find useful to ourselves and to the people we interact with, in certain group contexts, contribute to collective epistemic vice. A significant suggestion arising from the fact that character epistemic traits can be emergent in the strong sense is that it is necessary to study them in a group context to be able to fully understand them.

Conclusion

In this research, we introduced the criteria for emergence of collective epistemic traits. We also differentiated between weak and strong forms of emergence. The search for criteria for emergence, as we have stated, can have two goals: dialectical and metaphysical. When it comes to the dialectical goal, our criteria provide a unified and comprehensive approach to discussion with summativism. In addition to being a useful dialectical tool in the discussion with summativism, it is our hope that the criteria we introduced, will also be the basis for a metaphysical understanding of group traits emergence in general.

References

Bird, Alexander. 2014. "When Is There a Group That Knows? Distributed Cognition, Scientific Knowledge, and the Social Epistemic Subject." In *Essays in Collective Epistemology*, edited by J. Lackey, 42–63. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199665792.003.0003

Byerly, T. Ryan. and Byerly, Meghan. 2016. "Collective Virtue." *The Journal of Value Inquiry*, 50(1): 33–50. https://doi.org/10.1007/s10790-015-9484-y

Code, Lorraine. 1984. "Toward a 'Responsibilist' Epistemology." *Philosophy and Phenomenological Research*, 45(1): 29–50. https://doi.org/10.2307/2107325

Cohen, Jonathan. 1989. "Belief and Acceptance." *Mind*, 98(391): 367–89. https://doi.org/10.1093/mind/XCVIII.391.367

Faria, Domingos. 2021. "Group Belief, Defending a Minimal Version of Summativism." *Epistemology & Philosophy of Science*, 58(1): 82–93. https://doi.org/10.5840/eps202158111

Fricker, Miranda. 2010. "Can There be Institutional Virtues?" In *Oxford Studies in Epistemology 3*, edited by T. S. Gendler and J. Hawthorne, 235–52. Oxford: Oxford University Press.

Gilbert, Margaret. 1989. *On Social Facts*. Princeton, New Jersey: Princeton University Press.

Greco, John. 1993. "Virtues and Vices of Virtue Epistemology." *Canadian Journal of Philosophy*, 23(3): 413–32. https://doi.org/10.1080/00455091.1993.10717329

Greco, John. 2002. "Virtues in Epistemology." In *Oxford Handbook of Epistemology*, edited by P. Moser, 287–312. New York: Oxford University Press. https://doi.org/10.1093/0195130057.003.0010

Harris, Celia. B., Barnier, Amanda, Sutton, John, Keil, Paul. G., and Dixon, Roger. A. 2017. "'Going Episodic': Collaborative Inhibition and Facilitation When Long-Married Couples Remember Together." *Memory*, 25(8): 1148–59. https://doi.org/10.1080/09658211.2016.1274405

Hookway, Christopher. 2003. "How to be a Virtue Epistemologist." In *Intellectual Virtue. Perspectives from Ethics and Epistemology*, edited by M. DePaul and L. Zagzebski, 183–202. Oxford: Clarendon Press. https://doi.org/10.1093/acprof:oso/9780199252732.003.0009

Huebner, Bryce. 2014. *Macrocognition: A Theory of Distributed Minds and Collective Intentionality*. New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199926275.001.0001

Huebner, Bryce. 2016. "Transactive Memory Reconstructed: Rethinking Wegner's Research Program." *Southern Journal of Philosophy* 54 (1): 48–69. https://doi.org/10.1111/sjp.12160

Jones, Todd. 2007. "Numerous Ways to Be an Open-Minded Organization: A Reply to Lahroodi." *Social Epistemology*, 21(4): 439–48. https://doi.org/10.1080/02691720701746565

King, Nathan L. 2014. "Responsibilist Virtue Epistemology: A Reply to the Situationist Challenge." *The Philosophical Quarterly*, 64(255): 243–53. https://doi.org/10.1093/pq/pqt047

Lackey, Jennifer. 2014. "Socially Extended Knowledge." *Philosophical Issues*, 24(1): 282–98. https://doi.org/10.1111/phis.12034

Lackey, Jennifer. 2020. "Group Belief: Lessons from Lies and Bullshit." *Aristotelian Society Supplementary*, 94(1): 185–208. https://doi.org/10.1093/oso/9780199656608.003.0002

Lahroodi, Reza. 2007. "Collective Epistemic Virtue." *Social Epistemology*, 21: 281–97. https://doi.org/10.1080/02691720701674122

Lahroodi, Reza. 2019. "Virtue Epistemology and Collective Epistemology." In *The Routledge Handbook of Virtue Epistemology*, edited by H. Battaly, 407–20. New York: Taylor & Francis. https://doi.org/10.4324/9781315712550-34

Meade, Michelle. L., Nokes, Timothy. J., and Morrow, Daniel. G. 2009. "Expertise Promotes Facilitation on a Collaborative Memory Task." *Memory*, 17(1): 39–48. https://doi.org/10.1080/09658210802524240

Palermos, Spyridon Orestis. 2022. "Epistemic Collaborations: Distributed Cognition and Virtue Reliabilism." *Erkenntnis*, 87(4): 1481–1500. https://doi.org/10.1007/s10670-020-00258-9

Pritchard, Duncan. 2017. "Extended Virtue Epistemology." *Inquiry*, 61(5–6): 632–47. https://doi.org/10.1080/0020174X.2017.1355842

Quinton, Anthony. 1976. "*—The Presidential Address: Social Objects." *Proceedings of the Aristotelian Society*, 75: 1–27. https://doi.org/10.1093/aristotelian/76.1.1

Ritchie, Katherine. 2020. "Social Structures and the Ontology of Social Groups." *Philosophy and Phenomenological Research*, 100(2): 402–24. https://doi.org/10.1111/phpr.12555

Schmitt, Frederick. F. 1994. "The Justification of Group Beliefs." In *Socializing Epistemology: The Social Dimensions of Knowledge*, edited by F. F. Schmitt, 257–87. Lanham: Rowman & Littlefield.

Sosa, Ernest. 1980. "The Raft and the Pyramid: Coherence versus Foundations in the Theory of Knowledge." *Midwest Studies in Philosophy*, 5(1): 3–25. https://doi.org/10.1111/j.1475-4975.1980.tb00394.x

Sosa, Ernest. 1991. *Knowledge in Perspective*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511625299

Szanto, Thomas. 2014. "How to Share a Mind: Reconsidering the Group Mind Thesis." *Phenomenology and the Cognitive Sciences* 13 (1): 99–120. https://doi.org/10.1007/s11097-013-9323-1

Theiner, Georg. 2013. "Transactive Memory Systems: A Mechanistic Analysis of Emergent Group Memory." *Review of Philosophy and Psychology*, 4(1): 65–89. https://doi.org/10.1007/s13164-012-0128-x

Tollefsen, Deborah. 2015. *Groups as Agents*. Cambridge: Polity Press.

Wegner, D. M. 1987. "Transactive Memory: A Contemporary Analysis of the Group Mind." In *Theories of Group Behavior*, edited by B. Mullen and G. R. Goethals, 185–208. New York: Springer. https://doi.org/10.1007/978-1-4612-4634-3_9

Wegner, Daniel M., Erber, Ralph., and Raymond, Paula. 1991. "Transactive Memory in Close Relationships." *Journal of Personality and Social Psychology*, 61(6): 923–29. https://doi.org/10.1037/0022-3514.61.6.923

Wegner, Daniel M., Giuliano, Toni., and Hertel, Paula. T. H. 1985. "Cognitive Interdependence in Close Relationships." In *Compatible and Incompatible Relationships*, edited by W. Ickes, 253–76. New York: Springer. https://doi.org/10.1007/978-1-4612-5044-9_12

Wimsatt, William. 1986. "Forms of Aggregativity." In *Human Nature and Natural Knowledge*, edited by A. Donagan, J. Perovich, N. Anthony, and M. V. Wedin, 259–91. Dordrecht: Springer. https://doi.org/10.1007/978-94-009-5349-9_14

Woodard, Cooper. R., and Pury, Cynthia. L. S. 2007. "The Construct of Courage: Categorization and Measurement." *Consulting Psychology Journal: Practice and Research*, 59(2): 135–47. https://doi.org/10.1037/1065-9293.59.2.135

Wray, K. Brad. 2007. "Who Has Scientific Knowledge?" *Social Epistemology*, 21(3): 337–47. https://doi.org/10.1080/02691720701674288

Zagzebski, Linda. 1996. *Virtues of the Mind*. New York: Cambridge University Press. https://doi.org/10.1017/CBO9781139174763

Call for Papers

ISSUES ON THE (IM)POSSIBLE XII
May 20-22, 2026 (Bratislava, SLOVAKIA)

Keynote Speakers

Kathrin Koslicki (University of Neuchâtel)

David Vander Laan (Westmont College)

We invite submissions for a 30-minute presentation, discussion excluded. The conference focuses on analytic philosophy. The areas of interest include, but are not limited to modal metaphysics, hyperintensionality, modal epistemology, the metaphysics of fiction, the philosophy of religion, and metaontology.

An abstract of approximately 500 words should be prepared for blind review and include a cover page with the full name, institution, and contact information. Abstracts can be submitted in pdf or doc(x) and should be sent to impossible@savba.sk

Informal queries: martin.vacek@savba.sk

Deadline for submission: **December 31, 2025**
Notification of acceptance: **February 28, 2026**

Venue: 19 Klemensova, Bratislava, Slovakia
Webpage: www.impossible.sk