Kantian Review (2026), **31**, 27–49
doi:10.1017/S1369415426101344

KANTIAN REVIEW
CAMBRIDGE UNIVERSITY PRESS

ARTICLE

What Is This Thing Called a Transcendental Argument?

Stanisław Jędrczak

Faculty of Philosophy, University of Warsaw, Poland
Email: s.jedrczak@uw.edu.pl

Abstract

Section 1 introduces the notion of transcendental arguments as arguments from the necessary conditions of possibility. What distinguishes them from other forms of deductive reasoning is the so-called transcendental conditional of the logical form '*p* is possible only if *q*'. Section 2 outlines three models for such arguments, analysing transcendental conditionals in terms of material implication, strict implication, and presupposition (semantic and pragmatic). Section 3 considers competing interpretations of Kant's Refutation of Idealism presented in the B-edition of the *Critique of Pure Reason*. Section 4 summarises the main results of the paper.

Keywords: transcendental argument; necessary condition of possibility; material implication; strict implication; presupposition; modal logic; *a priori* knowledge; Refutation of Idealism

> *'I can't explain myself, sir,' said Alice, 'because I'm not myself, you know.'*
> *'I don't know,' said the Caterpillar.* (Carroll 1998: 40–1)

1. Introduction

Let us begin with a handful of terminological remarks. A *transcendental argument* is any instance of the following pattern of inference:

(1) It is the case that *p*;
(2) *p* is possible only if *q*;
(3) therefore, it is the case that *q* (or necessarily, *q*).

This structure resembles the argument form known as *modus ponendo ponens*. Here, however, the major premise takes the form of a conditional statement asserting that some *q* is a necessary condition for the *possibility* of a given *p* rather than for *p* itself. Should *q* not be the case, *p* would not be possible. As we shall see, this seemingly minor shift – from the notion of *necessary condition* to the notion of *necessary condition of possibility* – opens a Pandora's box of profound philosophical problems.

© The Author(s), 2026. Published by Cambridge University Press on behalf of The Kantian Review. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S1369415426101344 Published online by Cambridge University Press

The starting point of a transcendental argument is the *factual premise* (1). As the name suggests, it states a certain fact: 'It is the case that *p*'. This premise is tacitly assumed to be indubitable or, to use a legal *term,* 'beyond reasonable doubt'. At the same time, *p* is considered a contingent fact (*esse contingens*) – one that might not have been the case. Furthermore, it is usually regarded as subjective, i.e., concerning the mind, and thus accessible from a first-person perspective. As Robert Stern and Tony Cheng ([2011] 2023) note in their encyclopaedia entry, 'Because of the need to find an uncontentious starting point, transcendental arguments will also then characteristically be first-personal, by beginning from how *I* or *we* experience, think, judge, and so on'. Their supposition is based on the thesis of privileged cognitive access to the mind as opposed to 'things outside us' (Bxxxix), particularly the body. The cognition of the external is always fraught with uncertainty. While this thesis might be justified in Cartesian epistemology, beyond this historical context, being a relic of mind-body dualism, it has no foundation. Therefore, the requirement that the starting point of a transcendental argument be subjective may simply be set aside.

What distinguishes transcendental arguments from other modes of deductive reasoning is the *transcendental premise* (2). Generally, a formula '*p* is possible only if *q*' is called a *transcendental conditional*, defined as a conditional with a modal operator of possibility, interpreted *de dicto*, in its antecedent. In a transcendental argument, one moves from a contingent fact, through its possibility, towards its necessary conditions. Accordingly, it might also be referred to as an argument from the necessary conditions of possibility. The notion of possibility admits of various readings: typically, it is not logical possibility (which merely amounts to the consistency of a set of propositions) but rather a metaphysical or nomological one. The relation '*x* is a necessary condition for the possibility of *y*' is hypothetical, analogous to cause and effect. Hence, each transcendental conditional calls for independent justification through a distinct line of reasoning – a *transcendental sub-argument* – which is, in principle, abductive and thus relies on defeasible rules of inference. Transcendental arguments, *sensu largo*, comprise both the main argument itself, as set out in steps (1)–(3), and the justification of premise (2).

The structure of a transcendental argument is simple, if not trivial. Nevertheless, a transcendental conditional is of particular interest, since it exemplifies an inference from the possible to the actual (*a posse ad esse)*. Yet, in line with the scholastic maxim, the consequence does not follow from the possible to the actual (*a posse ad esse consequentia non valet*). Clearly, such an inference is not warranted in any system of modal logic. For this reason, it must rely on an extra-logical principle, bridging the gap between these two different modes of existence; otherwise, it would commit the fallacy of *non sequitur*.

In their most ambitious form, transcendental arguments yield a strong conclusion – 'necessarily, *q*' – which, however, raises the question of the enthymematic premises that make the subsequent reasoning formally valid:

(1) It is the case that *p*;
(2) *p* is possible only if *q*;
(3) therefore, necessarily, *q*.

This question is thoroughly examined in Section 2.2.

https://doi.org/10.1017/S1369415426101344 Published online by Cambridge University Press

As an exercise, the reader is invited to identify transcendental conditionals in the following excerpts from Lewis Carroll's children's novels, *Alice's Adventures in Wonderland* and *Through the Looking Glass*:

(a) The executioner's argument was that *you couldn't cut off a head unless there was a body to cut it off from*: that he had never had to do such a thing before, and he wasn't going to begin at his time of life. The King's argument was that anything that had a head could be beheaded and that you weren't to talk nonsense. The Queen's argument was that, if something wasn't done about it in less than no time, she'd have everybody executed all round. (Carroll 1998: 76, emphasis added)

(b) 'I can't believe that!' said Alice. 'Can't you?' the Queen said in a pitying tone. 'Try again: draw a long breath, and shut your eyes'. Alice laughed. 'There's no use trying', she said: '*one can't believe impossible things*'. 'I daresay you haven't had much practice', said the Queen. 'When I was your age, I always did it for half-an-hour a day. Why, sometimes I've believed as many as six impossible things before breakfast. There goes the shawl again!' (Carroll 1998: 174, emphasis added)

(c) It is a very inconvenient habit of kittens (Alice had once made the remark) that, whatever you say to them, they always purr. 'If they would only purr for 'yes', and mew for 'no', or any rule of that sort', she had said, 'so that one could keep up a conservation! But *how can you talk with a person if they always say the same thing*?' On this occasion the kitten only purred: and it was impossible to guess whether it meant 'yes' or 'no'. (Carroll 1998: 238, emphasis added)

For clarity, let us paraphrase the emphasised formulas: (a) a necessary condition for the possibility of cutting off a head is that there is a body to cut it off from; (b) a necessary condition for the possibility of believing in something is that it is possible; and (c) a necessary condition for the possibility of communication with another person is that they do not always say the same thing. Each of them takes the form of a transcendental conditional: p is possible only if q. Statement (a) appears to be an analytic judgement, true by virtue of the meanings of the words 'cutting off', 'head', and 'body'. It is, therefore, a self-evident conceptual truth. Statement (b) – hastily advanced by Alice – is a bold thesis in the epistemology of modality. This thesis is false, since we are notoriously prone to modal illusions whereby certain possibilities appear impossible or, conversely, certain impossibilities appear possible, leading us to 'believe in impossible things'. By contrast, statement (c) could serve as a would-be foundation for a 'transcendental theory of communication', similar to the commitment to aiming at the truth which underpinned the work of Karl-Otto Apel and Wolfgang Kuhlmann (cf. Illies 2003: 65–7). Alice was probably unaware that she was arguing transcendentally, much like Monsieur Jourdain in Molière's comedy *Le Bourgeois Gentilhomme* was unaware that he was speaking prose. Still, these examples from Carroll's novels demonstrate that transcendental arguments are widely employed, even outside the ivory tower of academic philosophy.

As previously noted, the distinctiveness of transcendental arguments consists in their specific inference pattern, with the logical form of their major premise at its

https://doi.org/10.1017/S1369415426101344 Published online by Cambridge University Press

core. This is both a necessary and sufficient condition for an argument to qualify as transcendental. The criterion is purely formal insofar as it appeals solely to the syntactic properties of the inference pattern rather than to the semantic features of the propositional variables p and q (or the domain of their interpretation). Thus, the study of transcendental arguments is reduced to an analysis of transcendental conditionals. In this context, three fundamental questions arise:

(Syntactic Question) What is the logical form of a transcendental conditional?

(Semantic Question) What are its truth conditions?

(Epistemological Question) How do we determine the logical value of this modal judgement?

The aim of this paper is to provide convincing answers to these questions (though the Epistemological Question cannot be fully addressed *in abstracto*).

2. The logic of transcendental conditionals

As I have shown elsewhere (Jędrczak 2025), at least four models for transcendental arguments can be distinguished. In these models, transcendental conditionals are defined in terms of material implication, strict implication, semantic and pragmatic presupposition, and a unique form of abductive hypothesis. In this section, I briefly examine three of these models: material implication theory, strict implication theory, and presupposition theory. My enquiries, situated within the field of philosophical logic, serve as preliminaries to Section 3, which is devoted to Kant's Refutation of Idealism. For the purposes of this paper, abduction theory is set aside, as it does not apply to this particular transcendental argument (though it is useful in analysing other key transcendental arguments put forward in the *Critique*, such as the Transcendental Exposition of the Concept of Space and the Transcendental Deduction of the Pure Concepts of the Understanding; cf. Jędrczak 2026). I primarily consider which of the aforementioned theories might function as an adequate basis for interpreting the Kantian proof of the existence of the external world. This is not a purely historical or exegetical undertaking, as I engage with important issues in metaphysics and the epistemology of modality. That said, throughout these reflections, I follow Kant's lead as an expert guide.

2.1 Material implication theory

I adopt the traditional square ($\Box$) and diamond ($\Diamond$) symbols to denote necessity and possibility operators. These operators are mutually definable: $\Box p \equiv \neg\Diamond\neg p$. In material implication theory, the transcendental conditional is construed as a material implication – a truth-functional connective of classical propositional calculus – with a possible antecedent:

$$\Diamond p \Rightarrow q. \tag{i}$$

https://doi.org/10.1017/S1369415426101344 Published online by Cambridge University Press

In schema (1)–(3) outlined in Section 1, it is tacitly assumed that the consequence follows from the actual to the possible: $p \Rightarrow \Diamond p$. As the scholastic maxim states, *ab esse a posse consequentia valet*. This enthymematic premise is an uncontroversial theorem of system **T**, the weakest system of alethic modal logic, based on the following characteristic axiom:

(T) $\Box p \Rightarrow p$.[1]

Modal logic can be understood in a narrow and a broad sense. On the one hand, the narrow sense refers to alethic modal logic, which is, as James Garson ([2000] 2024) puts it, 'the study of the deductive behaviour of the expressions "it is necessary that" and "it is possible that"'. It comprises various axiomatic systems that determine the interaction of these operators, with each system providing what Moritz Schlick called an 'implicit definition' of the core modal notions. On the other hand, in the broad sense, in addition to alethic modal logic, modal logic includes temporal logics, epistemic logics, deontic logics, and provability logics, among other things. All these systems share a common foundation: system **K**, which consists of two basic components – the *Necessitation Rule* and the *Distribution Axiom*:

(Necessitation Rule) If p is a theorem of **K**, then so is $\Box p$.
(Distribution Axiom) $\Box(p \Rightarrow q) \Rightarrow (\Box p \Rightarrow \Box q)$.

The Distribution Axiom constitutes the 'deductive minimum' for modal logic, allowing one to derive more than just trivial substitutions from the formula $\Box(p \Rightarrow q)$. From this point on, I deal exclusively with modal logic in the narrow sense – that is, alethic modal logic. System **K** provides the foundation for all modal logics that Saul A. Kripke calls *normal*. System **T** is an extension of **K** with Axiom T, which is indispensable for grasping the intuitive notion of necessity.

According to material implication theory, transcendental arguments proceed as follows:

(1) p (assumption);
(2) $\Diamond p \Rightarrow q$ (assumption);
(3) $\Diamond p$ (from (1): *ab esse a posse consequentia valet*);
(4) q (from (2) and (3): *modus ponens*).

Here, a certain difficulty emerges: given premise (2), conclusion (4) may just as well be derived by weakening premise (1) to merely $\Diamond p$. At first glance, the above schema, though logically valid, appears to be pragmatically flawed. This presents a thorny dilemma: we must either weaken premise (1) or simplify premise (2) by dropping the possibility operator, which becomes superfluous, thus settling for nothing more than a standard *modus ponens*. However, the possibility involved in transcendental arguments should not be conflated with logical possibility, which amounts to the consistency of a set of propositions. Rather, it is a metaphysical possibility (or, as Kant puts it, a *real possibility*[2]). Indeed, the actual serves as the *ratio cognoscendi* of the possible. We lack an *a priori* criterion for metaphysical possibility; we can only know that something is possible from the fact that it is actual. In exploring the realm of

https://doi.org/10.1017/S1369415426101344 Published online by Cambridge University Press

possibility, our imagination – or conceptual intuition – regularly leads us astray. For this reason, the factual premise is not, despite appearances, redundant in transcendental arguments.

Nevertheless, argument (1)–(4) is not genuinely modal, as it does not make use of any specifically modal inference rules, such as the Distribution Axiom. In sum, system **T** appears too coarse-grained to capture the logic of transcendental conditionals.

Let us now turn to the Semantic Question: What are the truth conditions for modal judgements of the form (i)? The answer follows directly from the truth table for material implication. The logical value of this truth-functional connective depends solely on the logical values of its antecedent and consequent, regardless of their content. A transcendental conditional ($\Diamond p \Rightarrow q$) is always true except when $\Diamond p$ is true, and q is false. This gives rise to a counterintuitive result, since every necessary truth – for example, that two plus two equals four – constitutes a transcendental condition for the possibility of any state of affairs whatsoever. Should two plus two not equal four, the Earth orbiting the Sun would not be possible. Similarly, any conditional with an impossible antecedent is true. If two plus two did not equal four, then the Sun would orbit the Earth. These examples illustrate particular instances of the so-called paradoxes of material implication.

2.2 Strict implication theory

Clarence I. Lewis and Cooper H. Langford ([1932] 1959) introduced the notion of strict implication in their seminal work, *Symbolic Logic*. It is a relation of entailment characterised by the axiom systems **S**1–**S**5, arranged from the logically weakest to the strongest (where '**S**' stands for 'strict'). The conditional was defined as strict implication in light of the aforementioned paradoxes of material implication. These paradoxes – consisting in the counterintuitive results of evaluating conditionals in the classical propositional calculus – arise from treating the 'if . . . , then . . .' connective as a truth-functional operator. As a solution, Lewis and Langford sought to impose a semantic link between a conditional's antecedent and consequent, which led them to the development of a family of distinct logics of entailment. Over time, these systems proved to be effective axiomatisations of modal propositional logic, with strict implication defined as a necessitated material implication: $(p \dashrightarrow q) \equiv \Box(p \Rightarrow q)$.

In strict implication theory, the transcendental conditional is represented as follows:

$$\Box(\Diamond p \Rightarrow q). \tag{ii}$$

Philosophers, especially metaphysicians, often fail to specify the modal logic system underlying their reasoning, implicitly assuming **S**5 – the strongest system of alethic modal logic, which best aligns with our intuitive understanding of necessity and possibility. Axiom 5 admits at least three interpretations. Syntactically, it is an iteration rule: a sequence of alternating modal operators is logically equivalent to the final operator in the sequence. Semantically, it is a condition imposed on the accessibility relation between possible worlds, namely that it is *Euclidean* (meaning that if world w_1 accesses worlds w_2 and w_3, then w_2 and w_3 also access each other).

https://doi.org/10.1017/S1369415426101344 Published online by Cambridge University Press

Metaphysically, it is the thesis that there are no merely contingent possibilities.[3] In **S**5, the following proof by assumption can be carried out:

(1) p	(assumption);
(2) $\Box(\Diamond p \Rightarrow q)$	(assumption);
(3) $\Diamond p$	(from (1): *ab esse a posse consequentia valet*);
(4) $\Diamond p \Rightarrow \Box\Diamond p$	(Axiom 5);
(5) $\Box\Diamond p$	(from (3) and (4): *modus ponens*);
(6) $\Box(\Diamond p \Rightarrow q) \Rightarrow (\Box\Diamond p \Rightarrow \Box q)$	(Distribution Axiom);
(7) $\Box\Diamond p \Rightarrow \Box q$	(from (2) and (6): *modus ponens*);
(8) $\Box q$	(from (5) and (7): *modus ponens*).

Under this interpretation, transcendental arguments become full-blooded modal reasoning. Moreover, this highlights the inferential role played by the possibility operator in the transcendental premise's antecedent. Whereas the implication $\Box(p \Rightarrow q) \Rightarrow (p \Rightarrow \Box q)$ is not a theorem of **S**5, the implication $\Box(\Diamond p \Rightarrow q) \Rightarrow (p \Rightarrow \Box q)$ is.

A recurring interpretative error in formalising transcendental arguments is the omission of the modal operator. A notable example of this problem is found in the influential papers of Moltke Gram (1971) and Ross Harrison (1982). Both authors reconstruct the schema of transcendental arguments as $[(p \Rightarrow q) \Rightarrow p] \Rightarrow q$. According to Harrison, strengthening the major premise to $\Box(p \Rightarrow q)$ is superfluous. This move fails to establish the stronger conclusion $\Box q$, since in alethic modal logic implication, $\Box(p \Rightarrow q) \Rightarrow (p \Rightarrow \Box q)$ does not hold. To derive that conclusion, one would have to accept an *ad hoc* assumption $\Box p$, which is questionable, thus undermining the argument's persuasive force. The problem at hand stems from Harrison's misrepresentation of transcendental conditionals as $p \Rightarrow q$ rather than $\Diamond p \Rightarrow q$. Indeed, the inference from $\Box(p \Rightarrow q)$ and p to $\Box q$ is invalid. The foregoing analysis conclusively demonstrates the superiority of strict implication theory over material implication theory. It is only within the former that the usefulness of the transcendental argument is fully apparent, though these brief remarks, of course, do not delve into the depths of the matter under discussion.

David Lewis ([1973] 2005: 16) stipulates that (counterfactual) conditionals of the form 'If p were the case, q would be the case' are true in the actual world if and only if either (i) there are no possible worlds in which p holds – 'p-worlds', for brevity – or (ii) there is a p-world in which q holds that is *more similar* (or *closer*) to the actual world than any p-world in which q does not. Although the notion of similarity is not explicitly defined here, we presume that possible worlds are compared with reference to facts and the laws of nature. In virtue of (i), all conditionals with an impossible antecedent are, as it were, 'vacuously true' (see Lewis [1973] 2005: 24–6). Under clause (ii), while there might be p-worlds where q does not hold, they are less similar to the actual world than any p-world where q does. Let us now examine the truth conditions of *transcendental* conditionals in light of Lewis's semantics. Suppose that transcendental conditionals are expressed as 'if q were not the case, then p would not be possible' (it is worth noting that, unlike some counterfactuals, transcendental conditionals conform to the law of contraposition). However, since p is contingent, we know it is possible; moreover, given Axiom 5, p is possible in every world that is accessible from the actual one. Therefore, in **S**5, the statement 'p is possible' is

https://doi.org/10.1017/S1369415426101344 Published online by Cambridge University Press

necessarily true (what is contingent is *p* itself, not its possibility). The set of worlds in which the antecedent of a transcendental conditional is true includes all possible worlds; however, this does not imply the necessity of its consequent. There may exist *p*-worlds where *q* does not hold. Such worlds, nonetheless, would be exceedingly 'exotic'. Returning to Carroll's example, a world in which 'one could cut off a head even though there was no body to cut it off from', to paraphrase the executioner's argument, would have to differ from the actual world not only with respect to the laws of nature but also with respect to the very laws of logic.

By contrast, strict implication – formalised as $\Box(p \Rightarrow q)$ – is true in the actual world if there is no *p*-world in which *q* does not hold (Lewis [1973] 2005: 7). Its truth conditions are thus relative to the accessibility of possible worlds rather than their comparative similarity. Hence, a *p*-world where *q* is not the case cannot merely be less similar to the actual world than some *p*-world where it is the case that *q*; it must not exist at all. In material implication theory, transcendental arguments aim to show that *q* – as a necessary condition for the possibility of a contingent fact *p* – is necessary in itself (*simpliciter*). Consequently, *q* must be true in every metaphysically possible world. Even if we could conceive of a world in which *q* is false, it would constitute nothing more than a logical possibility or perhaps a fictional one.

2.3 Presupposition theory

In the two preceding sections, I examined transcendental arguments of a *positive* nature. These arguments seek to establish modal judgements concerning the mind or external world that hold in all or at least some possible worlds accessible from the actual one. Here, the modal operators are interpreted alethically, referring to metaphysical or nomological necessity and possibility. Accordingly, such arguments may be classified as *transcendental proofs*, which must be distinguished from *transcendental refutations* – the focus of this section. Transcendental refutations are *negative* – arguments not 'for' but 'against'. Their purpose is to undermine certain philosophical positions, mainly various forms of scepticism: 'The point of transcendental arguments in general is an anti-sceptical point', claims Peter F. Strawson (1985: 10). As a rule, transcendental refutations demonstrate that a given position is, in a sense, inconsistent – whether semantically or pragmatically. Upon closer examination, such refutations reveal that a position at issue entails that the conditions for its own possibility are not satisfied, where 'possibility' is understood in semiotic terms as the meaningfulness of a sentence – its truth-aptness – or the felicity of a speech act. The inconsistency of a standpoint refers to the untenability or impossibility of adopting it from the outset. Scepticism thus appears not as a genuine philosophical problem but as a pseudo-problem to be dispelled rather than resolved.[4] As Strawson (1985: 19) notes, 'The correct way with the professional sceptical doubt is not to attempt to rebut it with argument, but to point out that it is idle, unreal, a pretence'.

In line with semantic presupposition theory, transcendental conditionals take the following form:

$$p \text{ is possible only if } q \equiv p \text{ semantically presupposes } q \qquad \text{(iii)}$$

https://doi.org/10.1017/S1369415426101344 Published online by Cambridge University Press

Semantic presupposition is a relation between sentences: p presupposes q if both p entails q and $\neg p$ entails q. In short, q must hold for p to have a truth value (Levinson [1983] 1991: 175). If q does not hold, then p would be neither true nor false. Semantic presupposition is a fixed point of negation as a truth-functional connective. Let us consider the sentence 'The painting that Vincent made pleased his brother'. Both this sentence and its negation – 'It is not the case that the painting that Vincent made pleased his brother' – presuppose that Vincent made a painting.[5] Like implication, presupposition is transitive. Unlike implication, however, presupposition is irreflexive, since no sentence presupposes itself (the truth of a sentence is not a condition for its being false). Furthermore, a transitive and irreflexive relation is asymmetric, so if p presupposes q, then q does not presuppose p. When a sentence's presupposition fails to hold, the sentence becomes meaningless, lacking what Kant would call *objective validity*. Stephen C. Levinson notes that this leads to the rejection of *modus tollendo tollens*, which is accepted in classical logic (Levinson [1983] 1991: 176). Nevertheless, the weaker schema remains valid: (i) if p, then q; (ii) $\neg p$; (iii) therefore, p is either false or meaningless.

Thus, transcendental refutations demonstrate that the sentence p is meaningless, if not outright false, since its presupposition q is not satisfied. Such a line of reasoning aims to show that scepticism – regarding the existence of the external world, other minds, or free will, among other things – is 'idle, unreal, a pretence'. Scepticism emerges not as a *logical* contradiction but as a *semantic* contradiction. A semantic contradiction arises when the meaning of a sentence (or what it claims) is inconsistent with the *sine qua non* conditions for its meaningfulness. The most famous example of reasoning of this kind is Hilary Putnam's brain-in-a-vat argument. Its transcendental premise states that the sentence 'We are brains in a vat' is truth-apt only if ... we are not brains in a vat! As Putnam (1981: 15) puts it, 'If we are brains in a vat, then "We are brains in a vat" is false. So it is (necessarily) false'. The quoted sentence presupposes its own negation, constituting a semantic contradiction – a thesis that refutes itself. Here, semantic externalism embodies the principle that bridges the logical gap between the possible and the actual, whereas scepticism relies on internalism.

In pragmatic presupposition theory, transcendental conditionals state that a certain fact q is a necessary condition for the felicity of a speech act p:

$$p \text{ is possible only if } q \equiv p \text{ pragmatically presupposes } q. \tag{iv}$$

Pragmatic presupposition is a relation between a speech act, such as the performance of an utterance (like an assertion), and its context. As Andrew Cling (as cited in Bardon 2005: 75) observes, 'Semantic presuppositions of T are the necessary conditions of "T is meaningful", and the pragmatic presuppositions of T are the necessary conditions of "T is asserted"' (where 'T' stands for a theory). When a pragmatic presupposition fails to hold, the speech act becomes infelicitous despite its logical and semantic well-formedness. Pragmatic transcendental refutations aim to demonstrate that a sceptical thesis is a *pragmatic contradiction*. A pragmatic contradiction arises when the propositional content of a speech act is inconsistent with the *sine qua non* conditions for its being felicitous. In the case of assertions, the propositional content of a speech act might be true but unassertible in the given context. The paradox of a pragmatic contradiction is that for

https://doi.org/10.1017/S1369415426101344 Published online by Cambridge University Press

the speech act to be felicitous, its propositional content must be false. For example, writing the sentence 'I cannot write a single sentence in English' constitutes an act of *performative* contradiction, which consists in the dialectics of the propositional content of a speech act and the 'transcendental' condition for its being felicitous. A paradigmatic illustration of pragmatic transcendental refutation is René Descartes's *cogito ergo sum* argument from the *Second Meditation*. 'Descartes does not exist'[6] is a well-formed English sentence; as Jaakko Hintikka notes, it is no more objectionable than the 'moot' sentence 'Homer does not exist'. Nonetheless, Descartes' act of uttering this sentence is self-destructive – or 'pointless' in Hintikka's (1962: 10–4) terms. Should Descartes not exist, he could not assert a thing. Here, the principle that bridges the gap between the possible and the actual is simple: 'Thinking involves existing' (Palmer 1985: 18).

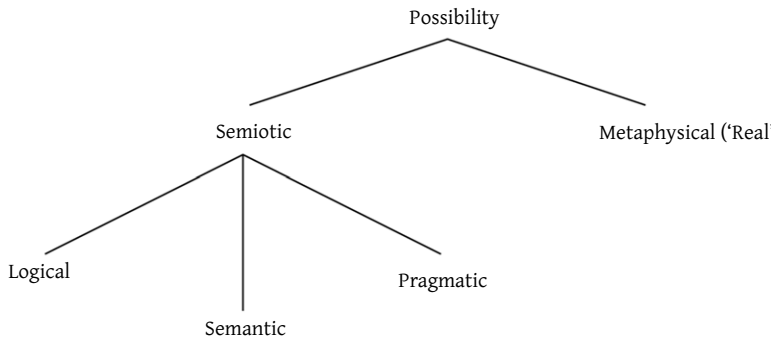


Figure 1. Notions of possibility: a classification.

Finally, I would like to make a few concluding remarks. First, it is important to distinguish between three degrees of inexpressibility in thought or language corresponding to logical, semantic, and pragmatic consistency. (I assume that whatever is inconsistent is inexpressible in at least one of these senses.) Logically and semantically well-formed sentences may still be unassertible, and sentences that are unassertible may still be true. This observation provides our modest contribution to the discussion of the 'limits of cognition', which, at least since Kant, has long captivated philosophers' imaginations. Second, we can differentiate several notions of possibility. In interpretations (i)–(ii), transcendental conditionals refer to real possibility, either of a metaphysical or nomological nature. (I presume that the notion of metaphysical possibility is broader, since the laws of nature *do not appear* metaphysically necessary.) In interpretations (iii)–(iv), transcendental conditionals refer to what I have termed 'semiotic' possibility. This notion encompasses logical as well as semantic and pragmatic possibility. This tentative classification, which supplements Kant's distinction between real and logical possibility, is illustrated in Figure 1. Third, the metaphysical and semiotic modalities can intersect. For instance, we can examine the truth-aptness of sentences and their assertibility across different possible worlds, varying with respect to the general laws that govern them, with the laws of nature being a special case. After all, the truth and assertibility conditions of sentences are not metaphysically invariant, i.e., not fundamental, unchanging aspects of reality. However, that is a matter for another discussion.

https://doi.org/10.1017/S1369415426101344 Published online by Cambridge University Press

3. Kant's proof of the external world: a case study

This section is divided into two main parts. In the first part, I examine the 'refutation of idealism' found in the A-edition of the *Critique*, in one of the passages of the Transcendental Dialectic concerning the Fourth Paralogism, of Ideality or Outer Relation (A367-80). In the second part, I address the proper Refutation of Idealism of the B-edition, laid out in the Transcendental Analytic, at the close of its second chapter, devoted to the 'system of all principles of pure understanding'. The Refutation of Idealism is based on a transcendental conditional stating that a necessary condition for the possibility of possessing temporal representations is the actual possession of perceptions of 'objects in space outside us' (B275) – and, as a result, their existence (since one cannot perceive something which does not exist). In the following pages, I analyse this bold modal judgement in light of the three theories of transcendental arguments outlined in Section 2. Although much has been written about the Refutation of Idealism in recent decades,[7] its depth remains far from exhausted.

3.1 On 'The fourth Paralogism: Of Ideality'

Kant calls *idealism* a philosophical position that claims that we have no knowledge of the external world, understood as the totality of 'objects in space outside us'. Under the classical definition, knowledge is a justified true belief. Consequently, the philosopher makes a useful distinction between two types of idealism: *dogmatic* and *problematic*. On the one hand, dogmatic idealism, represented primarily by George Berkeley, is a positive ontological position regarding the basic furniture of our world. It is, in Strawson's terms, an instance of 'revisionary metaphysics' as opposed to 'descriptive metaphysics'.[8] According to this position, ordinary beliefs about the external world, including the belief that it exists, do not constitute knowledge simply because they are false when taken literally. The concept of an external object of perception is internally contradictory, analogous to the concept of a square wheel, and therefore empty; nothing could fall under it. Objects that we mistakenly take to be material, located in space, spatially extended in three dimensions, and possessing certain 'primary qualities' are merely 'collections of ideas' in the mind. For Berkeley, to exist is to be perceived or to perceive (*esse est percipi aut percipere*). On the other hand, problematic idealism, represented primarily by Descartes, is a negative epistemological position regarding the limits of our knowledge. Here, the belief that the external world exists can never be sufficiently justified, even if it is true. In this context, *sufficiently* means 'beyond reasonable doubt'. In short, pursuant to dogmatic idealism, *we know that* the external world does not exist, while pursuant to the problematic one, *we do not know whether* it exists. As Kant (A377) remarks in the *Critique*, 'The dogmatic idealist would be one who denies the existence of matter; the sceptical idealist one who doubts them because he holds them to be unprovable'. Note that, logically speaking, these two positions are mutually exclusive: while *they cannot both be true*, they can both be false.

Kant's response to dogmatic idealism can be found in the section of the Transcendental Aesthetic devoted to space. Dogmatic idealism lapses into scepticism, which he describes as 'the euthanasia of pure reason' (A407/B434). Berkeley reduces space to an idea abstracted from sensory data, which are subjective in nature.

https://doi.org/10.1017/S1369415426101344 Published online by Cambridge University Press

Therefore, there would be no single space but as many spaces as there are cognising subjects. Berkeley is thereby unable to explain the objective validity of geometrical judgements, calling into question the scientific status of geometry – a branch of mathematics concerned with the properties of space, such as distance, shape, size, and the relative position of figures. By contrast, Kant argues that space is a form of the outer sense. In other words, the representation of space – that is, the medium in which certain geometric concepts such as point, line, or plane are constructed – is a pure intuition. Thus, geometry is a science worthy of the name, which for Kant constitutes synthetic *a priori* knowledge. It is synthetic because the representation of space is an intuition rather than a discursive concept. It is *a priori* because that intuition is pure, 'without the least empirical admixture' (A638/B666). These few sentences summarise the vindication of geometry presented in the *Critique*.

But Kant emphasises that the concept of the 'external' is ambiguous. In a relative sense, objects are called external if one is outside another and separated by a certain distance. In other words, 'outer' means 'spatially related'. In an absolute sense, however, external objects are those outside the perceiving subject itself. Space, as a form of the outer sense, is the way in which outer objects are given to us. As Kant says, 'It is itself nothing other than an inner mode of representation, in which certain perceptions are connected with one another' (A378). Metaphorically speaking, space 'is in the eye of the beholder'. It is worth recalling that this Kantian theory dates back to his Inaugural Dissertation of [1770] 1958, *De mundi sensibilis atque intelligibilis forma et principiis* (*On the Form and Principles of the Sensible and the Intelligible World*), published more than a decade before the first edition of the *Critique* and almost two decades before the second. If space is 'in us', then – with even stronger reason – the same applies to the objects in space; they are external only in a relative sense. Outer objects are those that we represent as spatially related. In the A-edition of the *Critique*, Kant equates them with objects of the outer sense, characterising objects of knowledge as 'objectively valid' representations. Yet, objective validity does not entail mind-independent existence; rather, it entails conformity with the principles of pure intellect, including, in particular, the axioms of intuition. Nevertheless, Kant's position differs from that of Berkeley in at least two key respects: first, he advocates formal, as opposed to material, idealism (the matter of cognition, affecting the senses, remains mind-independent); second, he holds that the representation of space is *a priori*, not *a posteriori*.[9]

Moreover, in the A-edition of the *Critique*, Kant addresses the issue of problematic idealism – or, more succinctly, scepticism – in the Transcendental Dialectic. The transcendental dialectic is described as the 'logic of illusion', as opposed to the transcendental analytic, which is the 'logic of truth' (A131/B170). This is telling, as Kant initially viewed scepticism not as a genuine problem to be solved but rather as an illusion of a problem to be dispelled through conceptual analysis. In the *Critique*, he reconstructs the paralogism of ideality that underlies Cartesian scepticism as follows:

> That whose existence can be inferred only as a cause of given perceptions has only a doubtful existence. Now all outer appearances are of this kind: their existence cannot be immediately perceived but can be inferred only as the cause of given perceptions. Thus, the existence of all objects of outer sense is doubtful. (A367)

https://doi.org/10.1017/S1369415426101344 Published online by Cambridge University Press

This line of reasoning is based on a pair of premises: (1) inference from effect to cause is fallible, and (2) the justification for the belief that the external world exists relies on that kind of inference. From the effect – namely, the stimulation of our outer senses – we infer the cause, that is, the existence of outer objects. Therefore, the sceptical conclusion is that (3) this belief is dubitable, since it might be false (or, to be more precise, its falsity is conceivable). Premise (1) is unquestionably true: backward reasoning is fallible, since the same effect could be caused by many different causes. This type of reasoning – best exemplified by Sherlock Holmes' *art of deduction* – yields only probable (hypothetical) and not certain (apodictic) conclusions. Kant, however, challenges premise (2). Let us stress, once again, that in the A edition of the *Critique*, outer objects are identified with objects of the outer sense. Hence, they are 'nothing other than a species of my representations' (A371). As 'a species of my representations', their reality is not inferred – that is, derived *from more fundamental premises* – but immediately perceived. This 'immediate perception' is, as Kant puts it, 'a sufficient proof of their reality' (A371). The directness of representation leaves no room for proof. There is no logical gap between the representation of outer things and their existence that could be filled by additional premises that a sceptic might call into question. Directness implies certainty, as one might put it aphoristically. Thus, proof of the external world is not so much redundant as infeasible; it is just as impossible to prove what is certain from the start as it is to climb onto a carpet using a ladder. For this reason, problematic idealism turns out to be a trivially false position, a misunderstanding, or even a grammatical error.

That said, the problem of idealism, as highlighted in the *Critique*, concerns solely the existence of outer things in a relative sense. The thesis that outer things in an absolute sense are located in space would be internally contradictory, given that space is a form of intuition.[10] Kant makes this point explicitly:

> The strictest idealist cannot demand that one prove that the object outside us (in the strict sense) corresponds to our perception. For if there were such a thing, then it still could not be represented and intuited outside us, because this would presuppose space; and reality in space, as a mere representation, is nothing other than perception itself. The real in outer appearances is thus actual only in perception and cannot be actual in any other way. (A376)

From the first review of the *Critique*, Kant was accused of phenomenalism, according to which we only know 'collections of ideas', not mind-independent objects. This situation did not change even after the publication of the *Prolegomena* in 1783. It was claimed that the position, which Kant himself compared to the Copernican Revolution in astronomy (Bxvi), was merely the 'emperor's new clothes' in which Berkeley's dogmatic idealism was disguised. There was a grain of truth in these accusations: for Berkeley, to be is to be perceived, whereas for Kant, objects are objectively valid representations, with intelligibility being their *essential* property.

3.2 On the 'Refutation of Idealism' (B275-9)

Kant's standpoint changed over time. In the second edition of the *Critique*, published in 1787, objects of knowledge come to be identified with 'things outside us', for which

https://doi.org/10.1017/S1369415426101344 Published online by Cambridge University Press

intelligibility is an *accidental* property. These are 'full-blooded' external objects in the absolute sense, located outside the thinking subject.[11] Since they are not given to us directly, their existence can be called into question (Poręba 2008: 341). Perhaps, one might wonder, objects of the outer sense appear as outer objects merely due to erroneous conceptualisation. The logical gap between representation and what is represented invites sceptical doubt. Unlike in the A-edition of the *Critique*, Kant cannot content himself with the claim that scepticism is false by virtue of the meaning of words. Instead, proof 'that the object outside us (in the strict sense) corresponds to our perception', as he phrases it, is required. That proof is precisely the Refutation of Idealism, which was added to the B edition of the *Critique*. Already in the Preface of 1787, he writes with a hint of irritation:

> No matter how innocent idealism may be held to be as regards the essential ends of metaphysics (though in fact it is not so innocent), it always remains a scandal of philosophy and universal human reason that the existence of things outside us (from which we after all get the whole matter for our cognitions, even for our inner sense) should have to be assumed merely on faith, and that if it occurs to anyone to doubt it, we should be unable to answer him with a satisfactory proof. (Bxxxix)

These statements contrast with the excerpt quoted at the end of Section 3.1. The argument from the passage on the fourth paralogism has been replaced in the B-edition with a more concise line of reasoning that excludes the phenomenalist interpretation. Henceforth, when referring to objects of the outer sense, Kant uses terms such as 'objects in space outside us', 'things outside me', or 'things that I perceive outside myself' (B275-6). This is all the more puzzling given that the doctrine of space, first formulated in the 1770 dissertation, remained intact. But now, almost two decades later, the philosopher supported it with a sophisticated argument found in the Transcendental Exposition of the Concept of Space (B41). On the one hand, according to the Transcendental Aesthetic, space is transcendentally ideal, constituting the form of the outer sense. On the other hand, according to the Refutation of Idealism, since there are 'objects in space outside us', space itself is mind-independent (or 'transcendentally real', in Kantian jargon). To resolve this apparent contradiction, we must postulate two distinct spaces: one belonging to the mind and the other existing in the external world. Here, then, we come to what has been called 'Trendelenburg's gap', referring to the famous 'neglected alternative', which stipulates that 'space could have something of a dual nature'. As Andrew Specht (2014: 521) clarifies, '[Space] could exist both as an *a priori* intuition in our minds and as an "objective form" – a structure that orders the things in themselves, the objects that exist outside of us and independently of us'.[12] For Trendelenburg, an ontological conclusion about space does not follow from epistemological premises about the representation of space (which is an *a priori* intuition). If the existence of transcendentally real space is demonstrated in the Refutation of Idealism, might there be no gap in the B-edition of the *Critique* after all? Although this is an interesting question, its exploration must be left for another occasion.

https://doi.org/10.1017/S1369415426101344 Published online by Cambridge University Press

Let us then finally consider a standardisation of the main argument of the Refutation of Idealism, which, in the *Critique*, is contained in a single paragraph of seven sentences.

(1) I am conscious of my existence as determined in time.
(2) The existence of the external world is a necessary condition for the possibility of being conscious of one's own existence as determined in time.
(3) Therefore, the external world exists.

This line of reasoning constitutes a transcendental argument against Descartes' problematic idealism. Premise (1) states that I am conscious of my persistence in time, which Kant otherwise calls 'empirically determined' existence. More specifically, I am conscious of being the same subject of many representations, occurring at different moments of time and ordered linearly by an earlier–later relation. Thus, I am able to measure the passage of time – for example, to use physical clocks. Premise (2) is a transcendental conditional, asserting that the necessary condition for the possibility of physical clocks is . . . the existence of the external world! The subsequent paragraphs explore Kant's justification for this intriguing thesis.

First, Kant points out that time measurement is relative. On the one hand – elaborating on the second analogy of experience, i.e., the principle of temporal sequence according to the law of causality – he claims that absolute time 'is no object of perception' (A200/B245). We perceive the passage of time through the succession of states of affairs or their representations (recall Aristotle's observation that time is a measure of change). There is no such thing as a 'mental clock' that indicates absolute time and is accessible to us in introspection. On the other hand, representations are not temporally indexed by themselves. 'Unlike images from sports telecasts', as Paul Guyer (1987: 244) strikingly notes, 'they have no digital timers in their corners'. The time of a given representation must be determined in reference to something external, even in a relative sense, such as other representations.

Second, as Kant specifies, time measurement is relative to 'something persistent in perception' (B276). This can be illustrated with familiar examples of physical clocks. Every clock has a stationary part and a component that moves relative to it. For example, in a sundial, the stationary gnomon casts a shadow on the face of the clock; as the sun appears to move, the shadow aligns with different hour lines, enabling time measurement. As a rule, we represent time with respect to 'something persistent in perception': a year is the time that the Earth takes to travel around the Sun; a day is the time that it takes the Earth to complete one rotation about its axis; and a second, the SI base unit of time, is defined in terms of the radiation frequency at which atoms of the element caesium change from one state to another. Importantly, this 'persistent something' must exist continuously during the measurement. The continuous existence of a clock is a necessary condition for the unity of measurement and, by extension, for the unity of time itself, as time is, by its operational definition, what clocks measure. In the Transcendental Aesthetic, Kant stresses that the representation of time is singular: 'Different times are only parts of one and the same time' (A32/B47). If the clock's existence were discontinuous, there would be a number of different 'times' (one for each continuous measurement), and the concept of 'one and the same time' would be meaningless (cf. Poręba 2008: 343). As

https://doi.org/10.1017/S1369415426101344 Published online by Cambridge University Press

a result, the succession of our representations would not form a single temporal order; different times would be incommensurable (akin to different currencies without a fixed exchange rate). Thus, 'the mere, but empirically determined, consciousness of my own existence', as Kant puts it, 'presupposes something persistent in perception' (B275). This leads us to the following question: What is this 'persistent something'?

Third, this 'persistent something' could be the subject of cognition, an internal object (that is, a representation, including an external object in a relative sense), or an external object in an absolute sense. In the course of the argument, Kant excludes the first two 'candidates' as implausible. One lesson that Kant learns from Hume is that our idea of the *self* is not based on a continuous impression. The subject persisting in time itself is not perceived. Rather, self-consciousness arises from consciousness of certain second-order representations (or, to be more precise, from consciousness of mutual relations among representations of the first order). As Kant notes in the beginning of the Transcendental Deduction of the Pure Concepts of the Understanding (in the second edition of the *Critique*):

> The empirical consciousness that accompanies different representations is by itself dispersed and without relation to the identity of the subject. The latter relation therefore does not yet come about by my accompanying each representation with consciousness, but rather by my adding one representation to the other and being conscious of their synthesis. Therefore, it is only because I can combine a manifold of given representations in one consciousness that it is possible for me to represent the identity of the consciousness in these representations itself. (B133)

Furthermore, the 'persistent something' in question cannot be identified with a representation because no representation endures as long as the subject itself. No mental state accompanies us throughout our entire existence. Consciousness functions as a container for changing content (somewhat like a hard drive that stores digital data). Thus, through this eliminative inference, only the third option remains: time measurement must be relativised to an object in space outside us. To prove that the external world exists is to prove that clocks are possible. To sum up, one could contend, in a light-hearted manner, 'Do not ask if the external world exists; ask what time it is'. Consider the standardisation of the transcendental sub-argument, outlined in the last few paragraphs:

(1) I am conscious of my existence as determined in time.
(2) Therefore, I am able to measure the passage of time.
(3) Time measurement is relative.
(4) Time measurement is relative to 'something persistent in perception'.
(5) This persisting something could be the subject of cognition, an internal object, or an external object in an absolute sense.
(6) It is neither the subject of cognition nor an internal object.
(7) 'Something persistent in perception', to which time measurement is relativised, is an external object in an absolute sense.
(8) Therefore, if clocks are possible, then the external world exists.

https://doi.org/10.1017/S1369415426101344 Published online by Cambridge University Press

To conclude, let us address the central question of this paper. Which of the theories outlined in Section 2 provides an adequate interpretation of Kant's proof of the external world as set out in the Refutation of Idealism? Note that material implication theory cannot serve this purpose, since the truth value of the transcendental conditional depends on the meanings of its antecedent and consequent. The *transcendental operator* 'it is possible that *p* only if *q*', as one might call it, is not truth-functional (unlike the connectives of the classical propositional calculus). Thus, the logic of transcendental conditionals, i.e., conditionals with a possible antecedent, is intensional – as is modal logic in general. Abduction theory also cannot serve this purpose, since abduction is a form of defeasible reasoning that leads to a mere hypothesis q that best explains a given fact *p*. The Refutation of Idealism is an anti-sceptical argument; the existence of the external world must be justified 'beyond a reasonable doubt' (apodictically certain, not hypothetically probable). As Kant himself remarks on hypotheses in general, 'Anything that even looks like a hypothesis is a forbidden commodity, which should not be put up for sale even at the lowest price but must be confiscated as soon as it is discovered' (A15). Therefore, I now examine the transcendental conditional – according to which the existence of the external world is a necessary condition for the possibility of being conscious of one's own existence as determined in time – in light of strict implication theory and presupposition theory (whether semantic or pragmatic), starting with the latter.

Instead of propositions and their truth conditions (or speech acts and their success conditions), we will refer, in accordance with Kant's terminology, to 'representations' and their 'objective validity'. In line with presupposition theory, the transcendental conditional under consideration states that temporal representations – or, in Kant's (B275) words, 'time-determinations' – presuppose perceptions of outer objects ('something persistent in perception').[13] To put it differently, the possibility of temporal representations entails actual perceptions of 'things in space outside us'. If we were lacking such perceptions – that is, if there were no outer objects (for the act of perceiving implies the existence of the perceived) – we could not make objectively valid judgements concerning the temporal order of states of affairs or their representations. In short, there are no temporal representations without perceptions of external objects in an absolute sense. Using Richard Rorty's expression, the Refutation of Idealism can be characterised as a 'parasitism argument' (Rorty 1971: 5): it reveals certain representations (namely, those of the inner sense) as 'parasitic' on others (namely, representations of the outer sense). Importantly, this is not merely a dependence of temporal representations on spatial representations in general; of course, even the simplest representation of time – for example, a timeline – is construed within a spatial medium. Rather, this dependence concerns specific spatial representations – that is, perceptions of physical bodies. In a possible world composed solely of sensory data without any material objects, time measurement would be *de re* (that is, factually) impossible. In that world, there would be no physical clocks whatsoever.

Nevertheless, in presupposition theory, rather than establish the falsity of scepticism, transcendental refutations establish its untenability. Note that the following reasoning, though formally valid, presents us with a vicious circle which renders it unsound:

https://doi.org/10.1017/S1369415426101344 Published online by Cambridge University Press

(1) I have objectively valid temporal representations.
(2) If I did not have perceptions of external things, I would not have objectively valid temporal representations.
(3) Therefore, I have perceptions of external things (and, hence, the external world exists).

As Humphrey Palmer (1985: 2) points out, 'Scepticism ... cannot be shown false by transcendental arguments, as that would make them circular. But it may be made untenable'. Given premise (2), it becomes clear that premise (1) is an implicit assumption of conclusion (3): (2) explicates (3) as a presupposition of (1). Thus, one cannot derive (3) from (1) – 'the conclusion of the argument would need to be known in advance, for the purpose of establishing the premises', summarises Palmer (1985: 1).[14] However, transcendental refutation can still demonstrate that problematic idealism is, in a sense, inconsistent (though not that it is logically contradictory). While acknowledging that we have objectively valid temporal representations, Descartes denies that the external world exists. By contrast, as Kant (B276) attempts to prove, the objective validity of temporal representations presupposes the existence of the external world: 'The consciousness of my own existence is at the same time an immediate consciousness of the existence of other things outside me'. If this presupposition is semantic, then problematic idealism turns out to be either false or meaningless. If it is pragmatic, then Descartes' position turns out to be unassertable, even if it is true. Moreover, the truth of this position would imply its unassertibility. In any case, transcendental refutation allows us to 'silence the sceptic', to use Stern's expression once again. The inconsistency of scepticism lies in the gap between its explicit claims and implicit assumptions. The aim of transcendental refutation is to bring the semantic and pragmatic commitments of scepticism to light. Strictly speaking, scepticism holds that the necessary conditions for its own objective validity are not met. For this reason, Kant asserts that

> the game that idealism plays has with greater justice been turned against it. Idealism assumed that the only immediate experience is inner experience and that from that outer things could only be inferred, but, as in any case in which one infers from given effects to determinate causes, only unreliably, since the cause of the representations that we perhaps falsely ascribe to outer things can also lie in us. (B276)

That is, the external world is not inferred from our spatial representations; it is already presupposed by our temporal representations.

One might wonder whether a stronger positive conclusion can be drawn from transcendental arguments. The transcendental conditional in question can be interpreted counterfactually: If external objects did not exist, time measurement would be *de re* impossible. Following David Lewis, we might say that this statement is true only if every possible world of purely sensory data is less similar to the actual one than any possible world 'populated' with physical bodies. This seemingly trivial condition does not exclude the existence of those possible worlds, well known from the phenomenalist fairy tale, that are governed by the principle of *esse est percipi*. Nor does it exclude the possibility of hypothetical 'mental clocks'. The only requirement is

https://doi.org/10.1017/S1369415426101344 Published online by Cambridge University Press

that these worlds operate under different general laws than those of the actual world. They would be, thus, 'exotic' possible worlds in which at least one premise of Kant's transcendental sub-argument is false. In such worlds, one would (i) perceive 'time in itself', (ii) have time-indexed representations, or (iii) possess sufficiently persistent mental states that endure throughout one's life and that are accessible to the inner sense. Put simply, one would have a different kind of mind from one's actual mind (for instance, a divine intellect of intuitive understanding). These are the truth conditions that the transcendental conditional must meet in light of the analysis in terms of the similarity of possible worlds. Strict implication theory requires, in addition, that phenomenalist possible worlds do not exist at all. The transcendental argument becomes, as it were, a philosopher's stone capable of turning base metals into gold. From a self-evident factual premise, using the **S**5 system of modal logic, one can, in a few simple steps, derive a bold modal claim concerning all possible worlds accessible from the actual one. First, time measurement is possible (because clocks actually exist). Second, for time measurement to be possible, there *must* exist material objects to which it can be relativised. Therefore, third, there are material objects in every metaphysically possible world. The existence of the external world is not only a necessary condition for the possibility of clocks; it is necessary *simpliciter*. This is precisely the point of Kantian idealism: to equate metaphysical possibility with intelligibility (for finite rational beings).

These investigations merit a brief commentary. In each case, the concept of *intelligibility* must be relativised to a certain type of mind.[15] The fact that we perceive things in a particular way is contingent (and no more essential than the fact that we breathe, digest, or reproduce as we do). In the *Critique*, Kant considers at least three types of mind to be distinct from human cognitive capacity. These are (i) the infinite divine mind, equipped with an intuitive intellect (*intellectus archetypus* or *intellectus originarius*); (ii) the Humean mind, whose operations are governed by the laws of the association of ideas; and (iii) the Cartesian mind, which has privileged and immediate access to its own mental states. In the Transcendental Doctrine of Method, Kant explicitly states that possible experience is 'something entirely contingent' (A737/B765). This claim, however, does not conflict with the assumptions underlying strict implication theory. It does not determine the modal status of the factual premise, describing – at least with respect to Kant's transcendental arguments – how our mind actually functions. Indeed, this premise typically asserts a contingent fact about the actual world. What is distinctive about transcendental arguments in strict implication theory is that they permit us to move from such a weak premise to a strong conclusion concerning all possible worlds accessible from the actual one. This transition is enabled by the transcendental conditional, which simply ascribes *de dicto* necessity to a certain conditional rather than to its antecedent itself. Both the fact that we have knowledge of the external world and the 'manner in which it is given in the mind' (B68) could be otherwise. Nevertheless, the possibility of cognising the 'objects in space outside us' in this manner is not itself contingent. Kant's transcendental question, to slightly paraphrase the main question in the Transcendental Aesthetic, is as follows: what must the external world be like for a cognition of it, such as ours, to be possible? Regardless of whether other possible worlds accessible from the actual one contain minds capable of knowledge, the cognition of these worlds remains possible for our minds (as guaranteed by the

https://doi.org/10.1017/S1369415426101344 Published online by Cambridge University Press

characteristic axiom of **S**5: 'there are no merely contingent possibilities'). Thus, in all metaphysically possible worlds, there must be certain facts in which the *real* possibility of cognition is grounded. These facts are, to put it bluntly, metaphysically invariant.

The above remarks suffice as an answer to the Semantic Question and the Syntactic Question posed in Section 1. Let us now turn to the Epistemological Question: How can we justify that a given transcendental conditional is *de dicto* necessary? According to Kant's theory of knowledge, the solution to this problem is straightforward: a judgement is necessary if and only if it is justified *a priori* (usually by way of a thought experiment). Kant equates necessary truths with truths known *a priori*. As is well known, this identity has been undermined in recent decades by the discovery of necessary truths *a posteriori* and contingent truths *a priori*. Therefore, within strict implication theory, the answer to the Epistemological Question remains open, presenting a challenge for future inquiry.

4. Conclusion

Transcendental arguments, i.e., arguments from the necessary conditions of possibility, are ubiquitous in philosophy: from ancient through early modern to contemporary thought; from metaphysics through epistemology and the philosophy of mind and language to the philosophy of mathematics and physics; and across traditions – from the analytic to the continental paradigms. I claim that what makes transcendental arguments distinctive is the so-called *transcendental premise* – or *transcendental conditional* – which states that a certain fact q is a necessary condition for the possibility of a given state of affairs p. The specificity of a transcendental argument lies in one of its premises, namely the transcendental conditional, which is a conditional with a possible antecedent. What is common to all transcendental arguments is their characteristic 'trajectory of thought' – from the possible to the actual (*a posse ad esse).* The notion of possibility, in this context, may be understood in two ways: ontologically, as the metaphysical possibility of a state of affairs; or semiotically, as the pragmatic or semantic consistency of a statement or an utterance, respectively. On the preceding pages, I roughly outlined three models for transcendental arguments: (1) material implication theory, (2) strict implication theory, and (3) presupposition theory (whether semantic or pragmatic). The above models should not be regarded as mutually exclusive. The notion of a transcendental argument does not designate a natural kind that might be grasped through a genus-differentia definition. Rather, the variety of its models illustrates the historical richness of transcendentalism as a philosophical tradition.

The main theses presented and defended in the text are as follows:

(1) Transcendental conditionals assert that a certain fact q, typically regarded as indubitable, is a necessary condition for the possibility of a given state of affairs p: 'p is possible only if q'.
(2) The possibility involved in transcendental conditionals should not be conflated with logical possibility, which amounts to the mere consistency of a set of propositions; rather, it is metaphysical possibility (or, as Kant puts it, *real possibility*).

https://doi.org/10.1017/S1369415426101344 Published online by Cambridge University Press

(3) According to strict implication theory, transcendental arguments become full-blooded modal reasoning since they rely on Axiom K, which underlies all systems of normal modal logic.
(4) Moreover, transcendental arguments turn out to be deductive within the strongest of Lewis systems, **S**5, since they assume its characteristic Axiom 5, which states – in somewhat colloquial terms – that there are no merely contingent possibilities.
(5) Kant's transcendental arguments can be adequately interpreted in strict implication theory.
(6) In the *Critique of Pure Reason*, Kant is concerned above all with the necessary conditions for the possibility of experience and, more specifically, with the necessary conditions for the possibility of making synthetic *a priori* judgements. In his view, the inference from the possible to the actual coincides with the inference from mind to world.
(7) The Refutation of Idealism admits both a weak and a strong interpretation. According to the weak interpretation formulated in the framework of presupposition theory, this transcendental argument leads to the conclusion that problematic idealism (that is, external world scepticism) is semantically or pragmatically inconsistent: not so much false as untenable. According to the strong interpretation formulated in the framework of strict implication theory, it leads to the conclusion that material objects exist out of necessity (that is, in all possible worlds accessible from the actual one). Phenomenalism, therefore, represents merely a logical possibility – insofar as it is conceivable – rather than a metaphysical or *real* one.

Financial support. This inquiry was supported by the National Science Centre, Poland (grant no. 2022/45/N/HS1/03379).

Notes

1 To be precise, the theorem $p \Rightarrow \Diamond p$ is a straightforward consequence of the mutual definability of the necessity and possibility operators and Axiom T.

2 Kant repeatedly addresses the distinction between logical and real possibility in the *Critique*, especially in the Appendix to the Transcendental Analytic titled 'On the Amphiboly of Concepts of Reflection' (A290-2/B346-9).

3 That Kant tacitly accepted this axiom is evidenced, for example, by the following passage from the Introduction to the second edition of the *Critique of Pure Reason*: 'About these sciences, since they are actually given, it can appropriately be asked how they are possible; *for that they must be possible is proved through their actuality* [emphasis mine – S. J.]' (B21). The sciences Kant refers to are, of course, pure mathematics and pure natural sciences.

4 Christian Illies (2003: 44–63) calls this type of transcendental argument 'retorsive'. As Stern (2017: 9) points out, one appeals to retorsive transcendental arguments to 'silence the sceptic'.

5 Consider, however, the well-known example: 'Bertrand has stopped beating his wife' (Levinson [1983] 1991: 174). The negation of this statement is 'It is not the case that Bertrand has stopped beating his wife.' Under a standard reading, this negation is true either if Bertrand is still beating his wife or if he never started beating his wife in the first place. Therefore, one needs to distinguish between *sentential* negation and *predicative* negation, since only the latter – 'Bertrand has not stopped beating his wife' – preserves the presupposition that Bertrand is still beating his wife.

6 Here, the name 'Descartes' is not an abbreviation of a definite description, but a rigid designator.

https://doi.org/10.1017/S1369415426101344 Published online by Cambridge University Press

7 See, in particular, Allison ([1983] 2004: 290ff); Bader (2012); Bennett (1966: 202–18); Chignell (2010); Dicker (2008); Guyer (1987: 279–329); Pereboom (2016: 455–7); Poręba (2014: 87–114); and Strawson (1966: 122–33).

8 'Descriptive metaphysics', as Strawson ([1959] 1996: 9) clarifies in the Introduction to *Individuals*, 'is content to describe the actual structure of our thought about the world, revisionary metaphysics is concerned to produce a better structure'. In place of 'our actual conceptual scheme, in which material bodies are basic' (Strawson [1959] 1996: 118), the metaphysical revisionist proposes an exotic conceptual scheme, involving sense-data as a fundamental category.

9 This point is made very clear in the *Prolegomena*: 'That space (and time as well, to which Berkeley gave no attention), together with all its determinations, can be cognised by us *a priori*, since space (as well as time) inheres in us before all perception or experience as a pure form of our sensibility and makes possible all intuition from sensibility, and hence all appearances. From this it follows: that, since truth rests upon universal and necessary laws as its criteria, for Berkeley experience could have no criteria of truth, because its appearances (according to him) had nothing underlying them *a priori*; from which it then followed that experience is nothing but sheer illusion, whereas for us space and time (in conjunction with the pure concepts of the understanding) prescribe *a priori* their law to all possible experience, which law at the same time provides the sure criterion for distinguishing truth from illusion in experience' (4: 375).

10 The external world, understood as the totality of physical objects, simply could not fit into the subject's mind.

11 In this paper – following Marcin Poręba – I adopt a realist interpretation of the Refutation of Idealism. While this interpretation is controversial, it should be noted that the reference to the Kantian Refutation only serves to illustrate the similarities and differences between various models for transcendental arguments.

12 The term 'Trendelenburg's gap' is derived from the name of German philosopher Adolf Trendelenburg, who formulated and defended the 'neglected alternative' (Specht 2014: 515).

13 This interpretation is, in a sense, supported by the German original of the *Critique*. There we read, 'Alle Zeitbestimmung setzt etwas Beharrliches in der Wahrnehmung voraus' (B275). The verb *voraussetzen* is the precise equivalent of the English *presuppose* (both, in turn, corresponding to the Latin *praesupponere*). For this reason, Paul Guyer renders the passage literally, 'All time-determination presupposes something persistent in perception' (B275).

14 For more details, see Palmer (1985: 19–22).

15 For a discussion of *relative* vs. *absolute* intelligibility, see McGinn (1989: 360–1).

References

Allison, H.E. ([1983] 2004) *Kant's Transcendental Idealism*. New Haven: Yale University Press.
Bader, R.M. (2012) 'The role of Kant's refutation of idealism'. *Archiv für Geschichte der Philosophie*, 94, 53–73.
Bardon, A. (2005) 'Performative transcendental arguments'. *Philosophia*, 33, 69–95.
Bennett, J. (1966) *Kant's Analytic*. Cambridge: Cambridge University Press.
Carroll, L. (1998) *Alice's Adventures in Wonderland and Through the Looking Glass, and What Alice Found There*. Edited with an Introduction and Notes by H. Haughton. London: Penguin Books.
Chignell, A. (2010) 'Causal refutations of idealism'. *Philosophical Quarterly*, 60, 487–507.
Dicker, G. (2008) 'Kant's refutation of idealism'. *Noûs*, 42, 80–108.
Garson, J. ([2000] 2024) 'Modal logic'. In E.N. Zalta and U. Nodelman (eds.), *The Stanford Encyclopedia of Philosophy* (Spring 2024 Edition), https://plato.stanford.edu/entries/logic-modal/.
Gram, M. (1971) 'Transcendental arguments'. *Noûs*, 5, 15–26.
Guyer, P. (1987) *Kant and the Claims of Knowledge*. Cambridge: Cambridge University Press.
Harrison, R. (1982) 'Transcendental arguments and idealism'. *Royal Institute of Philosophy Lecture Series*, 13, 211–24.
Hintikka, J. (1962) 'Cogito, ergo sum: inference or performance?'. *The Philosophical Review*, 71, 3–32.
Illies, C. (2003) *The Grounds of Ethical Judgement: New Transcendental Arguments in Moral Philosophy*. Oxford: Oxford University Press.

https://doi.org/10.1017/S1369415426101344 Published online by Cambridge University Press

Jędrczak, S. (2025) *Argumentacje Transcendentalne. Modele i zastosowania [Transcendental Arguments. Models and Applications]*. Warsaw: Scholar Publishing House.
—— (2026) 'Kantowskie argumentacje transcendentalne' ['Kant's transcendental arguments'], *Przegląd Filozoficzny – Nowa Seria*, 134, forthcoming.
Kant, I. ([1770] 1958) *De mundi sensibilis atque intelligibilis forma et principiis. Über die Form und die Prinzipien der Sinnen- und Geisteswelt [On the Form and Principles of the Sensible and the Intelligible World]*. Translated by K. Reich. Hamburg: Felix Meiner Verlag.
—— ([1781/1787] 1956) *Kritik der reinen Vernunft [Critique of Pure Reason]*. Hamburg: Felix Meiner Verlag.
—— ([1781/1787] 1998) *Critique of Pure Reason*. Translated and edited by P. Guyer and A. Wood. Cambridge: Cambridge University Press.
—— ([1783] 1997/2004) *Prolegomena to Any Future Metaphysics That Will Be Able to Come Forward as Science. With Selections from the Critique of Pure Reason*. Translated and edited by G. Hatfield. Cambridge: Cambridge University Press.
Levinson, S.C. ([1983] 1991) *Pragmatics*. Cambridge: Cambridge University Press.
Lewis, C.I., and C.H. Langford ([1932] 1959) *Symbolic Logic*. New York: Dover Publications.
Lewis, D. ([1973] 2005) *Counterfactuals*. Oxford: Blackwell Publishers.
McGinn, C. (1989) 'Can we solve the mind–body problem?'. *Mind*, 98, 349–66.
Palmer, H. (1985) *Presupposition and Transcendental Inference*. New York: St. Martin's Press.
Pereboom, D. (2016) 'Transcendental arguments'. In J. Hawthorne, H. Cappelen, and T. Gendler (eds.), *The Oxford Handbook of Philosophical Methodology* (Oxford: Oxford University Press), 444–62.
Poręba, M. (2008) *Możliwość rozumu. Ćwiczenia z metafizyki [The Possibility of Reason. Exercises in Metaphysics]*. Warsaw: Polish Scientific Publishers PWN.
—— (2014) *Granice względności. Opis metafizyczny [Limits of Relativity. A Metaphysical Account]*. Warsaw: Polish Scientific Publishers PWN.
Putnam, H. (1981) *Reason, Truth, and History*. Cambridge: Cambridge University Press.
Rorty, R. (1971) 'Verificationism and transcendental arguments'. *Noûs*, 5, 3–14.
Specht, A. (2014) 'F. A. Trendelenburg and the neglected alternative'. *British Journal for the History of Philosophy*, 22, 514–34.
Stern, R. and T. Cheng ([2011] 2023) 'Transcendental arguments'. In E.N. Zalta and U. Nodelman (eds.), *The Stanford Encyclopedia of Philosophy* (Fall 2023 Edition), https://plato.stanford.edu/entries/transcendental-arguments/.
Stern, R. (2017) 'Silencing the sceptic? The prospects for transcendental arguments in practical philosophy'. In J.P. Brune, R. Stern, and M.H. Werner (eds.), *Transcendental Arguments in Moral Theory* (Berlin: De Gruyter), 9–23.
Strawson, P.F. ([1959] 1996) *Individuals: An Essay in Descriptive Metaphysics*. London and New York: Routledge.
—— (1966) *The Bounds of Sense: An Essay on Kant's Critique of Pure Reason*. London: Methuen.
—— (1985) *Scepticism and Naturalism: Some Varieties*. New York: Columbia University Press.

Cite this article: Jędrczak, S. (2026). What Is This Thing Called a Transcendental Argument? *Kantian Review* **31**, 27–49. https://doi.org/10.1017/S1369415426101344

https://doi.org/10.1017/S1369415426101344 Published online by Cambridge University Press