

The Consequence Argument: Examining the Conditions for a Counterexample to Rule β

Matija Bajić

Povzetek

Članek proučuje problem, ki se pojavi pri poskusu formulacije protiprimena pravilu sklepanja β . Po kratkem uvodu, v katerem so predstavljeni ključni koncepti in struktura modalnega argumenta, je pozornost usmerjena na protiprimer, ki ga predstavi Johnson in McKay, pa tudi na pogoje, ki naj jim zadosti veljaven protiprimer. Preko analize predloga Erica Carlsona članek raziše možnost šibkejše interpretacije teh pogojev. V zadnjem delu se članek osredotoči na interpretativno napetost med pogojema (C1) in (C2) ter s tem izpostavi širši kontekst razprave med kompatibilizmom in inkompatibilizmom. Sklep članka je, je konstruktivna razprava zelo otežena, kolikor sta (C1) in (C2) interpretirana v močnem smislu.

Ključne besede: svobodna volja, determinizem, Peter van Inwagen, argument iz posledice, pravilo β , Eric Carlson

Summary

The article examines the problem that arises in the attempt to formulate a counterexample to inference rule β . Following a brief introduction to the key concepts and the structure of the modal argument, the focus shifts to the counterexample proposed by Johnson and McKay, as well as the conditions that a valid counterexample should satisfy. Through the analysis of Eric Carlson's proposal, the article explores the possibility of a weaker interpretation of these conditions. The final section highlights the interpretive tension between conditions (C1) and (C2), underscoring the broader context of the debate between

compatibilism and incompatibilism. The article concludes that a constructive debate is significantly hindered if (C1) and (C2) are interpreted in a strong sense.

Keywords: free will, determinism, Peter van Inwagen, the Consequence Argument, rule β , Eric Carlson

Introduction

It has been suggested that debates on the consequence argument may seem like a dead end.¹ Authors have exhausted the possibilities of logical argument analysis, thus making its strength rely on rhetorical tools. Although such a view can be justified, I believe that the logical structure of the argument deserves full attention, since inference rule β raises philosophically relevant questions. For this reason, the purpose of this paper is to identify the problems encountered in the debates on the modal version of the Consequence Argument. By presenting one of the possible lines of criticism, I point out the challenges that arise when formulating counterexamples to inference rule β .

After a brief introduction of the fundamental concepts and the structure of the modal argument, the focus shifts to the counterexample proposed by Johnson and McKay, as well as the conditions that a valid counterexample should satisfy. Through the analysis of Eric Carlson's proposal, the article explores the possibility of a weaker interpretation of these conditions. The final section highlights the interpretive tension between conditions (C1) and (C2), underscoring the broader context of the debate between compatibilism and incompatibilism. The article concludes that a constructive debate is significantly hindered if (C1) and (C2) are interpreted in a strong sense.

Determinism and Free Will

The problem with the relationship between determinism and free will mostly depends on the way these two concepts are defined. The classic definition of determinism can be formulated as follows: "From a complete description of the relevant properties of the elements of a closed system at the moment t_1 and an exhaustive listing of all the laws that apply to the elements of that system, logically follows an unambiguous and precise description of the state of that system at

1 Boran Berčić, *Filozofija*, vol. 1, p. 181.

any other moment in time, before or after t_1 .”² A similar definition was proposed by Peter van Inwagen:

We shall apply this term [determinism] to the conjunction of these two theses:

(a) For every instant of time, there is a proposition that expresses the state of the world at that instant.

(b) If A and B are propositions that express the state of the world at some instants, then conjunction of A with the laws of physics entails B .³

It should be noted that this definition does not favor one direction of time, which is consistent with the consequence of the classical definition, in which a precise description of the system follows before or after t_1 . Peter van Inwagen thoroughly explains the concepts he relies on when defining determinism.⁴ Here, it is relevant to point out that van Inwagen does not mean to say that the laws of nature in a broader sense are implied by the term “laws of physics.” Rational agents are part of nature, but psychological laws do not fall under the concept of laws of physics. In other words, psychological laws (laws of will) are not reducible to the laws of physics. Definitions of determinism that rely on a broader notion of the laws of nature leave room for compatibility with free will.⁵

The basis of the dissent between compatibilists and incompatibilists lies in their respective definitions of free will. Philosophers who believe that determinism and free will are compatible generally suggest that the absence of obstacles when acting is implicit in the concept of free will. On the other hand, incompatibilists (which include strong determinists and libertarians) define free will as the possibility to act otherwise. A philosopher of libertarian orientation will consider such an act possible, while a representative of a strong deterministic position will reject its possibility.

Van Inwagen provides a definition close to the intuitions that dictate that free will is determined by the power or ability of the subject to act differently. Therefore, we can speak of the abilities of the subject in sentences such as: “ S can make (could have made) p wrong,” where p stands for the name of proposition.⁶ Sentences about abilities in everyday speech can also be translated using this paraphrase.

2 Neven Sesardić, *Fizikalizam*, p. 118.

3 Peter van Inwagen, “The Incompatibility of Free Will and Determinism,” p. 186.

4 *Ibid.*, pp. 185–188.

5 *Ibid.*, p. 187.

6 *Ibid.*, p. 189.

For example, the sentence “ S could have written a book in three months,” can be translated as “ S could have rendered the claim that S did not write a book in three months false.”

Arguments Against Compatibilism

Consequence Argument

The Consequence Argument is the central incompatibilist argument. In his “An Essay on Free Will,” van Inwagen puts it forth in its informal form:

If determinism is true, then our acts are the consequences of the laws of nature and events in the remote past. But it is not up to us what went on before we were born, and neither is it up to us what the laws of nature are. Therefore, the consequences of these things (including our present acts) are not up to us.⁷

This formulation of the argument entails the principle of the transfer of practical necessity.⁸ This is the principle on which the entire argument rests, and its examination in a formal form will be the main task of this article. In its informal form, the principle tells us that since S cannot influence the laws of nature and the facts of the past, it follows that S cannot influence any of the consequences of that conjunction. Since the subject’s actions also belong to the set of consequences of that conjunction, S has no influence over their own actions either. Van Inwagen formalizes the consequence argument in three ways, with the first two relying on first-order logic,⁹ while the third is based on the principles of modal logic.

Modal Argument

The modal argument represents the third argument of incompatibilism.¹⁰ Before presenting the argument, it is necessary to define operator N , as well as the α and β inference rules. Operator N is defined as follows: $Np =_{\text{def}}$ “ p and no one has, or ever had a choice about p ”. The key part of the argument is represented by two rules of inference:

7 Inwagen, *An Essay on Free Will*, p. 16.

8 Berčić, *Filozofija*, vol. 1, pp. 174–177.

9 Inwagen, *An Essay on Free Will*, pp. 68–93.

10 Inwagen, *An Essay on Free Will*, pp. 93–105.

$$(\alpha) \quad \Box p \vdash Np$$

$$(\beta) \quad N(p \rightarrow q), Np \vdash Nq$$

Rule α implies that from “it is necessary that p ”, we conclude p , and no one has, or ever had, a choice about p . Rule β states that if $p \rightarrow q$ holds and no one has, or ever had, a choice about $p \rightarrow q$ and Np holds, it follows that we can conclude Nq . It is not difficult to construct an example in favor of this rule. Let’s say that no one has a choice about the fact that the state of the atmosphere at t_1 entails the soil being waterlogged at t_2 due to precipitation. Also, no one has a choice about the state of the atmosphere at t_1 . Therefore, we can conclude that the Earth is waterlogged and that no one has a choice about the waterlogging of the Earth at t_2 . It seems possible to object to and challenge this example. If a device that could absorb the entire amount of precipitation existed, the conclusion drawn would not be correct. Nevertheless, in counterexamples of this kind, the falsity of a premise is pointed out. The existence of a liquid-absorbing device indicates the falsity of the first premise, not of the reasoning principle itself. Similarly constructed examples do not present an obstacle to the acceptance of rule β . Therefore, an adequate counterexample must show that the premises of rule β ($N(p \rightarrow q)$, Np) are true, while the conclusion (Nq) is not.

Let P_0 denote the complete state of the world at the moment T_0 in a distant past. We mark the conjunction of natural laws with L , while P is the state of things at the moment T . The setting of the modal argument looks like this:

- | | | |
|----|---|------------------------------|
| 1) | $\Box((P_0 \wedge L) \rightarrow P)$ | a consequence of determinism |
| 2) | $\Box(P_0 \rightarrow (L \rightarrow P))$ | from 1), by standard logic |
| 3) | $N(P_0 \rightarrow (L \rightarrow P))$ | from 2), by rule α |
| 4) | NP_0 | premise, fixity of past |
| 5) | $N(L \rightarrow P)$ | from 3) and 4), by β |
| 6) | NL | premise, fixity of laws |
| 7) | NP | from 5) and 6), by β |

The modal argument possesses an enviable degree of logical plausibility. It seems that by accepting the initial premises, the conclusion in the form of “ P , and nobody has, or ever had a choice about P ” inevitably follows. Consequently, compatibilists do not have much room for maneuver: the criticism can be divided into

two groups.¹¹ The first tries to point out the logical invalidity of the argument, and mainly relies on the construction of counterexamples to inference rule β . The second group tries to challenge the relevance of the argument itself in the context of the discussion about the relationship between determinism and free will. Considering the complexity of the debate and the essential role of rule β in the modal version of the argument, the remaining sections of the article will be devoted to the development of the first line of criticism through the analysis of relevant counterexamples.

The Problem of Inference Rule β

Counterexamples to Rule β

There are many counterexamples to rule β in the literature.¹² One of the examples cited by Berčić reads:¹³ I cannot influence the fact that I have had an elevated level of cholesterol in my blood over the past three years. I also cannot influence the fact that high cholesterol increases the risk of a heart attack. However, I can help reduce the risk of a heart attack by going on a diet, watching what I eat, and starting a healthy lifestyle routine.

An example that gets to the heart of the problem concerning rule β is given by Johnson and McKay.¹⁴ They note that one of the logical consequences of the conjunction of inference rules α and β is the principle of agglomeration:

$$(PA) \ Np, Nq \vdash N(p \wedge q)$$

The formulation of the counterexample is now different. Since the conjunction of two van Inwagen rules implies (PA), it is necessary to construct an example that will contradict this principle. From the premises Np and Nq , a valid counterexample will derive a conclusion of the form $\neg N(p \wedge q)$. Rule α is valid, therefore, the counterexample to the principle of agglomeration will show that rule β is not valid.

11 Berčić, *Filozofija*, vol. 1, p. 174. Here, I should also mention a third group of criticism, which relies on the issue of the necessity of the past. An example of this approach can be found in Joseph K. Campbell, "Free Will and the Necessity of the Past," pp. 105–111.

12 See David Widerker, "On an Argument for Incompatibilism," pp. 37–41; Eric Carlson, "Counterexamples to Principle Beta: A Response to Crisp and Warfield," p. 731; David Johnson and Thomas McKay, "A Reconsideration of an Argument against Compatibilism," pp. 115–116. Additional interesting constructions are provided by Berčić, *Filozofija*, vol. 1, pp. 175–176, and Sesardić, *Fizikalizam*, pp. 141–143.

13 Berčić, *Filozofija*, vol. 1, pp. 175–176.

14 Johnson and McKay, "A Reconsideration of an Argument," p. 115.

Johnson and McKay provide the following example.¹⁵ Suppose I did not flip a coin, but I could have. Let $p =_{\text{def.}}$ “the coin does not land on heads” and $q =_{\text{def.}}$ “the coin does not land on tails.” Therefore, both premises of agglomeration are true: no one can choose to make the coin land on heads (it can also land on tails), so Np , and similarly, no one can choose to make the coin land on tails, so Nq . However, $N(p \wedge q)$ does not follow from Np and Nq ! While it is true that no one has a choice about these statements taken individually, the truth of their conjunction is not necessarily beyond the subject’s control. By choosing to flip the coin, the agent would make the coin land either on heads or tails, thus making $N(p \wedge q)$ false. Therefore, the principle of agglomeration is not valid. It is a consequence of the conjunction of rules α and β , and since rule α is valid, rule β must be invalid.

Conditions for the Formulation of Counterexamples

Johnson and McKay provide a strong counterexample to the inference rule β . However, it seems that there are reasons that call their scenario into question. An adequate example should fulfill three relevant conditions found in the literature. The first condition was stated by van Inwagen:¹⁶

(C1) A counterexample to rule β must not assume the compatibility of determinism and free will.

The assumption of an indeterministic world is another inadequate way of arguing against rule β . The reason lies in the fact that incompatibilists can easily defend themselves against this type of counterexample. In this case, the incompatibilist can introduce a replacement for rule β in the form of rule δ :¹⁷

(δ) $D, N(p \rightarrow q), Np \vdash Nq$, where “D” stands for determinism.

Therefore, a second condition is necessary. It is implicitly present in Johnson and McKay and explicitly articulated by Crisp and Warfield:¹⁸

(C2) A valid counterexample to rule β must not assume the truth of indeterminism.

¹⁵ *Ibid.*, pp. 115–116.

¹⁶ Inwagen, *An Essay on Free Will*, p. 102.

¹⁷ Johnson and McKay, “A Reconsideration of an Argument,” p. 118.

¹⁸ Thomas M. Crisp and Ted A. Warfield, “The Irrelevance of Indeterministic Counterexamples to Principle Beta,” pp. 179–180.

A modification of rule β in the form of (β') is also present in the literature. This modification assumes the role of the third condition. Timothy O'Connor makes the following demand:¹⁹

(C3) (β') $N(p \rightarrow q), Np \vdash Nq$, for every p and q such that q is made true later than p .

The question arises: Do the previously presented examples meet these three conditions? Crisp and Warfield present Johnson and McKay with a dilemma.²⁰ Coin tossing takes place either within a deterministic world or within an indeterministic world. If tossing a coin is an act in an indeterministic world, then clearly (C2) is violated. Conversely, if the example is set in a deterministic world, the possibility of acting differently (i.e., flipping a coin) leads to the assumption of compatibilism and the violation of (C1).

We can similarly criticize Berčić's example. Does the possibility of my lifestyle change take place in a deterministic or nondeterministic world? If it is a deterministic world, then this example violates (C1). More precisely, either it assumes, through the possibility of taking different actions, that free will is compatible with determinism (which needs to be proven), or it assumes that psychological laws are subsumed under deterministic laws on which it bases the explanation for the possibility of an agent's actions within the deterministic world. The first disjunct does not fulfill (C1), while the second does not accept the initial definition of determinism from which psychological laws are excluded. On the other hand, if the example takes place in an indeterministic world, then (C2) is not fulfilled.

Is a Counterexample Possible?

It is clear that the set conditions significantly shape the construction of counterexamples. The task of formulating an example can also be set in a formal form:

(C1 \wedge C2 \wedge C3 \wedge A) $\rightarrow \neg Nq$, where A denotes the premises of β , i.e., $Np, N(p \rightarrow q)$

While analyzing this task further, I will rely on Carlson's example.²¹ In his counterexample to rule β , Carlson constructs the following scenario. At t_0 , subject S may or may not press a button that is connected to a coin-tossing

19 Timothy O'Connor, "On the Transfer of Necessity," p. 209.

20 Crisp and Warfield, "The Irrelevance of Indeterministic Counterexamples," p. 180.

21 Carlson, "Counterexamples to Principle Beta," p. 731.

machine. If S presses the button, the machine flips the coin twice, at t_1 and at t_2 . If S does not press the button, the machine flips the coin only once, at t_1 . Suppose the subject presses the button and the coin lands on heads both times. Therefore, $p =_{\text{def.}}$ "A coin is tossed at t_1 and lands on heads" and $q =_{\text{def.}}$ "A coin is tossed at t_2 and lands on heads." Even if S had not pressed the button at t_0 , the machine would have tossed a coin at t_1 and the coin could have landed on heads; therefore, S has no influence in making p false. So, Np is true. On the other hand, if S had not pressed the button, the coin could have landed on tails at t_1 . If this were the case, the implication $p \rightarrow q$ would be true, since its antecedent would be false, therefore making $N(p \rightarrow q)$ true. However, Nq is not true. If S had not pressed the button, the coin would not have been tossed at t_2 , and q would not be true.

The essence of Carlson's example lies in providing a different interpretation of the conditions. We have seen Crisp and Warfield criticize the example presented by Johnson and McKay, facing them with a dilemma. Carlson argues that the dilemma is false. By implication, Carlson offers a weaker reading of the terms. He takes the first condition to mean that the example could take place in a world where compatibilism is false. Similarly, (C2) requires that the counterexample be formulated so that it could take place in a deterministic world, not that it does. The reason for this interpretation of the conditions is an attempt to formulate counterexamples in such a way that they do not presuppose a specific metaphysical position on the nature of the world. In this sense, the wording "could be" leaves room for neutrality concerning the question of the metaphysical properties of the world. Crisp and Warfield's mistake, Carlson believes, is their assumption that the example takes place in exactly one world which is either deterministic or indeterministic.²² As the example could occur in several possible worlds (either deterministic or indeterministic), there is no reason to assume that they are all deterministic or that they are all indeterministic.

Let us see how Carlson's example reacts to the set conditions. If it takes place in a deterministic world, then it satisfies (C2). However, it also fulfills (C1) since it could also occur in a world where compatibilism is false. On the other hand, if it takes place in an indeterministic world, then it fulfills (C1). As in the previous case, the second condition is also fulfilled because it could also occur in a deterministic world. The third condition is not a problem, since q is made true later than p .

22 *Ibid.*, p. 733.

Interpretive Tension of Conditions (C1) and (C2)

Carlson's argument points to a problem that arises when trying to formulate a counterexample to the inference rule β . I have already mentioned that Carlson implicitly raises the question of the interpretation of the set conditions. The initial requirement is that the example be restricted to one world, which is deterministic (as dictated by (C2)) and in which there is no assumption of compatibilism (C1). However, the problem with the interpretation of the first condition can already be seen in the dilemma that Crisp and Warfield present Johnson and McKay with. As a reminder, if the example of tossing a coin is part of a deterministic world, the subject's ability to do this is interpreted as a compatibilist assumption. Therefore, (C1) is not fulfilled.

Such an interpretation of the conditions does not leave much room for formulating counterexamples. Under the interpretation whereby the example takes place in exactly one world, it is permissible to assume determinism (C2) and incompatibilism (C1). However, under these assumptions, as Carlson correctly notes, rule β is trivial!²³ If one starts from determinism and incompatibilism, then subjects by definition do not have the possibility of alternative actions, and the outcome of rule β is no longer subject to proof but a direct consequence of the assumed framework.

I would add that the strongly interpreted conditions implicitly set tasks that go beyond the basic goals of the compatibilist position. It is about proving freedom from the assumption of determinism. Therefore, I believe that the discussion should take a step back and reconsider its goals. The compatibilist's task in the context of a broader discussion is to point to a possible world in which determinism and free will are true. In his first formal argument against compatibilism, van Inwagen, under the assumptions of determinism and freedom, points to their incompatibility. Using the same assumptions, a compatibilist should point out the compatibility of the two phenomena. Therefore, Carlson's reading of the conditions is a persuasive first step toward the possibility of further confronting the arguments.

Bibliography

- Berčić, Boran. *Filozofija*. Vol. 1. Zagreb: Ibis grafika, 2012.
Campbell, Joseph K. "Free Will and the Necessity of the Past." *Analysis* 67, no. 2 (2007): pp. 105–111.

23 *Ibid.*, p. 734.

- Carlson, Erik. "Counterexamples to Principle Beta: A Response to Crisp and Warfield." *Philosophy and Phenomenological Research* 66, no. 3 (2003): pp. 730–737.
- Crisp, Thomas M., and Ted A. Warfield. "The Irrelevance of Indeterministic Counterexamples to Principle Beta." *Philosophy and Phenomenological Research* 61, no. 1 (2000): pp. 173–184.
- Inwagen, Peter van. *An Essay on Free Will*. Oxford: Clarendon Press, 1983.
- . "The Incompatibility of Free Will and Determinism." *Philosophical Studies* 27, no. 3 (1975): pp. 185–199.
- Johnson, David, and Thomas J. McKay. "A Reconsideration of an Argument against Compatibilism." *Philosophical Topics* 24, no. 2 (1996): pp. 113–122.
- O'Connor, Timothy. "On the Transfer of Necessity." *Noûs* 27, no. 2 (1993): pp. 204–218.
- Sesardić, Neven. *Fizikalizam*. Belgrade: Istraživačko-izdavački centar SSO Srbije, 1984.
- Widerker, David. "On an Argument for Incompatibilism." *Analysis* 47, no. 1 (1987): pp. 37–41.