

Rearguing Consequential Questions: A Reply to Gustafsson

Christopher Taylor and Daniel Dennett

Several months before the publication of “Dennett and Taylor’s Alleged Refutation of the Consequence Argument,” Johan Gustafsson was kind enough to send us an advance copy of that article, and the resulting correspondence was, in our opinion, quite entertaining and instructive. Since, regrettably, few of the ideas that came up in this exchange made it into the final form of Gustafsson’s paper, we propose to explore them here.

Gustafsson’s attempt to resuscitate the Consequence Argument centers on the following claim: the version of the Argument against which we directed the appendices in the two editions of “Who’s Afraid of Determinism” misrepresented the real, “modern” form of said Argument, and that the latter is not subject to the objections we raised. Now the exegetical question, about whether our account captured the thinking of Van Inwagen and others accurately, is not in fact one that we intend to spend time debating here; even if we grant that the reasoning in “Who’s Afraid of Determinism” rested on a misunderstanding, it seems likely to us that others who encounter the Consequence Argument may well share our (mis)construal of it, and that it was worthwhile to expose the logical error upon which this version of the Argument depends. At the least, our debunking exercise reveals unmistakably the dangers that sloppy treatment of notions related to causation and counterfactuals may entail.

In all events, the first task when assessing Gustafsson’s article is undoubtedly to examine the differences between his interpretation of the Argument and the one we originally criticized. The crux of the distinction, according to Gustafsson, involves the Argument’s fourth line, where he contrasts a deductive process called “The Power-Transfer Rule” (the rule that Taylor/Dennett refuted) with a more elaborate variant that he dubs “Rule Gamma”:

(The Power-Transfer Rule) From $\Box(\alpha \text{ occurs} \supset \beta \text{ occurs})$, and that some person S has the power to cause α , deduce that S has the power to cause β .

(Rule Gamma) From $\Box(p \supset q)$ and that some person S has the power to cause some event α such that, if S were to cause α , then p *would* be true, deduce that some person S' has the power to cause some event β such that, if S' were to cause β , then q *would* be true.

The first of these rival statements seems relatively straightforward, though we should note that even in this case it required nontrivial effort, while writing “Who’s Afraid of Determinism,” to reduce it into the rigorous language of possible worlds; but by contrast, we consider “Rule Gamma” as Gustafsson presents it a treacherously obscure statement, more apt to trip readers up than to resolve philosophical difficulties. Leaving aside some minor peculiarities (such as his penchant for italicizing the word “would”), the fundamental trouble we see stems from the way that Gustafsson stirs together with such abandon not only counterfactuals but concepts like “power” and “causation,” to produce in the end a complex brew whose bewildering chemistry never receives proper analysis.

In light of these problems, we encouraged Gustafsson over the course of our correspondence to adopt the approach advocated from the outset by Taylor/Dennett, and develop a more explicit formulation of Gamma in terms of possible worlds. As a preliminary step, we suggested that he attempt to clarify the following sentence:

$Z = \text{if } S \text{ were to cause } \alpha, p \text{ would be true.}$ (1)

This task did not seem to be one that Gustafsson viewed with much enthusiasm, but when one of us (Taylor) came up with the following proposal, he appeared amenable to it:

$Z \equiv$ there exists a sentence γ (describing an action of S) and a universe f fairly close to actuality, such that γ obtains in f and is a necessary condition for α in all the universes most similar to f — i.e., $\alpha \supset \gamma$ throughout f 's immediate neighborhood; furthermore, $\gamma \supset p$ for all universes within some suitable set X of universes similar to actuality¹ (2)

Now this resolution for Z undeniably raises several issues, most relating to the vague way in which the various sets of universes have been specified — for example, it remains unclear whether any special relationship is supposed to exist between X and the previously mentioned set of universes in f 's neighborhood.² Readers of “Who’s Afraid of Determinism” will recognize quandaries of this sort as an inevitable byproduct of all attempts to formalize counterfactual and causal statements. As users of this method, however, we have never viewed such difficulties as

¹Those unaccustomed to this manner of speaking are referred to “Who’s Afraid of Determinism” for a tutorial in the idiom. Notice that here, as in our original appendix, we have simplified the notion “event γ causes event α ,” extracting from this concept its one most critical ingredient, counterfactual necessity.

²In what follows, we do not bother to assume any particular relationship between these sets of universes. Also left open is the question, whether we should take into account the likelihood that there exist other ways for S to “cause α ,” distinct from action γ .

flaws in the system; instead, we believe that in passing from statement (1) to statement (2), we have succeeded in bringing the vagueness and ambiguity inherent in the former out into the open. While still bearing these complications in mind, we can proceed by setting them provisionally to one side and advancing into the next phase of the discussion.

The above resolution of Z has for Gustafsson the pleasant result that it confirms the logical truth of Rule Gamma. For suppose that S has the power to cause some α , where “if S were to cause α , p would be true”; reformulating as above, there exists an action of S, γ , which (1) is a necessary condition for α in the cluster surrounding nearby universe f , and (2) entails p in set-of-nearby-universes X. If we now postulate that $p \supset q$ in every universe, then obviously $\gamma \supset q$ within X; translating back, using once more the causal connection between γ and α , we obtain “if S were to cause α , q would be true.” And in that way, we find that the consequent clause in Rule Gamma has been satisfied.³

When we attempt to diagnose why Rule Gamma succeeds where the Power-Transfer Rule failed, one of the first points we notice is the quite marginal position that α occupies in the preceding paragraph. The starring role really belongs to the action γ , whose relationship with sentence p, one of counterfactual entailment, is something that can be automatically transferred to sentence q as well. (The contrast with the causal relationships found in the Power-Transfer Rule is clear: with the latter, as Taylor/Dennett originally demonstrated, the arrows of counterfactual implication basically point in the wrong directions, blocking one’s ability to make valid inferences.) Meanwhile, whatever bond γ enjoys with α has no real impact, neither on the

³We choose α to assume the role of β in Gustafsson’s formula, and S to assume the role of S’. These surplus symbols then mercifully drop out of the story for good.

current nor on the forthcoming discussion. And thus we begin to discern Gustafsson's basic recipe for revitalizing the Consequence Argument: although his formulation of Rule Gamma mentions causation, he is effectively shelving that concept and is assembling a concoction whose dominant ingredient is a simpler type of counterfactual implication.

Accordingly, it appears that one can streamline Gustafsson's variant of the Consequence Argument as follows, without, as far as we can see, losing anything of significance:

- (1) $\Box(P_0 \ \& \ L \supset P)$, where P_0 represents a complete specification of the initial state of the universe, L stands for some imagined deterministic laws, and P represents an event in the life of some agent S that actually occurred — this statement holds, simply by virtue of the meaning of “determinism”
- (2) $\Box(\sim P \supset \sim(P_0 \ \& \ L))$ — the contrapositive of (1)
- (3) There exists a sentence γ , describing an action by S , that obtains in some possible world close to actuality, and such that $\gamma \supset \sim P$ throughout a suitable set X of universes close to actuality — this is a hypothesis, the first step in a proof by contradiction
- (4) $\gamma \supset \sim(P_0 \ \& \ L)$ throughout X — follows immediately from (2) and (3), using the same reasoning by which we validated “Rule Gamma” previously
- (5) But there cannot exist any sentence describing a possible action of S that implies the falsity of $(P_0 \ \& \ L)$ throughout X — this is a premiss, based according to Gustafsson on the “Fixity of the Past and the Laws”
- (6) Therefore hypothesis (3) is false — from which we are to draw the intuition that, when laws are deterministic, S never has the power to bring about the falsity of any events that

actually occur, or the truth of any non-actual events.

When we phrase matters this way, it becomes particularly obvious that Gustafsson's entire edifice rests upon the premiss in (5) — an altogether wobbly foundation, as we shall see. The weakness in this line represents the cost that he must pay for strengthening (4).

If Gustafsson's premiss (5) seems initially plausible (“*surely* no action I could possibly take would imply either the falsity of the laws of nature, or a history of the Big Bang different from actuality!”), one should pause a moment to consider a distinction that many people may well overlook in their pursuit of a quick intuitive verdict:

(Claim A) Agent S's possible action γ counterfactually implies (within set of universes X) the falsity of the laws of nature, or the falsity of the Big Bang's actual history — this is identical to line (4) above.

(Claim B) Agent S's possible action γ would *cause* the falsity of the laws of nature, or the falsity of the Big Bang's actual history.

Claim B is one from which our intuitions immediately recoil, and indeed if we analyze it carefully using the Taylor/Dennett methods we find this distaste well-justified.⁴ And yet, however fishy the superficial similarity to B may make it appear, Claim A occupies a far superior position. A number of arguments not only lend support to claims of this sort, but make them appear logically invincible, the simplest of which just refers us back to line (2) above: given any

⁴We leave the analysis as an exercise for the interested reader.

sentence whatsoever referring to a non-actual event (including any hypothetical action by S), a universe in which that event occurs *necessarily* contains either a violation somewhere in its history of its postulated deterministic laws, or else has a beginning different from actuality's. This simple observation makes Gustafsson's rejection of (A) — which is to say, his premiss (5) — look increasingly untenable.

In his article and in our correspondence, Gustafsson tried a couple of strategies in defense of (5). A bit of time was spent debating the “Fixity of the Past,” which in his words rests on the notion that “we try to hold the remote past fixed when we evaluate counterfactuals.” To put it mildly, this “rule” needs considerable refinement if it is to become intellectually respectable. The most glaring gap concerns how we are supposed to apply it when we are dealing with sentences whose antecedents or consequents themselves talk hypothetically about the remote past, a category that includes most of the counterfactuals recently considered. Is Gustafsson really prepared to defy the last paragraph's logic and dismiss them as false (or perhaps meaningless) *en masse*? Or are exceptions made to his rule in these instances? His article makes no effort to clarify. If in fact the “Fixity of the Past” principle is allowed to have exceptions in such cases, then notice that its power to refute line (4) in the above argument collapses.

Even when we restrict the “Fixity of the Past” idea to more ordinary counterfactuals, it encounters grave difficulties. Picture yourself engaged in some everyday counterfactual exercise, like trying to determine what the world would be like if your mother had fallen ill on the day of her first date with your father. Quite likely you will be focusing your attention on hypothetical events in the months and years following that day, and as the calendar advances in your imagined world the scenario becomes ever more speculative, with differences relative to actuality

presumably tending to grow. If pressed, you might also decide to extend your exercise in the opposite direction, imagining time periods prior to the fateful date. As you run the clock backwards, chances are you will be envisioning a world that gets ever closer to reality (one week before the date, you might alter the positions of a few bacteria; one week before that, just a few DNA molecules...) While pursuing this procedure all the way back to the Big Bang is obviously something only a philosopher would attempt (at that point, you might be dealing with a situation where, say, a single electron has been shifted over by 10^{-1000} meters), we see no logical absurdity in doing so.⁵ We see no principled reason to choose some particular date in the past, and insist that, prior to that point, only universes *identical* to reality will be considered; it is particularly unlikely that ordinary language users evaluating sentences like “If my mother had gotten sick, I would not have been born” adhere with any consistency to such a principle, even though thinkers like J.L. Austin sometimes liked to imagine that they do. (The case of Austin, along with his notorious missed putt, receives lengthy attention in “Who’s Afraid of Determinism”; this paragraph echoes much of that discussion, though presented now in a somewhat novel form.)⁶

Faced with objections along these lines, Gustafsson conceded during our correspondence that his proposals for ruling out whole categories of counterfactuals might need some tweaking,

⁵Gustafsson during our correspondence derided this procedure as “backtracking compatibilism,” but offered no substantive critique. A quantum physicist might raise various objections to the universe you have imagined, but such objections should carry little weight among philosophers of free will (who spend much of their time contemplating worlds where non-quantum rules prevail anyway).

⁶Strictly speaking, it was unnecessary for us to write this paragraph: the preceding remarks already show how poor the prospects are for the Fixity of the Past to be useful when handling the sorts of counterfactuals found in Gustafsson’s Consequence Argument. Nonetheless, we considered this last thought-experiment valuable, as an antidote to the misguided thinking motivating “Fixity of the Past” proposals generally.

but that regardless, his main objection to claim A concerns not the counterfactual implication, but the part of the statement asserting that action γ is *possible* — to use his terms, that S actually “is able” to bring γ about. In other words, he is falling back on the standard incompatibilist strategy, addressed in the first sections of “Who’s Afraid of Determinism,” of restricting the types of universes we are allowed to include in set X when making assertions about what is possible, permitting only those worlds whose pasts are identical to reality. As it happens, the objections raised in the preceding paragraph to artificial restrictions on counterfactuals apply equally well to equivalent restrictions on possibility. Although it sounds like a nice principle to insist that people “hold the past fixed” down to the last submicroscopic detail, there is no reason to think that anyone in everyday reasoning really obeys it. If Jane Q. Public asserts that Austin “could have” made his famous failed putt, and philosopher Phil contradicts her with the argument that the only lawful universes in which Austin makes the putt are ones where a single electron’s position during the Big Bang differs from reality by 10^{-1000} meters, Public is absolutely entitled to shrug off this beyond-undetectable historical circumstance as irrelevant to her understanding of Austin and his golfing capabilities. Phil’s restricted notion of “possible” bears no relation to what matters to a normal rational agent.

At any rate, with all the extra smoke now blown away and the crucial premiss for Gustafsson’s Argument fully exposed — that an imagined action by S is only “possible” if it forms part of a universe where both the laws and the *exact* initial state are the same as actuality — we can see how very similar it is to the Argument’s conclusion in line (6): that, in a deterministic setting, S never has the power to bring about any non-actual event. By the definition of determinism (to reiterate), only *one* universe matches reality’s initial state while

satisfying deterministic laws; hence for anyone who accepts incompatibilist restrictions on possibility, it is utterly trivial to deduce that under such laws only the actual is possible.

Although Gustafsson's rescue attempt for the Consequence Argument stands revealed as a rather pointless, question-begging endeavor, and although responding to it only required that we reiterate points already evident in our earlier writings, his article deserved a reply for a couple of reasons. First, as we have seen, his (re)interpretation of the Consequence Argument effectively jettisons notions relating to causation in favor of simpler, purely counterfactual statements, thereby shoring up the Argument's line (4); he then defends line (5) by imposing familiar restrictions on the word "possible." He thus transforms, subtly, an Argument whose original riposte centered on concepts developed in later sections of "Who's Afraid of Determinism" into one that forces us to invoke the concepts covered in the opening sections. Overall, his paper illustrates in an instructive fashion how the various strands of the dispute may become intricately intertwined. Secondly, it must be said that although Gustafsson clothes his arguments in rigorous-seeming logical garb, many of his definitions involve superfluous, confusing complications that are left unexplained (like his Rule Gamma), or else they contain gaping holes (like his Fixity of the Past principle for counterfactuals). While we would not accuse him of conflating Claims A and B above, it is likely that his approach will appeal particularly to those who do suffer from mix-ups of this sort. Gustafsson's decision not to address any of these problems, despite our alerting him to them in correspondence, does suggest a certain nonchalance about the possibility of his work fostering confusion, and for this reason we felt obliged to issue a corrective.

References

Gustafsson, J. 2020. Dennett and Taylor's Alleged Refutation of the Consequence Argument.

Analysis. <https://doi.org/10.1093/analys/anz048>

Taylor, C. and D. Dennett. 2002. Who's afraid of determinism? Rethinking causes and possibilities. In *The Oxford Handbook of Free Will*, ed. R. Kane, 257–77. New York: Oxford University Press.

Taylor, C. and D. Dennett. 2011. Who's still afraid of determinism? Rethinking causes and possibilities. In *The Oxford Handbook of Free Will*, 2nd edn. R. Kane, 221–40. New York: Oxford University Press.