Is Sylvan's Box a Threat to Classical Logic Norms?

Winner of the Gerritt and Edith Schipper Undergraduate Award
for Outstanding Undergraduate Paper at the
57th Annual Meeting of the Florida Philosophical Association

Theodore D. Locke, *University of North Florida*

I. Introduction

Consider a world in which every proposition obtains (i.e., under a given consequence relation *every* proposition holds); call this world "the *trivial* world."[1] Next, consider an inference of classical logic *ex contradictione quodlibet*, or, as it is often referred to in the contemporary literature, the principle of *explosion*. Loosely stated, this principle allows *any* proposition to follow from a contradiction.[2] Let an *impossible world* be a world in which a logical rule or logical truth is violated (e.g., a world in which the conjunction of a sentence and its negation holds).[3] It follows from classical logic that such a world, by the principle of explosion, is the trivial world. However, in "Sylvan's Box: a Short Story and Ten Morals," Graham Priest presents a fictional story that is intended to be an example of a non-trivial impossible world (i.e., an impossible world in which not every proposition obtains).[4] Though motivations for discussing impossible worlds vary, an important motivation driving "Sylvan's Box" and Priest's morals is to show that the logic used to reason about the story is *paraconsistent*, that is, the consequence relation used to make inferences from the content of the story is not one that validates arguments resulting from the classical principle of explosion. In his explication of ten morals that follow from the story, Priest argues that the story contains non-trivial information from which one can draw relevant, non-arbitrary inferences despite the existence of logical falsehoods within the story that are *essential* to the plot.

Advocates of certain non-classical logics are willing to take impossible fiction and inconsistent theories, such as the old quantum theory, as evidence that classical logic fails to provide correct norms for reasoning when inconsistencies are encountered.[5] Moreover, they take this failure as a sufficient reason to revise classical accounts of consequence. To this end, Priest offers "Sylvan's Box." However, in "A Consistent Reading of *Sylvan's Box*," Daniel Nolan suggests that the story may instead be read consistently by considering the literary phenomenon of unreliable narrators. Nolan claims that such a consistent reading renders Priest's claim that the story is essentially inconsistent

false and, as such, potentially undermines Priest's objective.[6] In this paper, I will explain and evaluate an argument for the conclusion that classical logic[7] provides incorrect norms for reasoning about impossibilities, which is supported by Priest's claim that "Sylvan's Box" is essentially inconsistent and intelligible. I will also explain and evaluate an objection to this argument based on Nolan's suggested consistent reading. I will argue against objections based on Nolan's suggested reading. However, I will conclude that it is not the case that classical logic provides incorrect norms for reasoning about impossible situations. I will discuss logical tools, offered by Daniel Nolan, compatible with classical accounts of consequence that allow for non-vacuous reasoning about impossibilities. Finally, I will offer an explanation for seeming tensions between certain classical logic principles and reasoning about impossibilities. First, it will be instructive to discuss, in more detail, the classically derived principle of explosion and classical accounts of consequence.

II. Classical Consequence and Explosion

Logic is often seen as a discipline that studies patterns of reasoning and making inferences. Ideally, our formal accounts of logical relations such as consequence will adequately relate to and capture our pre-theoretic conceptions of those relations. These conceptions are often engendered by our semantic intuitions about the language we are studying, our epistemic intuitions about what norms should govern our reasoning, or our alethic intuitions about how truth is preserved from the premises to a conclusion of an argument.[8] Formally, we define consequence in a language via a deductive system and/or a model theory. Often, consequence in a deductive system is defined in terms of proofs involving premises and conclusion(s), and consequence in a model theory is defined in terms of the satisfaction of premises and conclusion(s) across a domain of models. However, in his paper "Logical Consequence, Proof and Model Theories," Stewart Shapiro offers pre-theoretic counterparts to the formal resources of a given logic that are meant to capture our intuitions about consequence. Let '**T**' stand for a set of premises and let '*B*' stand for an arbitrary proposition. First, our epistemic intuitions are captured by the following pre-theoretic definitions of consequence,

> (R) *B* is a logical consequence of **T** if it is irrational to maintain that every member of **T** is true and that *B* is false. The premises **T** alone justify the conclusion.[9]
>
> (Ded) *B* is a logical consequence of **T** if there is a deduction of *B* from **T** by a chain of legitimate, gapfree (self-evident) rules of inference.[10]

Next, our intuitions about truth preservation are generally explicated in terms of modality (e.g., through all possible worlds, or in terms of semantics, i.e., through all interpretations of a language). In this case, Shapiro offers a blended pre-theoretic definition of consequence,

(PwI)   *B* is a logical consequence of **T** if *B* is true in every possible world under every interpretation of the non-logical terminology in which every member of **T** is true. [11]

Shapiro argues that definitions of consequence in terms of model theory should correctly and adequately relate to these pre-theoretic definitions. [12] It is worth pointing out that (PwI) also relates to our epistemic intuitions about reasoning insofar as it illustrates the usefulness of finding counterexamples to arguments that utilize fallacious reasoning. [13] Consider the claim that the deductive system and model theory of classical logic (including model theories expanded to include possible worlds) adequately and correctly relate to Shapiro's pre-theoretic definitions of consequence. The question will become this: in light of impossible worlds or inconsistent theories, do these pre-theoretic definitions need to be revised? In other words, should we weaken (PwI) to allow for impossible worlds or weaken (R) to allow one to rationally hold that **T** is true when **T** is inconsistent?

Consider again the principle of classical logic *ex contradictione quodlibet*, or, simply, explosion. The principle of explosion is represented formally as the inference $\{A, \neg A\} \vDash B$ [14]; given a set of premises, if at least two are contradictory, *any* proposition logically follows. Referring back to our pre-theoretic definitions of consequence, it is easy to see that there can be no possible world (or interpretation of a language [15]) in which the above premises come out true and the conclusion false. Moreover, it doesn't seem rational to maintain a contradiction is true, pace dialetheism. Hence, according to both of the classical accounts of consequence, these arguments are logically and deductively valid. Next, let's consider a world as a set of propositions closed under the consequence relation for a specific logic. [16] Recall that a world in which every proposition obtains is trivial. Evaluated under a classical consequence relation, impossible worlds, those worlds in which a logical rule or logical truth is violated, are trivial by the principle of explosion since any arbitrary proposition follows from the logical falsehood. [17] Worlds intended not to be trivial and closed under classical consequence fail in at least one of these respects when they house logical inconsistencies. An interesting question follows: how well does such a principle reflect how we should reason about worlds or theories where contradictions arise? On the one hand, evaluated with a semantics consistent with classical logic, it seems that inconsistent worlds can never provide counterexamples to arguments since every proposition is true (or false) in those worlds. Furthermore, one might argue that the principle of explosion motivates a fallacious norm that one has reason to remove a contradictory proposition of an impossible world in order to draw any nontrivial conclusions about that world; when confronted with a contradiction one no longer has the means to draw meaningful conclusions about that world, since anything follows, and must walk away. [18] We will now look at a proposed counterexample to this principle.

III. Sylvan's Box

Impossible fiction is often cited as a salient counterexample to explosion and counted as evidence that classical accounts of validity fail to provide norms for making non-vacuous inferences from the information given in the story.[19] Graham Priest attempts to give such a counterexample with "Sylvan's Box." In the story, which is told in the first person by Priest, Priest visits the home of the recently deceased philosopher Richard Sylvan.[20] Waiting at the residence, is Nick Griffin, Sylvan's literary executor. Both Priest and Griffin are there to organize the large volume of intellectual work left behind by Sylvan, much of which is on paraconsistent logic and Meinongian metaphysics. While mining through all of the material, Priest happens upon a box labeled "Impossible Object." Priest opens the box and finds that the box is empty and has, fixed to the base, a small figurine. In other words, he finds that the box is empty and not empty at the same time. A considerable amount of time passes, during which Priest reflects on his discovery. Eventually, Griffin, who was handling Sylvan's affairs at the university, returns to the bungalow and has nearly the same reaction as Priest when shown the box. The story ends inconsistently with Priest placing the box in the trunk of his car and Griffin burying the same box, both before they turn to one another for farewells.

The morals that Priest proceeds to draw from the story concern truth in fiction, analysis of beliefs, competing paraconsistent logics, and philosophical positions regarding the ontological status of impossible worlds. Most relevant to our purpose is that Priest claims "Sylvan's Box" is an essentially inconsistent story – the story describes a logically impossible situation that is essential to the plot; moreover, the impossible world portrayed is non-trivial – not every proposition follows from the content of the story. An additional moral that Priest draws from his story is that the logic employed when reasoning about the story cannot be classical logic.[21] In other words, classical logic gives incorrect norms for reasoning about the story. Instead, the logic is or should be *paraconsistent*, that is, a logic in which the logical consequence relation does not allow an inference to be made from a contradiction to an arbitrary conclusion.[22] To illustrate, consider the following inference that might be made from the story:

> (**A1**) Priest put the box in his trunk and Griffin buried the same box together entail that the box was sent to the moon.

Note that employing a classical consequence relation would lead us to conclude that (A1) is valid, and trivially so, because the premises are false in every possible world. However, employing particular paraconsistent logics, many of which have model theories that utilize impossible worlds, the semantic interpretation would not be the same. Instead, (A1) would be not be logically valid since there are impossible worlds where both the premises come out true but the conclusion false.

In nuce, on this reading, "Sylvan's Box" provides a cogent example of a story that is, *prima facie*, essentially inconsistent, intelligible, and is an example of a non-trivial impossible world that we can reason about in a non-vacuous manner.

Advocates of paraconsistent logic would be willing to take "Sylvan's Box" and Priest's morals as evidence supporting the claim that classical logic provides incorrect norms for reasoning, which is then taken as motivation for revising classical accounts of validity.[23] Consider the following representative argument:

Impossible World Argument

(**P1**) "Sylvan's Box" is essentially inconsistent and intelligible.

(**P2**) If "Sylvan's Box" is essentially inconsistent and intelligible, then it is an example of a non-trivial impossible world that we can reason about in a non-vacuous manner.

(**C1**) "Sylvan's Box" is an example of a non-trivial impossible world that we can reason about in a non-vacuous manner. (P1, P2)

(**P3**) If classical logic gives us correct norms of reasoning, then it is not the case that "Sylvan's Box" is an example of a non-trivial impossible world that we can reason about in a non-vacuous manner.

(**C2**) Classical logic does not give us correct norms of reasoning. (C1, P3)

Under the assumption that a logic should provide correct norms for reasoning and that these norms are reflected by the consequence relation defined in the metalanguage of the logic, an additional conclusion is drawn: the classical notion of consequence is in need of revision. For now, the concern is with the first half of this argument. Premise (P1) is supported by the existence of an impossible object within the story. Priest comments that the box prevents the story from being divided up into consistent parts, a method known in the literature as *chunking*, which could give the story a consistent reading.[24] We will look at an objection to this premise shortly. Premise (P2) stands as it seems incorrect to claim that the story is essentially inconsistent and intelligible but also trivial and one from which we may legitimately infer whatever we please. For example, it seems that one may not legitimately infer "after discovering the box, Priest managed to square the circle," which never occurs in the story nor is even implicitly implied by the content of the story. Hence, "Sylvan's Box" is an example of a non-trivial impossible world that we can reason about in a non-vacuous manner.[25]

IV. Impossible Worlds and Unreliable Narrators

In "A Consistent Reading of *Sylvan's Box*," Daniel Nolan offers an appealing reading of the story that is supposed to make the content of the story consistent contra (P1) of the Impossible

World Argument.[26] Nolan claims that if there is a consistent reading, then Priest's claim that the story is essentially inconsistent is false. Obviously, this line of argument would be very useful for those who wish to deny premise (P1). Nolan's suggestion is to read the story, written in first person with Priest as narrator, as if Priest *believes* he has found an impossible object. Furthermore, the fact that Priest and Griffin have been sorting through and reading all of Sylvan's philosophical writings, much of which is on impossible objects and paraconsistent logic, explains *why* they would believe they have found such an object and *why* they would believe that they treated the box in an inconsistent manner at the end of the story. Nolan states, "It is of course, entirely consistent that a pair of logicians [especially Priest, who entertains true contradictions in real life] might come to believe contradictions as a result of some experiences," and, "a good case can be made that the story suggests to the hearer that Priest *thought* he had found a strange object, much more strongly than it suggests that Priest might really have found an impossible object." Reading the story by Nolan's suggestion would be very helpful for those who are concerned about the potential threat "Sylvan's Box" and Priest's derived morals pose to the classical notion of logical consequence. The reading might allow for classical logic to be employed in making inferences from the content of the story by way of dividing the story into consistent parts (in particular, to account for the inconsistent ending), since now the impossible object itself can be explained away as being the product of an incorrect belief.[27]

However, to show that the story is not essentially inconsistent proponents of this reading must also give convincing evidence via the content of the story that suggests this reading is sufficiently supported by that content. Nolan defends his suggested reading in two substantial ways. First, reading the story this way does not harm the content of the story – an impossible object is still involved, but only in the sense that it is an object *believed* to be impossible by the characters. Second, by this reading, appeal may be made to the literary phenomenon of *unreliable narrators*. An unreliable narrator incorrectly judges what occurs within the world the story unfolds.[28] However, there are concerns with this offered defense that reveal this reading is not as threatening to the Impossible World Argument as it first appears. First, there is little reason to think that reading the story solely as if Priest believes he found an impossible object is supported by the content of the story in any nontrivial manner. Furthermore, it looks as though the only reason to read fictional Priest as an unreliable narrator is external to the content of the story and begs the question with respect to impossible worlds.

Nolan argues that his reading does not harm the content of the story. An initial objection to this claim is that by Nolan's reading the content of the story that reflects Priest's incredible experience is left deflated. The impossible object is essential to the story – if the attitudes and behavior of the two men are to make any sense, then they must have found an impossible object.[29]

However, this objection does not recognize the subtly of Nolan's suggestion. The box needn't be an *impossible* object in order to explain the behavior and attitudes of Priest and Griffin when they encounter the box; the box need only be something *very strange*. So, at least this much content of the story *is* left intact by Nolan's reading. However, simply appealing to the beliefs of the narrator is not an interesting reason to claim the *entire* content of the story is unharmed, for any story can be trivially reinterpreted as being about what the narrator *believed* he experienced as opposed to what the narrator *actually* experienced.[30] Whatever support Nolan's suggested reading gets from the story is, so far, trivial. Appealing to Priest's beliefs doesn't provide good evidence that he did not actually experience an impossible object; moreover, this reading does not adequately explain the inconsistent ending. For Nolan's suggestion to gain full momentum, one must also claim that Priest is an unreliable narrator and have good reason for making that accusation.

If we are to charitably follow the narrator of a story, we should believe what they tell us about the world in which the narration takes place, unless we have an overriding reason to do otherwise such as evidence from the story that the narrator is deceitful or hallucinating. If we play along with the narration and take fictional Priest for his word, then the real Priest does provide a story that contains an essential inconsistency. In the story, Priest does not narrate that he was shocked to find some *very strange object*, instead he asserts, with prudent confidence, that the box he found was both empty and had something in it. Opponents who follow Nolan's advice will argue "no." They will claim that it is better to read the story with fictional Priest as an unreliable narrator; after all, Priest is making contradictory claims.[31] However, to be charitable one must acknowledge that, via the content of the story, Priest is not making a hasty judgment – he is analyzing his experience in *great detail* and with *cautious self-reflection*. He wrestles with his discovery for the rest of the story despite his confidence about what he has discovered. One who claims that Priest is an unreliable narrator must offer an explanation, via the content of the story, for why his testimony about his and Griffin's experience with the box is faulty.[32]

Priest is a *reliable* narrator just in case what he is recounting from within the story is his encounter with an impossible object and his experience of the inconsistent way in which he and Griffin dealt with it; this can be only if there is, in some sense, a non-trivial impossible world.[33] A proponent of Nolan's suggested reading needs to explain why Priest's beliefs are not supported by his experience and why he is thus an *unreliable* narrator. The proponent would, more than likely, be more than willing to grant possible worlds to consistent stories told by reliable narrators, that is, narrators who are reporting things they take to be known as fact. With respect to "Sylvan's Box", a proponent of Nolan's suggested reading can deny that Priest is a reliable narrator only by appealing to the claim that there can be no impossible objects or non-trivial impossible worlds in any sense, including fictional contexts. However, this begs the question with respect to impossible worlds and

whether or not "Sylvan's Box" is an essentially inconsistent and intelligible story. As is often the case when a logical inconsistency is in close proximity one may feel that she has a right to insist that such states of affairs are "nonsense." So, it does not matter that this line of reasoning is circuitous – there simply are no logically impossible objects. But "nonsense" seems close to "impossible," which is what is under discussion – of course there are no boxes that are both empty and not-empty. So, a proponent of Priest's reading will more than likely continue to hold her antecedent inclination to read "Sylvan's Box" as *impossible* yet non-trivial.[34] Hence, instead of rejecting the Impossible World Argument by denying (P1) with an antecedent assumption about impossible worlds, it would be better to make the rejection to this argument stronger by granting that premise.[35]

V. Impossible Worlds and Logical Norms

The Impossible World Argument concludes that classical logic does not give us correct norms of reasoning, and, further, that the classical notion of logical validity needs revision. Given that explosion follows from classical logic, if we employ classical logic when reasoning about the "Sylvan's Box", then it seems we should find that we are unable to draw non-vacuous conclusions about the content of the story. Moreover, given explosion, we may even find that the story is trivial insofar as the content of the story logically entails every proposition. But, we are able to draw non-vacuous conclusions from the story, and the story does not seem trivial at all. Given the intractable inconsistencies found in the story, it is concluded that classical logic provides incorrect norms for reasoning about impossibilities. However, claiming that classical logic yields incorrect norms of reasoning on the basis of impossible worlds is plausible only if one of the following is true:

(1) There is no alternative to reasoning with impossibilities that allows us to consistently keep a classical notion of validity; or

(2) There is no good explanation for why it is acceptable that the classical account of validity supports a norm that declares reasoning from contradictions trivial.

I will look at each of these in turn. If both are false, then we have good reason to deny (P3): we can reason non-vacuously about "Sylvan's Box" while claiming that classical logic provides correct norms of reasoning.

First, there *is* an alternative to reasoning with impossibilities that allows us to keep a classical notion of validity. It is found in another paper written by Daniel Nolan, "Impossible Worlds: A Modest Approach." In this paper Nolan argues that a notion of logical consequence that accounts for *any* counterpossible situation, or *any* impossible world, will be a very weak notion. In fact, too weak, for any violation of the alternative notion of consequence will itself be an impossible situation; thus, the already attenuated notion of consequence will need to be further adulterated.[36] Instead,

Nolan suggests that we should look to counterfactual conditionals as a way to deal with impossible worlds.[37] Consider that many impossible worlds are close enough to the actual world in appropriate respects to consider what would or would not follow from an impossible antecedent. For example, though impossible, the world in which Priest really did find a box that was both empty and not empty is more than likely not very different from the actual world in other appropriate respects. In fact, it is closer to the actual world than the one in which Priest found the box and had it sent to the moon of Jupiter, Io, let alone the trivial world in which every proposition is true (or every proposition is false). Following Nolan's suggestion of employing *counterpossible conditionals* to model impossible situations allows reasoning with inconsistent information *within the scope* of those conditionals; Nolan likens this to hypothetical reasoning in general and hypothetical reasoning about necessary falsehoods.[38] So, when reasoning about "Sylvan's Box" one may not legitimately infer that Priest sent the box to moon of Jupiter, Io, from the fact that the box is both empty and not empty – in the impossible world closest to the actual world where Priest finds the box, such a thing does not occur. The classical account of validity was never intended to account for impossible worlds; however, Nolan's suggestion requires only that the semantics for counterfactual conditionals be expanded to fill this needed role.[39] Hence, a revision of classical validity, which is only defined over possible worlds, on the basis that it provides no mechanism for reasoning with impossible worlds has a spurious motivation.

However, there are two potential worries with Nolan's account of counterpossible conditionals relevant to this discussion. First, our models would now require sets of impossible worlds and things like similarity relations. This might engender more metaphysical discourse than many conservative advocates of classical logic would prefer. However, many who challenge classical notions of consequence on the basis of impossible worlds are already comfortable with non-standard model theories, so this should not be a concern to them. Second, and more concerning, is that by Nolan's account, reasoning about impossibilities must occur *within* the scope of the conditionals and so depend on the content of the antecedent and the consequent (i.e., on the truth of the counterpossible conditional).[40] By Nolan's account, many of the deductive theorems that hold for material or counterfactual conditionals fail for counterpossible conditionals. So what makes an inference rationally justified depends heavily on the content of the conditionals and the similarity relations employed.[41] However, this seems to complicate how we determine the relationship between what is correct when reasoning about impossibilities and our notions of consequence, since it is the consequence relation, in our case at least, that is viewed as the primary indicator of what is correct to reasoning. Keep in mind, Nolan still defines consequence in terms of truth preservation of all possible worlds in all models[42] – the introduction of the counterpossible conditional is a way to cheaply allow for counterpossible reasoning while keeping classical the definitions of deductive and

logical consequence. Nonetheless, critics of classical logic that endorse the Impossible World Argument might claim that while Nolan's alternative to reasoning with impossibilities allows us to keep a classical notion of validity, there is still no good explanation for why it is acceptable that the classical account of validity supports a norm that declares reasoning from impossibilities trivial. After all, even under Nolan account, explosion is still a valid argument. In fact, Nolan's account seems to generate conflicting norms: one from the similarity relation in the counterpossible semantics which does allow non-arbitrary conclusions to be drawn from inconsistent antecedents, and another from the consequence relation which requires revisions of inconsistent premises in the face of absurdity. I will next argue that there are normative explanations for counterintuitive rules of inference such as explosion and the appearance of conflicting norms just mentioned.

One might respond to either of these criticisms by denying that explosion is meant to provide a norm for reasoning. When establishing a normative theory (e.g., a theory of right and wrong action or a theory of good and bad reasoning), some aspects of that theory will more accurately represent what is being studied while other aspects will not. The latter aspects will be artifacts of the theory and sometimes unwanted.[43] For example, utilitarian ethics might deliver verdicts that more often than not coincide with our considered moral beliefs, but also have unwanted features, such as making no act supererogatory. In our case, one might respond to the above criticisms by claiming the principle of explosion is an unwanted artifact with respect to the normative aspect of classical logic. However, unlike unwanted features of utilitarianism, explosion is a desirable aspect of the theory insofar as it is useful in establishing certain metatheories of classical logic (e.g., establishing that the deductive system is complete for the model theory). If this is the case, then either explosion does not reflect a norm for reasoning or it is an unwanted norm that is worth keeping in light of its usefulness in the metatheory. However, this is not a good response. There does not seem to be a non-arbitrary way of deciding which rules of inference or arguments are mere artifacts of our theory or are actually representative of how we should reason unless we beg the question as to which ones reflect rationality. Certainly, some aspects of classical logic are artifacts, like certain notational devices[44], but consequence is more substantial than these insofar as our pre-theoretic definitions of consequence ostensibly reflect norms of reasoning. Ideally, we want to account for how every formal argument validated by our consequence relation relates to the norms we feel are generated by our logic.

Another way to resolve the seeming tension between classical logic and reasoning about impossibilities is to take the position that the study of logic and the study of reasoning are distinct and fundamentally unrelated. In this case, it does not matter to how we reason what theorems fall out of classical logic since there is no such thing as deductive reasoning.[45] Classical logic is really just a formal endeavor that only concerns truth preservation across interpretations of a language.

However, as mentioned previously, it seems that what largely motivates pre-theoretic notions of logical relations (e.g., our account of logical validity), is that these relations closely relate to how we should reason. Aside from the formal study of language and arguments, which produces interesting theorems and other mathematical results, discussion about logical inference seems to follow a rough methodology similar to other normative endeavors. That is, we look for common features in what seem to be good arguments or common inferential moves in what seem to be good chains of reasoning and then set those as criteria for what is an acceptable argument or what is an acceptable move in reasoning. Many of the contexts in which arguments are given and where chains of reasoning take place will not be ideal, but as with most normative endeavors a certain amount of distillation takes place before the sought after features become salient. In mathematics, a discipline that largely drove the establishment of classical logic, there is less need for distillation since the language employed in mathematical reasoning can be very precise.[46] Finding these features in reasoning that takes place in natural languages can, obviously, be more difficult, but we shouldn't conclude that this endeavor is fundamentally unrelated to our formal studies.

Instead, one could think of explosion as reflecting an ideal norm such that whenever an individual has overcommitted herself to a theory by believing a contradictory pair of propositions, then in a *deductively ideal* situation that individual has reason to give up one of the contradictory beliefs in order to make non-trivial inferences. Some features that make a situation deductively ideal are: a clearly defined language, one where every term has a referent; an extensional language, one where whether or not a thing under discussion has a certain property has a definite yes or no answer; one where an individual has a sufficient number of premises to draw out the consequences; and, pertinent to our purposes, one where the individual reasoning is clearly aware as to whether or not the premises or beliefs she is using in drawing a conclusion are consistent. Suppose that someone finds that they have inconsistent beliefs and attempts to explicitly reason with these beliefs. Then it is not difficult to draw any conclusion he pleases by other intuitive rules of inference such as disjunctive introduction and disjunctive syllogism. Thus, we have a legitimate chain of reasoning that is gap-free.[47] On the other hand, if one is explicitly aware of inconsistent beliefs then the correct verdict seems to be that no matter what other belief one comes to, it is irrational to maintain that all of those inconsistent beliefs are true. So on pain of absurdity, one has a good reason to give up as many beliefs as necessary to restore consistency.[48] However, given a situation that isn't deductively ideal, one may have an overriding reason not to accept inferences of classical logic such as explosion.

Before continuing it will be helpful to first draw an analogy with another, arguably, normative endeavor. In epistemology, the theory of evidentialism requires that a doxastic attitude taken towards a belief fits an individual's evidence and that to do otherwise is to have an irrational

belief.[49] However, as pointed out by proponents of evidentialism, this requirement for epistemic rationality does not imply that there are not any moral or theoretical reasons to take on a doxastic attitude towards a proposition not supported by the evidence. In fact, as Richard Feldman and Earl Conee point out, it may even be best to one's epistemic goals to take on a doxastic attitude not supported by the evidence.[50] When we consider the features mentioned above that make a situation logically ideal there are clear examples where one has reasons that may override her logical reasons for accepting certain inferences. For example, one may not have all of the premises required to draw a conclusion but also have that conclusion. Then through a process of abductive reasoning, that is reasoning backwards, one may *legitimately* conclude with a premise, and insofar as an analogy can be drawn between this process and affirming the consequent, it is not acceptable by the norms of classical logic.[51] Nonetheless, in a situation that is not deductively ideal, as is often the case when developing theories to explain observed phenomenon, it is logically permissible (and in some sense obligatory) to deviate for classical logical norms.

Under discussion here, by classical consequence, which ranges over *possible* worlds only, the inferences made from a necessary falsehood are trivial. What should one do when they find they have inconsistent beliefs or are dealing with inconsistent premises? First, one may attempt to revise her beliefs to exclude some of the beliefs. If this is not an immediate option, then while in deliberation, one can partition ones beliefs in a way that allows one to reason from consistent sets of beliefs until the inconsistency is resolved.[52] However, there may be cases where reasoning with impossibilities is unavoidable *and* the inconsistency irresolvable; such a context is not a deductively ideal situation. Despite this, nothing prevents there being theoretical or practical considerations that would motivate one to accept an inconsistent set of propositions or a proposition that is necessarily false. For example, when trying to draw conclusions from an inconsistent story, an inconsistent mathematical theory, a necessarily false philosophical theory, or even a necessarily false account of logic. In the last case, it may even be best to one's logical goals to consider other logical theories.[53] Given classical logic, in all of these cases the inconsistencies can remain within the scope of a logical device *compatible* with consequence in classical logic such as Nolan's counterpossible conditional. One would look to the truth of the counterpossible conditional to determine what one has reason to believe, not the principle of explosion. In situations that are not deductively ideal, the classical notion does not yield an incorrect norm but a norm that can be overridden by context.[54]

Recall the concern that Nolan's counterpossible method yields conflicting verdicts about how one should reason. Nolan's suggestion is that counterpossible reasoning is a type of hypothetical reasoning that occurs within the scope of the counterpossible conditional. The similarity relation in the counterpossible semantics allows non-arbitrary conclusions to be drawn from inconsistent antecedents. From this one has a reason to believe some non-arbitrary conclusion

would follow if they *were* to hold some impossibility. This does not conflict with explosion by which one has reason to believe any proposition if they *actually* hold some impossibility as true. Nolan provides a device that yields a theoretical or practical norm about reasoning in inconsistent situations that overrides logical obligations in deductively ideal situations. So when reasoning about "Sylvan's Box" the classical consequence relation dictates that on pain of absurdity we have reason to either believe the box is empty or it is not, but not both. However, since reasoning about this story puts us in a situation that is not deductively ideal and would cause us to lose what is important about the story, we have an overriding (perhaps, in this case, an aesthetic) reason to not reason according to this norm. Instead we would look to the truth value of the counterpossible conditional to find a reason to legitimately believe some non-arbitrary conclusion *would* follow if we *were* to hold that such a box existed.

VI. Conclusion

I have argued that since there are logical means for reasoning about impossibilities and since there is a good explanation for what only seems to be a counterintuitive norm derived from classical logic, the claim that classical logic fails to provide correct norms of reasoning is false. Hence, we can reason non-vacuously about "Sylvan's Box" while claiming that classical logic provides correct norms of reasoning. This response is motivated by the Impossible World Argument, an argument that concludes classical logic fails to provide correct norms of reasoning. I argued first that to defend classical logic and the corresponding notion of logical validity by denying the first premise is ineffective. Then I evaluated logical tools offered by Daniel Nolan that allow us to reason with impossibilities and keep a classical notion of consequence. There may be other reasons for revising the classical notion of consequence, but appeal to the need for reasoning about impossible worlds or inconsistent theories is not a good reason.

Notes

---

[1] In such a world *every* proposition is true or *every* proposition is false.

[2] Jennifer Fisher, *On the Philosophy of Logic* (Belmont, CA: Thomson Wadsworth, 2008) 119. The proposition may follow deductively or logically.

[3] See Graham Priest, "Sylvan's Box: A Short Story and Ten Morals," *Notre Dame Journal of Formal Logic* 38.4 (Fall 1997): 573-582.

[4] Note that Priest's intended morals could stand regardless of what is assumed about such status about these worlds.

[5] For example, Bryson Brown, in regard to Bohr's theory of the atom, claims, "…until [the inconsistencies] are removed, classical logic can tell us absolutely nothing about how to go on reasoning." Bryson Brown, "Old Quantum Theory: A Paraconsistent Approach," *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 2 (1992) 1.

[6] Since Nolan himself claims to prefer Priest's reading I will try my best to refer to his reading as "Nolan's suggested reading." That is, while Nolan might not subscribe to this particular reading, he offers it to defenders of classical logic; see Daniel Nolan, "A Consistent Reading of *Sylvan's Box*," *The Philosophical Quarterly* 57.229 (October 2007): 667-673.

[7] I will use the term "classical logic" broadly in this paper – I take the term to include the formal language, deductive system, and/or a model theory akin to classical elementary logic but also to include non-deviant extensions such as modal and conditional logics, which utilize possible worlds in the model theory.

[8] Stewart Shapiro, "Logical Consequence, Proof Theory, and Model Theory," in *The Oxford Handbook of Philosophy of Mathematics and Logic*, ed. Stewart Shapiro (New York: Oxford UP, 2005) 659-660.

[9] There is some ambiguity with this definition. The definition could be read as: it is irrational to maintain every member of **T** and it is irrational to maintain B, instead of it is irrational to maintain (every member of **T** and B). I am not completely positive that Shapiro meant the latter, but that is the reading of the definition I will assume for this paper. John MacFarlane, in an unpublished paper, gives excellent reasons against deontic operators, or in this case claims of rationality, applying to both the antecedent and consequent. Instead, he argues that the operators should range over the entire conditional or conjunctive equivalent. John MacFarlane, "In What Sense (If Any) Is Logic Normative for Thought," American Philosophical Association Central Division Meeting (2004) 10; cited with permission.

[10] Shapiro, "Logical Consequence, Proof Theory, and Model Theory," 659-660.

[11] Shapiro, "Logical Consequence, Proof Theory, and Model Theory," 663, and Stewart Shapiro, "Logical Consequence: Models and Modality," in *The Philosophy of Mathematics Today*, ed. Matthias Schirn (New York: Oxford Clarendon Press, 1998) 152. This definition of consequence, as Shapiro points out the only modality the model theory need register is size. However, as mentioned above I would also like to include logics with expanded model theories, that is, interpretations that include sets of possible worlds; see Graham Priest, *An Introduction to Non-Classical Logic: From If to Is*, 2nd Ed. (Cambridge UP, 2008) 309. However, expanded so, the definition of consequence will lose some of its modal "simplicity" and engender more metaphysical discussion.

[12] That is, setting aside issues that arise from considering second-order systems, every formal rule of inference or model correctly corresponds to moves in reasoning or arguments given in a natural language. See Shapiro, "Logical Consequence, Proof Theory, and Model Theory."

[13] See Shapiro, "Logical Consequence: Models and Modality," 141.

[14] For a deductive inference we would replace the double turnstile with a single turnstile.

[15] At least, that is, if we assume a classical account of negation.

[16] We needn't think of a world as being identical with a set of propositions. One might reasonably take worlds as *represented* by a set of propositions. Also, later it will be better to consider the world closed within a collection of worlds similar to that world in other, appropriate respects.

[17] Stephen Read talks about theories, but the idea easily holds for this account of worlds too. See Stephen Read, *Thinking About Logic* (New York: Oxford UP, 1995) 42.

[18] We can loosely translate all "world" talk with "theory" talk. For example, "inconsistent theories can never provide counterexamples to arguments about those theories since every proposition is true (or false) in those theories."

[19] By "non-vacuous inferences" I mean those that are not trivially valid by way of starting from a necessary falsehood.

[20] Priest emphasizes that while certain parts of the story actually took place, much of what happens in the story is fictitious.

[21] See Priest, "Sylvan's Box," 580.

[22] See Priest, *An Introduction to Non-Classical Logic*.

[23] Newton C. A. Da Costa and Stephen French point out a distinction between types of paraconsistent logic: those that are complementary to classical logic and those that are meant to replace classical logic, which they refer to as heterodox. The advocates I am primarily concerned with here are those who have advocated the latter, such as Graham Priest, Richard Routley, Bryson Brown, et al. See Da Costa and French, *Science and Partial Truth* (New York: Oxford UP, 2003) 87.

[24] Priest, "Sylvan's Box," 580.

[25] The Impossible World Argument takes as a premise a fictional story. A proponent of classical logic might feel that the argument is easily dismissed because of this, which is a criticism to be addressed later. However, a proponent of revising classical logic in light of inconsistent theories might feel that the argument could be made more general by including non-fictional examples. For example, there are claimed non-fictional examples often cited as non-trivial theories despite the existence of an inconsistency in the premises (e.g., the old quantum theory). The old quantum theory reigned from about 1910 to the mid-1920s. Niels Bohr's postulates, which underlie the theory, have been read as inconsistent in various ways. For example, it is claimed that, roughly, a charged electron was modeled as orbiting the nucleus using Coulomb's law and a classical view of electric charge, yet did not adhere to classical electrodynamics (CED) insofar as the electron did not continuously release energy (hence, keeping it from collapsing into the nucleus). Instead, the electron released energy only when changing states (e.g., from an excited state back to the ground state). So, the description of the electron adhered to CED and did not adhere to CED. Critics claim that, according to explosion, individuals who worked with the old quantum theory had license to infer whatever they liked from the inconsistent information found in the theory. Hence the theory was trivial. But, they continue, this is absurd. For example, in *An Introduction to Non-Classical Logic* Graham Priest claims, "Bohr did not infer, for example, that electron orbits are rectangular" (75). Advocates of paraconsistent logic take Bohr's model as evidence of a non-trivial theory from which many correctly reasoned despite the inconsistency in the theory, so classical logic should be revised insofar as it does not provide a way to reason through inconsistencies in a non-vacuous manner.

Despite the intuitive appeal of this argument, it is important to note that there are difficulties in interpreting the claims of logical inconsistency in the examples used to support it. First, there is the issue of providing historical evidence that those who worked with the theories really thought that they were essentially and logically inconsistent. Second, there are available explanations compatible with classical logic for the so called inconsistencies. For example, Peter Vickers gives convincing arguments that the old quantum theory was not inconsistent after all. Vickers argues that it is acceptable to take parts of classical electrodynamics to describe the electron's behavior while neglecting those that actually are at odds with quantum theory. Hence, the old quantum theory was not inconsistent, though why the community at the time demurred the theory, often with hostility, is a question left for the historians. See Peter J. Vickers, "Bohr's Theory of the Atom: Content, Closure and Consistency," in *EPSA07: 1ˢᵗ Conference of the European Philosophy of Science Association (Madrid, 15-17 November, 2007)*, 2008. So, given the difficulties surrounding historical examples like this one, in this paper, we will focus on the more idealized fictional examples of impossible worlds

(though they too have difficulties) and the question of whether or not they pose a threat to classical logic norms.

²⁶ A similar, though less detailed, argument is given by Richard Hanley in "As Good As It Gets: Lewis on Truth in Fiction," in *Lewisian Themes: The Philosophy of David K. Lewis* (New York: Oxford UP, 2004) 113-129.

²⁷ Recall that Priest comments that the box prevents the story from being divided up into consistent parts, a method known in the literature as chunking, which could give the story a consistent reading.

²⁸ It is important to keep in mind the distinction between real-life Priest, who is the author, and fictional Priest, who is the narrator. The story is not meant to give an example of a true contradiction; something that real-life Priest claims exists but for which many feel real-life Priest is unreliable.

²⁹ The impossible box, as Priest mentions in his analysis, is what makes sense of the behavior and attitudes of Griffin and Priest after they find the box; see Priest, "Sylvan's Box," 580.

³⁰ Of course, in this context "actually" is used somewhat loosely; certainly nobody (pace metaphysical dialetheists) has *actually* experienced a logically impossible object. Priest's story could not have taken place in a possible world, which is why it is impossible fiction.

³¹ One reason why Nolan's suggested reading might be effective is that the story is written in first person. But does this really matter? It could be argued that even if the story were written in the third person, some narrator would be implied, and hence this mysterious narrator might also be unreliable. So, it seems we are stuck with a narration that we may take either to be reliable or unreliable. To illustrate, David K. Lewis points out the "paradox of the narrator" implied in works written in the third person and that end in "…and so none were left to tell the tale"; see Lewis, "Truth in Fiction," in *Arguing About Metaphysics*, ed. Michael C. Rea (New York: Routledge, 2009) 81.

³² Perhaps, because Priest openly argues that there are true contradictions, even contradictions in the world, and we should assume his fictional counterpart would do the same. Furthermore, in the story, Priest has been reading through Sylvan's work about Meinongian metaphysics and paraconsistent logic, so, within the story, Priest is primed to believe he has experienced the impossible. However, the story could be retold in a different context, one where the main character is not a logician, let alone a dialetheist; the report could certainly be the same: "The box was really empty and occupied at the same time."

³³ All that is needed in this case is that there is an impossible world a fictional sense, which is at least what Nolan's suggestion threatens.

³⁴ This is the reading that strikes me as the most natural, though impossible.

[35] To reiterate the point made in note 24, the nice thing about "Sylvan's Box" and potentially other impossible fictions is that the question of its internal inconsistency will come down to more basic antecedent inclinations about logic and impossible worlds. In this way it is less messy than the non-fictional examples, which depend not only on antecedent inclinations about inconsistent theories, but also inclinations about more general interpretations of various theories found in mathematics, physics, and science, not to mention questions of historical accuracy and just what was cognitively accepted by those working with the theories.

[36] See Daniel Nolan, "Impossible Worlds: A Modest Approach," *Notre Dame Journal of Formal Logic* 38.4 (Fall 1997): 547-548.

[37] Robert Stalnaker and David Lewis are well known for their (independent) formulations of a conditional logic that utilizes counterfactual conditionals. Roughly, to see if a counterfactual conditional is true one looks to the closest world where the antecedent is true and sees if the consequent is true as well. All of this turns on a similarity relation and how that relation is defined. Roughly, one holds every proposition not being evaluated constant and finds the world within a sphere of that world similar to that world in those respects. See Priest, *An Introduction to Non-Classical Logic*, Chapter 5.

[38] Priest, *An Introduction to Non-Classical Logic*, 557.

[39] Priest, *An Introduction to Non-Classical Logic*, 546.

[40] Nolan states, "So I hope I may be excused for proposing only to demonstrate why some inferences [in his counterpossible semantics] involving conditional are rationally justified, rather than proving very many theorems…I think impossible worlds are too badly behaved for there to be many principles which hold distinctively of the conditional regardless of the content of the antecedent and consequent…" (Priest, *An Introduction to Non-Classical Logic*, 553-554).

[41] For example, Nolan has us consider the following theorem: "if A then B" entails "if A then C," where C is a logical consequence of B. Suppose A and B are both impossible. By classical accounts of consequence, if B is impossible then no matter what C is C is a consequence of B. However, depending on our similarity relation, the closest A world might not be a C world and so "if A then C" would be false. Hence, cases where it which would be rational to conclude "if A then C" would not depend on deductive theorems but instead on the similarity relations employed in the model theory. As Nolan states, "Nevertheless the inference will often be rationally justified, since the impossible worlds most similar to ours will by and large have the consequences of most of the propositions which are true at them also true at them" (Priest, *An Introduction to Non-Classical Logic*, 551).

[42] Priest, *An Introduction to Non-Classical Logic*, 563.

[43] See Shapiro, "Logical Consequence: Models and Modality," 138-139.

[44] Shapiro points out parentheses are artifacts of the formal language (Shapiro, "Logical Consequence: Models and Modality," 138).

[45] Gilbert Harman points out that such a distinction would make arguments calling for a new logic based on explosion lose plausibility. See Gilbert Harman, *Change in View: Principles of Reasoning* (Cambridge, MA: MIT Press, 1989) 6.

[46] See John P Burgess, "No Requirement for Relevance," in *The Oxford Handbook of Philosophy of Mathematics and Logic*, ed. Stewart Shapiro (New York: Oxford UP, 2005) 727.

[47] This is somewhat begging the question as many relevantists reject disjunctive syllogism. But the main point will not be lost.

[48] Many of Harman's arguments for maintaining a distinction between reasoning and deductive arguments, such as over-demanding obligations to believe consequences of ones beliefs, are allayed by putting the norms in terms of reasons for belief instead of strict obligations; see MacFarlane, "In What Sense (If Any) Is Logic Normative for Thought."

[49] See Richard Feldman and Earl Conee, "Evidentialism," in *Arguing about Knowledge*, ed. Ram Neta and Duncan Pritchard (New York: Routledge, 2009) 174-191.

[50] Feldman and Conee, "Evidentialism," 179.

[51] Attempts to formalize inductive and abductive reasoning result in logics that are non-monotonic and so are appropriately divergent from classical logic, which is monotonic. However, the Impossible World Argument is a direct criticism of classical logic and I am concerned with those who want to replace classical logic as *the* deductive logic on the basis of reasoning with inconsistencies, see note 22.

[52] This would be a doxastic application of Lewis's method of intersection and method of union both of which, Lewis explains, are available to deal with inconsistencies found in fiction. In the method of intersection one makes multiple revisions of a story and sees what is true in the intersection of those revisions. There is the more general method of union in which one divides the story up into maximally consistent parts and sees what is true in the union. However, in this case we should not close under implication, see postscripts in Lewis, "Truth in Fiction."

[53] See Nolan, "Impossible Worlds."

[54] Impossible situations are not the only cases where logical norms could be overridden. For example, MacFarlane points out that the preface paradox illustrates that there can be times when epistemic obligations conflict with our logical obligation; see MacFarlane, "In What Sense (If Any) Is Logic Normative for Thought."

Bibliography

Brown, Bryson. "Old Quantum Theory: A Paraconsistent Approach." *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 2 (1992): 397-411.

Burgess, John P. "No Requirement for Relevance." In *The Oxford Handbook of Philosophy of Mathematics and Logic*. Ed. Stewart Shapiro. New York: Oxford UP, 2005, 727-750.

Da Costa, Newton C. A. and Stephen French. *Science and Partial Truth*. New York: Oxford UP, 2003.

Feldman, Richard and Earl Conee. "Evidentialism." In *Arguing about Knowledge*. Ed. Ram Neta and Duncan Pritchard. New York: Routledge, 2009, 174-191.

Fisher, Jennifer. *On the Philosophy of Logic*. Belmont, CA: Thomson Wadsworth, 2008.

Hanley, Richard. "As Good As It Gets: Lewis on Truth in Fiction." In *Lewisian Themes: The Philosophy of David K. Lewis*. New York: Oxford UP, 2004, 113-129.

Harman, Gilbert. *Change in View: Principles of Reasoning*, Cambridge, MA: MIT Press, 1989.

Lewis, David K. "Truth in Fiction." In *Arguing About Metaphysics*. Ed. Michael C. Rea. New York: Routledge, 2009, 78-93.

MacFarlane, John. "In What Sense (If Any) Is Logic Normative for Thought." American Philosophical Association Central Division Meeting, 2004.

Nolan, Daniel. "A Consistent Reading of *Sylvan's Box*." *The Philosophical Quarterly* 57.229 (October 2007): 667-673.

Nolan, Daniel. "Impossible Worlds: A Modest Approach." *Notre Dame Journal of Formal Logic* 38.4 (Fall 1997): 535-571.

Priest, Graham. *An Introduction to Non-Classical Logic: From If to Is*. 2nd Ed. Cambridge UP, 2008.

Priest, Graham. "Sylvan's Box: A Short Story and Ten Morals." *Notre Dame Journal of Formal Logic* 38.4 (Fall 1997): 573-582.

Priest, Graham and Richard Routley. "Applications of Paraconsistent Logic." In *Paraconsistent Logic: Essays on the Inconsistent*. Eds. Graham Priest, Richard Routley, and Jean Norman Munchen: Philosophia, 1989, 367-381.

Read, Stephen. *Thinking About Logic*. New York: Oxford UP, 1995.

Shapiro, Stewart. "Logical Consequence: Models and Modality." In *The Philosophy of Mathematics Today*. Ed. Matthias Schirn. New York: Oxford Clarendon Press, 1998, 131-156.

Shapiro, Stewart. "Logical Consequence, Proof Theory, and Model Theory." In *The Oxford Handbook of Philosophy of Mathematics and Logic*. Ed. Stewart Shapiro. New York: Oxford UP, 2005, 659-661.

Vickers, Peter J. "Bohr's Theory of the Atom: Content, Closure and Consistency." In *EPSA07: 1st Conference of the European Philosophy of Science Association (Madrid, 15-17 November, 2007)*, 2008. Available at *Phil Sci Archive*: http://philsci-archive.pitt.edu/id/eprint/4005. Accessed January 21, 2013.