ORIGINAL RESEARCH

Logic talk

Alexander W. Kocurek[1]

Abstract
Sentences about logic are often used to show that certain embedding expressions (attitude verbs, conditionals, etc.) are hyperintensional. Yet it is not clear how to regiment "logic talk" in the object language so that it can be compositionally embedded under such expressions. In this paper, I develop a formal system called *hyperlogic* that is designed to do just that. I provide a hyperintensional semantics for hyperlogic that doesn't appeal to logically impossible worlds, as traditionally understood, but instead uses a shiftable parameter that determines the interpretation of the logical connectives. I argue this semantics compares favorably to the more common impossible worlds semantics, which faces difficulties interpreting propositionally quantified logic talk.

Keywords Metalogical claims · Hyperconvention semantics · Hyperintensionality · Propositional quantifiers · Hybrid logic · Impossible worlds · Counterfactuals

1 Introduction

Philosophers talk about logic all the time. They debate which logic is correct (if any), which laws of logic hold, which inferences are valid, and so on. But philosophical debates are not the only place where such talk occurs. We can attribute beliefs about logic to people ("My colleague thinks intuitionistic logic is correct"), counterfactually reason about alternative logics ("Teaching logic would be way harder if the law of noncontradiction failed"), and more besides. Such "logic talk" seems continuous with the way we talk about non-logical matters. Yet relatively little has been done to situate

✉ Alexander W. Kocurek
awk78@cornell.edu

1   Sage School of Philosophy, Cornell University, Ithaca, USA

such talk about logic within a broader semantic theory. My aim here is to make some initial steps toward such a theory—that is, to develop a semantics for logic talk.

By 'logic talk', I roughly mean any sentence that is, in some sense, about logic, i.e., a *meta*logical claim. Here are some simple examples:

(1)   Classical logic is correct.

(2)   The law of excluded middle holds.

(3)   Some contradiction entails everything.

These are stereotypical examples of the kinds of claims philosophers debate. But logic talk is not restricted just to these sorts of simple sentences. It includes, e.g., sentences that describe what holds *according to* other logics.

(4)   According to intuitionistic logic, the law of excluded middle fails.

(5)   Nothing is valid in the strong Kleene logic.

(6)   Everything that is intuitionistically valid is classically valid.

It also includes sentences that embed metalogical claims under, e.g., attitude verbs, conditionals, or modals.

(7)   Inej believes intuitionistic logic is correct.

(8)   If the Liar were true and not true, the law of noncontradiction would fail.

(9)   Classical logic might not be correct.

Embedded examples such as (7)–(9) pose two challenges for the theory of logic talk. First, there is the familiar hyperintensionality problem. The standard intensional semantic theories for attitude verbs, conditionals, and modals predict that these expressions validate the replacement of necessary equivalents. Yet this prediction seems incorrect, as illustrated by (10)–(12).

(10)   Inej believes paraconsistent logic is correct.

(11)   If the Liar were true and not true, the law of excluded middle would fail.

(12)   The trivial logic might be correct.

These do not seem equivalent to (7)–(9), even though (assuming classical logic is correct) their constituents are necessarily false, and thus necessarily equivalent. One challenge, then, is to explain what's going on here.

But these examples raise a second, more basic problem: it's not even clear how to *regiment* metalogical claims in the first place. Normally, metalogical claims are stated in the metalanguage—we talk of axiomatic proofs, models, sequents, and so on. Examples like (7)–(9) require bringing such talk into the object language: we need to assign compositional semantic values to metalogical claims. But how? We cannot, for instance, regiment (8) as '$(l \wedge \neg l) \,\square\!\!\rightarrow\, \neg \vDash \neg(\phi \wedge \neg \phi)$': the string '$\vDash \neg(\phi \wedge \neg \phi)$' is not a formula of the object language but rather an abbreviation in the metalanguage. What we need is an object-level regimentation of '$\vDash \neg(\phi \wedge \neg \phi)$' so that it can be given a compositional semantic value.

My goal is to motivate and develop a formal system called **hyperlogic** that can help address these problems. Hyperlogic solves the regimentation problem using three ingredients. First, it introduces a multigrade operator $\rhd$ representing entailment. So where $\alpha_1, \ldots, \alpha_n, \beta$ are formulas, $(\alpha_1, \ldots, \alpha_n \rhd \beta)$ is a formula which, intuitively, represents the claim that $\alpha_1, \ldots, \alpha_n$ entail β. Second, laws of logic are formalized with propositional quantifiers (Fine, 1970). Thus, we can regiment the law of excluded middle as $\forall p(\rhd(p \vee \neg p))$. Finally, claims about what laws hold *according to* a logic are formalized using an "according-to" operator @ borrowed from hybrid logic (Areces and ten Cate, 2006). So, we can regiment (4) as $@_{il} \neg \forall p(\rhd(p \vee \neg p))$ where *il* stands for intuitionistic logic.

To solve the hyperintensionality problem, theorists standardly invoke "logically impossible worlds" as arbitrary sets of formulas. However, I argue that there are difficulties interpreting propositional quantifiers in an impossible worlds semantics for embedding expressions like counterfactuals. Instead, I build off Kocurek and Jerzak's (2021) hyperintensional semantics, which introduces a shiftable parameter that provides the interpretation of the connectives (Muñoz, 2020; Muskens, 1991; Williamson, 2009). I will show how this theory can be expanded to a hyperintensional semantics with propositional quantifiers that fits more naturally with hyperlogic. As we will see, this semantics retains the flexibility of logically impossible worlds without the technical difficulties they bring.

Throughout, I will be working within a classical framework in the background. This is not because I think classical logic is the "one true" logic or that hyperlogic must be developed within a classical setting to be intelligible (Meyer and Routley, 1977).[1] The task of developing an adequate semantics for logic talk arises independently of which logic is *actually* correct (if any). My goal, rather, is to demonstrate one general strategy for developing such a semantics. The choice of a classical background logic is just a convenient starting point.

The plan is as follows. In Sect. 2, I examine the regimentation problem in more detail and present some desiderata that any adequate solution to it should satisfy. In Sect. 3, I propose the language of hyperlogic as a solution to the regimentation problem. In Sect. 4, I introduce the impossible worlds approach as a solution to the hyperintensionality problem and argue it faces difficulties interpreting propositional quantifiers. In Sect. 5, I show how to expand Kocurek and Jerzak's (2021) hyperconvention semantics into a semantics for hyperlogic that can more adequately interpret propositional quantifiers. I conclude in Sect. 6 with some ways we might extend hyperlogic to cover a broader range of phenomena.

---

[1] Indeed, this project is in many ways connected to work in relevant logic (Dunn and Restall, 2002; Mares, 2004, 2020). I view hyperlogic as a step towards developing what Routley (2019) calls a "universal logic": a logic that's "applicable in every situation whether realised or not, possible or not" (see Nolan (2018) for a critical overview of Routley's program). The approach taken here will be different from Routley's in that I aim to develop a universal logic within a classical framework, rather than revise the base logic. Other choices of background logic may be equally (or even more) fruitful as starting points (Girard and Weber, 2015; Weber, 2014; Weber et al., 2016). It would be also worth investigating how the hybrid aspects of hyperlogic change in nonclassical settings; see Braüner (2006, 2011); Braüner and de Paiva (2006); Chadha et al. (2006) for work on intuitionistic hybrid logic, as well as Standefer (2020) for work on relevant logics for 'actually'.

2 The regimentation problem

Let's start with the regimentation question: how do we regiment *metalogical* claims so that they can be interpreted in the *object* language?

To illustrate the problem, consider again (7)–(9). To provide a semantics for these sentences, we first need to regiment them in a formal language—one with a belief operator B, a counterfactual $\square\!\!\rightarrow$, or an epistemic modal $\diamondsuit_e$ respectively. We can partly regiment these claims as follows:

$$B(\text{Intuitionistic logic is correct})$$
$$(l \wedge \neg l) \,\square\!\!\rightarrow\, \neg(\text{The law of noncontradiction holds})$$
$$\diamondsuit_e \,\neg\,\text{Classical logic is correct}$$

To assign these sentences truth conditions, however, we need to finish the regimentation: we need to formalize the embedded sentences as some formulas (e.g., il, lnc, and cl respectively), and then write down a semantic clause for each. But how? What semantic clause do we give such formulas?

One way to avoid this problem is to formalize metalogical claims as distinct atomics p_{il}, p_{lnc}, and p_{cl}. The formal interpretation of these atomics is simply determined by a valuation function provided by a model. Then (7)–(9) become:

$$B\,p_{il}$$
$$(l \wedge \neg l) \,\square\!\!\rightarrow\, \neg\, p_{lnc}$$
$$\diamondsuit_e \,\neg\, p_{cl}$$

This atomic regimentation works well enough for many purposes. But it has two main disadvantages. First, it ignores the internal logical structure of metalogical claims. For example, metalogical claims can have quantificational structure, as illustrated by (3):

(3)   Some contradiction entails everything.

The reasons why we can't accurately regiment such quantificational structure with atomic formulas are familiar (e.g., they can't distinguish de dicto from de re sentences). Similarly, the atomic regimentation is blind to the structure of according-to claims, as in (4):

(4)   According to intuitionistic logic, the law of excluded middle fails.

This insensitivity to logical form is undesirable. It would be better if the syntax of our formal language at least somewhat matched (even if not perfectly) the actual syntactic structure of such talk.

Second, it does not capture differences in patterns of reasoning with metalogical claims. For example, consider the following argument:

(13)   a.   According to intuitionistic logic, the law of excluded middle fails.
       b.   Intuitionistic logic is correct.
       c.   Therefore, the law of excluded middle fails.

This seems like impeccable reasoning regardless of what one thinks about the premises. By contrast, the following argument is bad:

(14)   a.   According to intuitionistic logic, the law of excluded middle fails.
       b.   Quantum logic is correct.
       c.   Therefore, the law of excluded middle fails.

Yet the difference between (13) and (14) cannot be captured on the current approach, as both arguments are represented as having the form:

$$p, q \therefore r$$

For another example, contrast (15), which seems like good reasoning, and (16), which does not:

(15)   a.   Everything is entailed by every contradiction.
       b.   Therefore, everything is entailed by some contradiction (or other).

(16)   a.   Everything is entailed by some contradiction (or other).
       b.   Therefore, everything is entailed by every contradiction.

After all, given that there are contradictions, if all contradictions entail everything, then some do. So (15) seems like good reasoning. But there are logics on which everything is entailed by some contradiction or other (e.g., everything is entailed by the result of conjoining it with its negation) even though contradictions don't entail everything. So (16) does not seem like good reasoning.

    Let me be clear: I am not claiming that (14) and (16) are bad because they're *invalid*. On a standard (classical) notion of validity, these arguments are all trivially valid since they have either an impossible premise or a necessary conclusion.[2] My point, rather, is that there is some good-making feature that (13) and (15) have and that (14) and (16) lack. Very roughly, (13) and (15) seem to be good forms of reasoning regardless of one's views about which logic is correct, whereas (14) and (16) do not. Even if we do not call this feature "validity", it demands explanation nonetheless.

---

[2] Of course, (14) and (16) may be invalid on a stricter notion of validity. For example, these arguments could be deemed invalid in the sense of relevant logic (Anderson and Belnap, 1975; Dunn and Restall, 2002; Mares, 2004, 2020). Again, my assumption of a classical background logic is just a starting point. Alternative choices of background logic might lead to different views on what exactly is wrong with these arguments (see Weber (2014); Girard and Weber (2015); Weber et al. (2016) for semantic analyses within nonclassical metatheory). My claim is just that these arguments seem bad *in some sense* that is not accounted for by saying they're invalid on a classical conception of validity.

3 The language of hyperlogic

I now develop a solution to the regimentation problem called **hyperlogic**. The purpose of this section is simply to introduce the language of hyperlogic and explain its intended interpretation. I'll give a semantics in Sect. 5.

3.1 The language

To introduce the language of hyperlogic, we start with the language of propositional modal logic, which I'll call $\mathcal{L}$. The language contains an infinite stock of propositional variables $\mathsf{Prop} = \{p_1, p_2, p_3, \ldots\}$, boolean connectives ($\neg$, $\wedge$, $\vee$, $\rightarrow$), and modal operators ($\Box$, $\Diamond$).[3] The syntax is given below in Backus-Naur form:

$$\phi ::= p \mid \neg\phi \mid (\phi \wedge \psi) \mid (\phi \vee \psi) \mid (\phi \rightarrow \psi) \mid \Box\phi \mid \Diamond\phi.$$

To get the language of hyperlogic, we make three additions to $\mathcal{L}$.

First, we add a left-multigrade operator $\vartriangleright$ representing entailment. It is "left-multigrade" in that it takes a finite (possibly empty) list of formulas on the left and a formula on the right.[4] Informally, we can read $\phi_1, \ldots, \phi_n \vartriangleright \psi$ as "$\phi_1, \ldots, \phi_n$ entail ψ". Similarly, we can read $\vartriangleright \phi$ as "ϕ is valid". The notion of entailment that $\vartriangleright$ represents need not be solely logical per se. In principle, it can be interpreted as analytic, a priori, metaphysical, or whatever other notion of entailment one employs in characterizing validity for arguments. Thus, we can represent the claim that my shirt being red entails its being colored as $r \vartriangleright c$. Like any other operator, the interpretation of $\vartriangleright$ is something to be specified by a model.

Second, we'll introduce propositional quantifiers ($\forall p$, $\exists p$) that bind into sentence position.[5] This allows us to represent quantificational structure more faithfully. For example, here is how we can regiment (3):

$$\exists p \, \forall q((p \wedge \neg p) \vartriangleright q).$$

Furthermore, we can now regiment laws of logic as universally quantified sentences.[6] Thus, the law of excluded middle can be regimented as follows:

$$\forall p(\vartriangleright(p \vee \neg p)).$$

Third, we will introduce operators borrowed and modified from hybrid logic. Hybrid logic is an extension of modal logic with terms denoting individual worlds. These terms act both as special atomics that are true at exactly one world and as arguments

---

[3] We could add other connectives to the language, including those that arise in nonclassical logics such as "intensional" conjunction and disjunction operators from relevant logic (Dunn and Restall, 2002) or exponentials from linear logic (Di Cosmo and Miller, 2019). But I will stick with the standard "boolean" connectives for ease of exposition.

[4] We could make $\vartriangleright$ right-multigrade also to account for multiple-conclusion logics, but we'll set this complication aside.

[5] For more on propositional quantifiers, see Bull (1969); Fine (1969, 1970); Kaplan (1970); Kaminski and Tiomkin (1996); Holliday (2017); Holliday and Litak (2018); Fritz (2020).

[6] See Williamson (2013); Bacon (2020) for related approaches.

for certain operators.[7] Specifically, hybrid logic adds the following to propositional modal logic:

(i) two new kinds of atomic formulas, viz., **state variables** $s_1, s_2, s_3, \ldots$ and **state nominals** $n_1, n_2, n_3, \ldots$, which are true at exactly one world;

(ii) an operator $@_\sigma$ ('according to σ,...') for each state term σ, which resets the current world of evaluation to be the world denoted by σ

(iii) an operator $\downarrow s$ ('where s stands for the current world,...') for each state variable s, which resets the value of s to be the current world of evaluation.

Hybrid logic is useful for, among other things, modeling temporal language, which often contains names for specific times, e.g., 'It's 3 pm' or 'At 3 pm, it'll rain' (Yanovich, 2015). The idea here is to extend our language with hybrid operators for interpretations of the base language, including the connectives ($\neg$, $\wedge$, etc.), rather than for individual worlds.[8]

Putting this all together, here is the full language of hyperlogic, which I'll call $\mathcal{H}$. We introduce two new sets of atomic formulas: $\mathsf{IVar} = \{i_1, i_2, i_3, \ldots\}$ (interpretation variables) and $\mathsf{INom} = \{l_1, l_2, l_3, \ldots\}$. An **interpretation term** is a member of $\mathsf{ITerm} := \mathsf{IVar} \cup \mathsf{INom}$. We use $\iota_1, \iota_2, \iota_3, \ldots$ for interpretation terms. The syntax of $\mathcal{H}$ is given as follows:

$$\phi ::= p \mid \iota \mid \neg\phi \mid (\phi \wedge \psi) \mid (\phi \vee \psi) \mid (\phi \rightarrow \psi) \mid \Box\phi \mid \Diamond\phi \mid$$
$$(\phi_1, \ldots, \phi_n \rhd \psi) \mid \forall p\,\phi \mid \exists p\,\phi \mid @_\iota \phi \mid \downarrow i.\phi.$$

For instance, let *cl* stand for classical logic and let *il* stand for intuitionistic logic. Then we can regiment (1) and (4) as follows:[9]

(1)   Classical logic is correct.

$$cl$$

(4)   According to intuitionistic logic, the law of excluded middle fails.

$$@_{il}\,\neg\,\forall p(\rhd(p \vee \neg p))$$

In sum, here is how we can regiment logic talk in hyperlogic:

- $\ulcorner l$ is correct$\urcorner$ is regimented using l as an interpretation nominal.
- $\ulcorner$according to l, $\phi\urcorner$ is regimented as $@_l\,\phi$.
- $\ulcorner\phi_1, \ldots, \phi_n$ entail $\psi\urcorner$ is regimented as $(\phi_1, \ldots, \phi_n \,\triangleright\, \psi)$.
- $\ulcorner$it is a law that $\phi(q_1, ..., q_n)\urcorner$ is regimented as $\forall q_1 \cdots \forall q_n\,\phi(q_1, \ldots, q_n)$.

3.2 Solving the regimentation problem

Recall, there are two issues with the atomic regimentation: it does not faithfully represent the internal structure of metalogical claims, and it does not capture the relative goodness of patterns of reasoning with such claims. We are now in a position to see how hyperlogic addresses these concerns.

First, the language of hyperlogic clearly provides us with a more faithful representation of the internal syntactic structure of logic talk. We've already seen how propositional quantifiers allow us to represent quantificational structure and how hybrid operators allow us to represent the common structure in according-to sentences. This does not mean the language of hyperlogic matches perfectly with the syntactic structure of logic talk. But it does far better than the atomic regimentation.

Second, we can now explain the relative goodness of patterns of reasoning with logic talk. For example, consider (13) and (14) again:

(13) a. According to intuitionistic logic, the law of excluded middle fails.
      b. Intuitionistic logic is correct.
      c. Therefore, the law of excluded middle fails.

$$@_{il}\,\neg\forall p(\triangleright(p \vee \neg\,p)), \; il \; \therefore \; \neg\forall p(\triangleright(p \vee \neg\,p))$$

(14) a. According to intuitionistic logic, the law of excluded middle fails.
      b. Quantum logic is correct.
      c. Therefore, the law of excluded middle fails.

$$@_{il}\,\neg\forall p(\triangleright(p \vee \neg\,p)), \; ql \; \therefore \; \neg\forall p(\triangleright(p \vee \neg\,p))$$

Intuitively, the difference between (13) and (14) is that the former but not the latter has the form of a good inference regardless of one's views about logic. Specifically, it has the following form, which seems to be impeccable reasoning even if the premises are false:

$$@_l\,\phi, \; l \; \therefore \; \phi$$

Similarly, (15) has the logical form of a good inference regardless of one's views about logic, while (16) does not.

(15) a. Everything is entailed by every contradiction.
      b. Therefore, everything is entailed by some contradiction (or other).

$$\forall p \, \forall q ((q \wedge \neg q) \rhd p) \; \therefore \; \forall p \, \exists q ((q \wedge \neg q) \rhd p)$$

(16) a. Everything is entailed by some contradiction (or other).
      b. Therefore, everything is entailed by every contradiction.

$$\forall p \, \exists q ((q \wedge \neg q) \rhd p) \; \therefore \; \forall p \, \forall q ((q \wedge \neg q) \rhd p)$$

Up to this point, I've refrained from spelling out what "regardless of one's views about logic" means. This will be made more precise later (Sect. 5.4). To foreshadow, we can define two notions of validity in hyperlogic: a classical notion and a "universal" notion. An argument is **classically valid** if it's truth-preserving on a classical interpretation of the connectives. An argument is **universally valid** if it's truth-preserving on any interpretation of the connectives. The sense in which (13) and (15) are "good", then, is that both arguments are universally valid. By contrast, (14) and (16), though classically valid (for trivial reasons), are not universally valid.

With hybrid operators, we can move back and forth between classical validity and universal validity.

- $\phi_1, \ldots, \phi_n \therefore \psi$ is universally valid iff $@_l \phi_1, \ldots, @_l \phi_n \therefore @_l \psi$ is classically valid, where l does not occur in $\phi_1, \ldots, \phi_n, \psi$.
- $\phi_1, \ldots, \phi_n \therefore \psi$ is classically valid iff $cl, \phi_1, \ldots, \phi_n \therefore \psi$ is universally valid, where cl is a classical nominal.

So there's no need to settle which notion of validity is the "true" notion (though there can still be fruitful debate over which is better for which purpose). Both notions of validity can be seen as virtues of an argument.

4 Impossible worlds

Thus far, we've only focused on the regimentation problem. Now let's return to the hyperintensionality problem. While one could try to explain away hyperintensionality by appeal to pragmatics, I will simply assume that hyperintensionality is generally a semantic phenomenon. The challenge, then, is to develop a hyperintensional semantics for the relevant embedding expressions (e.g., attitude verbs, counterfactuals, etc.), i.e., a semantics on which necessarily (even logically) equivalent sentences are not intersubstitutable *salva veritate*.

The standard semantic approach to dealing with hyperintensionality is what I'll call **the impossible worlds approach**.[10] This approach modifies intensional semantic

---

[10] See Nolan (1997); Vander Laan (2004); Krakauer (2012); Brogaard and Salerno (2013); Jago (2014); Kment (2014); Berto et al. (2018); Berto and Jago (2019); French et al. (2020). There are other approaches to hyperintensionality that I do not have space to consider here. See Fine (2012); Schaffer (2016); Wilson (2018); Leitgeb (2019) for some prominent examples. For an overview of different approaches to hyperintensionality, see Berto and Nolan (2021).

theories by introducing "impossible worlds" where necessarily equivalent sentences can be distinguished.

In this section, I will consider whether the impossible worlds semantics can provide us with a simple solution to the hyperintensionality problem. I argue that the standard implementation of this semantics runs into problems when we try to extend it with languages with propositional quantifiers, which we've seen are crucial to solving the regimentation problem.

4.1 Impossible worlds semantics

The impossible worlds approach starts by expanding the notion of a "world" so as to include both possible and impossible worlds. We can think of a world as an ersatz entity—say, a set of formulas (Nolan, 1997). Truth-at-a-world reduces to membership: ϕ is true at an ersatz world w iff $\phi \in w$. A "possible" world is just a maximal compossible set of formulas, whereas an "impossible" world is a set that is not both maximal and compossible. So to transform an intensional semantics into a hyperintensional one, we simply replace possible worlds with worlds in this broader sense.

To illustrate, let's see how this approach applies to counterfactuals.[11] According to the standard "selection" semantics, a counterfactual is true iff at all the selected ("closest") possible worlds where the antecedent is true, the consequent is true. This is implemented formally by introducing a selection function f, which takes a set of possible worlds X and a possible world w, and "selects" a set of possible worlds $f(X, w)$ as the "closest" or "most similar" X-worlds to w (Stalnaker, 1968; Lewis, 1973). Thus, a counterfactual $\phi \,\square\!\!\rightarrow\, \psi$ is true at a possible world w iff $f(\llbracket \phi \rrbracket, w) \subseteq \llbracket \psi \rrbracket$ (where $\llbracket \phi \rrbracket = \{v \mid \phi \text{ is true at } v\}$).

Applied to this semantics, the impossible worlds approach allows selection functions to select sets containing impossible worlds. More precisely, we start with an extension of $\mathcal{L}$ with counterfactuals:

$$\phi ::= p \mid \neg \phi \mid (\phi \wedge \psi) \mid (\phi \vee \psi) \mid (\phi \rightarrow \psi) \mid \square \phi \mid \Diamond \phi \mid (\phi \,\square\!\!\rightarrow\, \psi).$$

An impossible worlds model is a quadruple $\mathcal{I} = \langle W, P, f, V \rangle$, where W is a nonempty set of worlds, $P \subseteq W$ is a nonempty set of possible worlds, $f : \wp W \times W \rightarrow \wp W$ is a selection function (perhaps satisfying certain constraints, e.g., $f(X, w) \subseteq X$), and V is a valuation function where:

---

[11] My implementation of the impossible worlds approach to counterfactuals is largely in line with that of French et al. (2020). See also Mares (1997); Nolan (1997); Brogaard and Salerno (2013); Kment (2014); Berto et al. (2018) for other examples. This semantics is closely related to the Routley-Meyer semantics for relevant logics; see Mares (2004) for an overview.

(i) $V(p, w) \in \{0, 1\}$ for each $w \in P$ and each $p \in \mathsf{Prop}$
(ii) $V(\phi, w) \in \{0, 1\}$ for each $w \in \overline{P}$ and each ϕ.[12]

Satisfaction ($\Vdash$) is defined as follows. If $w \in \overline{P}$, then $\mathcal{I}, w \Vdash \phi$ iff $V(\phi, w) = 1$. If $w \in P$, then $\Vdash$ is defined recursively:

$$\begin{aligned}
\mathcal{I}, w \Vdash p \quad &\Leftrightarrow \quad V(p, w) = 1 \\
\mathcal{I}, w \Vdash \neg \phi \quad &\Leftrightarrow \quad \mathcal{I}, w \nVdash \phi \\
\mathcal{I}, w \Vdash \phi \wedge \psi \quad &\Leftrightarrow \quad \mathcal{I}, w \Vdash \phi \text{ and } \mathcal{I}, w \Vdash \psi \\
\mathcal{I}, w \Vdash \phi \vee \psi \quad &\Leftrightarrow \quad \mathcal{I}, w \Vdash \phi \text{ or } \mathcal{I}, w \Vdash \psi \\
\mathcal{I}, w \Vdash \phi \rightarrow \psi \quad &\Leftrightarrow \quad \mathcal{I}, w \Vdash \phi \text{ only if } \mathcal{I}, w \Vdash \psi \\
\mathcal{I}, w \Vdash \Box \phi \quad &\Leftrightarrow \quad \text{for all } v \in P: \mathcal{I}, v \Vdash \phi \\
\mathcal{I}, w \Vdash \Diamond \phi \quad &\Leftrightarrow \quad \text{for some } v \in P: \mathcal{I}, v \Vdash \phi \\
\mathcal{I}, w \Vdash \phi \,\Box\!\!\rightarrow \psi \quad &\Leftrightarrow \quad f(\llbracket \phi \rrbracket^{\mathcal{I}}, w) \subseteq \llbracket \psi \rrbracket^{\mathcal{I}}.
\end{aligned}$$

Consequence is defined as preservation of truth over *possible* worlds: $\Gamma \vDash \phi$ if for all $\mathcal{I} = \langle W, P, f, V \rangle$ and all $w \in P$, if $\mathcal{I}, w \Vdash \Gamma$, then $\mathcal{I}, w \Vdash \phi$. Since possible worlds are all classical, this notion of consequence is an extension of classical **S5**. But counterfactuals are hyperintensional: even if ϕ and ψ are *logically* equivalent (i.e., true at all the same possible worlds in every model), there can still be impossible worlds in some models where ϕ and ψ differ in truth value.

Note, there are no recursive clauses for how the truth of a complex formula at an impossible world is related to the truth of its constituents. This is because, in order to model logic talk, some impossible worlds need to be *logically* impossible.[13] And when it comes to the logically impossible, anything goes: there are impossible worlds where $\wedge$ and $\vee$ are equivalent, where $\neg$ is a redundant operator, or even where everything is true. Such worlds are strange, to be sure, but nothing rules them out—in fact, such worlds are needed to fully model logic talk. So there simply is no single set of recursively defined truth conditions that applies to every logical impossibility. This means truth at impossible worlds must be determined *by fiat* via the valuation function. As we'll now see, this feature of the impossible worlds semantics causes it difficulties.

4.2 Quantifier problems

To provide a full semantics for metalogical claims, the impossible worlds semantics needs to be capable of being extended to a language that's capable of solving the

---

[12] Following Tanaka (2018), this clause could be generalized so that formulas can take multiple (or no) truth values at an impossible world. This additional complication does not affect the main arguments below, however, so I will set this aside.

[13] As Tanaka (2018) and Sandgren and Tanaka (2020) note, there are two senses in which a world can be logically impossible: it can be logically different, in that the laws of logic differ from the actual laws, or it can contain a violation of a logical law (Kocurek and Jerzak, 2021: pp. 22–23 on actual vs. counterfactual logical impossibility). Here, by "logically impossible", I mean to neutrally refer to either notion of logical impossibility. The semantics provided in Sect. 5 can represent either kind of logical impossibility.

regimentation problem, such as hyperlogic. I will now present a challenge to doing this. Specifically, I argue that the impossible worlds semantics, as stated, faces difficulties adequately interpreting propositional quantifiers.

The standard semantics for propositional quantifiers (**S5π+**) interprets them as quantifiers ranging over sets of worlds (Fine, 1970). So where $\mathcal{M} = \langle W, V \rangle$ is a possible worlds model and where $X \subseteq W$, let V_X^p be like V except that $V_X^p(p, w) = 1$ iff $w \in X$ and let $\mathcal{M}_X^p = \langle W, V_X^p \rangle$. Then here is the semantics for propositional quantifiers in **S5π+**:

$$\mathcal{M}, w \Vdash \forall p\, \phi \;\; \Leftrightarrow \;\; \text{for all } X \subseteq W \colon \mathcal{M}_X^p, w \Vdash \phi$$
$$\mathcal{M}, w \Vdash \exists p\, \phi \;\; \Leftrightarrow \;\; \text{for some } X \subseteq W \colon \mathcal{M}_X^p, w \Vdash \phi.$$

In words, $\forall p\, \phi$ is true iff ϕ is true on every interpretation of p, where "an interpretation of p" is just an assignment of p to some possible-worlds proposition (i.e., a set of possible worlds).

Importing this into the impossible worlds semantics, we get something like the following. Where $\mathcal{I} = \langle W, P, f, V \rangle$, let V_X^p be like V except that $V_X^p(p, w) = 1$ iff $w \in X$, and let $\mathcal{I}_X^p = \langle W, P, f, V_X^p \rangle$. Then:

$$\mathcal{I}, w \Vdash \forall p\, \phi \;\; \Leftrightarrow \;\; \text{for all } X \subseteq W \colon \mathcal{I}_X^p, w \Vdash \phi$$
$$\mathcal{I}, w \Vdash \exists p\, \phi \;\; \Leftrightarrow \;\; \text{for some } X \subseteq W \colon \mathcal{I}_X^p, w \Vdash \phi.$$

Thus, propositional quantifiers have essentially the same meaning as in **S5π+**, except now p can denote a more fine-grained proposition, viz., a set of worlds that may include some impossible worlds, too.[14]

The problem with this proposal is that V_X^p only changes the truth of p at impossible worlds, not the truth of complex formulas involving p. To see why this is an issue, consider the following inference (where $q \neq p$):

$$\exists p((p \wedge \neg p) \,\Box\!\!\rightarrow q) \;\; \therefore \;\; \forall p((p \wedge \neg p) \,\Box\!\!\rightarrow q).$$

Intuitively, this is invalid: just because *one* contradiction (e.g., $q \wedge \neg q$) counterfactually implies q, it does not follow that *all* contradictions do (at least, not if we want counterfactuals to be hyperintensional). Yet, according to the semantics for propositional quantifiers above, it is valid precisely because changing the interpretation of p does not change the interpretation of $p \wedge \neg p$, i.e., $V_X^p(p \wedge \neg p, w) = V(p \wedge \neg p, w)$.[15]

---

[14]　Nolan (1997: p. 563) argues that not every set of impossible worlds counts as a proposition. Thus, we may want to include in our models a domain of sets of worlds the quantifiers can range over. (So the semantics would be closer to **S5π**.) This won't affect the substantive points in what follows, however.

[15]　Proof: suppose $\mathcal{I}, w \Vdash \exists p((p \wedge \neg p) \,\Box\!\!\rightarrow q)$ where $\mathcal{I} = \langle W, P, f, V \rangle$ and $w \in P$. Let $X \subseteq W$. Since possible worlds are classically consistent, $[\![p \wedge \neg p]\!]^{\mathcal{I}} \subseteq \overline{P}$. But since V_X^p doesn't reassign the truth of $p \wedge \neg p$, $V_X^p(p \wedge \neg p, v) = V(p \wedge \neg p, v)$ for all $v \in \overline{P}$. Hence, $[\![p \wedge \neg p]\!]^{\mathcal{I}_X^p} = [\![p \wedge \neg p]\!]^{\mathcal{I}}$, and so $f([\![p \wedge \neg p]\!]^{\mathcal{I}_X^p}, w) = f([\![p \wedge \neg p]\!]^{\mathcal{I}}, w)$. Moreover, since $q \neq p$, $V_X^p(q, w) = V(q, w)$, and so $[\![q]\!]^{\mathcal{I}_X^p} = [\![q]\!]^{\mathcal{I}}$. Therefore, if $f([\![p \wedge \neg p]\!]^{\mathcal{I}_X^p}, w) \subseteq [\![q]\!]^{\mathcal{I}_X^p}$ for any $X \subseteq W$, then it holds for all X. In other words, $\exists p((p \wedge \neg p) \,\Box\!\!\rightarrow q)$ entails $\forall p((p \wedge \neg p) \,\Box\!\!\rightarrow q)$.

This problem suggests we need to define V_X^p so that it differs from V not just on the interpretation p but also on the interpretation of complex formulas involving p. The trouble is that it is not clear how to do this. Again, when it comes to the logically impossible, anything goes. How do we know whether some arbitrary impossible world is supposed to be one that allows contradictions? And even if it does, how do we know which interpretations of p generate contradictions? Without further guidance, we simply have no way of knowing how we are allowed to interpret $p \wedge \neg p$ at an impossible world given an arbitrary interpretation of p.

We might try to solve this problem by simply quantifying over all the ways of interpreting complex formulas from interpretations of their constituents. More precisely, let $V' \sim_p V$ iff V' is like V except for how it interprets formulas with p. Then the proposal is:

$$\mathcal{I}, w \Vdash \forall p\, \phi \;\; \Leftrightarrow \;\; \text{for all } V' \sim_p V : \mathcal{I}', w \Vdash \phi$$
$$\mathcal{I}, w \Vdash \exists p\, \phi \;\; \Leftrightarrow \;\; \text{for some } V' \sim_p V : \mathcal{I}', w \Vdash \phi.$$

Unfortunately, this won't work either. Consider the following argument (where p doesn't occur free in ϕ):

$$\neg \diamond \phi \; \therefore \; \exists p (\phi \,\square\!\!\rightarrow (p \wedge \neg p)).$$

Clearly, this argument is not valid: there is no contradiction that would be true if intuitionistic logic were correct. But on these truth conditions for the quantifiers, it is valid: assuming ϕ is impossible, our witness can be a V' like V in every way except $V'(p \wedge \neg p, u) = 1$ iff $u \in f(\llbracket \phi \rrbracket, w)$.

The problem is that some impossible worlds are meant to be governed by specific nonclassical logics. Intuitively, we only want to quantify over propositions that conform to the logic of a world. On the current truth conditions, though, there are no constraints on which interpretations of complex formulas we can quantify over. So we can quantify over interpretations of formulas that violate those logics at those worlds.

To solve this problem, we need to equip worlds with something like a *semantics* for the connectives, i.e., a set of rules that tell us how to interpret complex formulas given an interpretation of their constituents. That way, when we reinterpret the atomics, the semantics will automatically give guidance for how that changes the interpretation of complex formulas. As we'll see, that is exactly what the hyperconvention semantics provides.

5 The hyperconvention semantics

I now present an alternative semantics, **the hyperconvention semantics**, for hyperlogic. This semantics is based on Kocurek and Jerzak's (2021) "logical expressivist" semantics for counterlogicals, which relativizes truth to a shiftable parameter (a "hyperconvention") that provides the interpretation of the logical connectives (Muñoz, 2020; Muskens, 1991; Williamson, 2009). I build off this semantics to develop a seman-

tics for hyperlogic. In Sect. 5.5, I'll show how this semantics avoids the problems plaguing the impossible worlds semantics from Sect. 4.2.

5.1 Hyperconventions

Kocurek and Jerzak are primarily concerned with **counterlogicals**, i.e., counterfactuals with logically impossible antecedents. They argue that counterlogicals can only be semantically nonvacuous if we are allowed to shift the interpretation of the logical connectives. This because, they argue, if we hold fixed what words like 'not' and 'or' *actually* mean when evaluating (17), then we would expect (for the usual Kripkean reasons) that (17) is trivially true, whereas it clearly seems false.

(17)   If intuitionistic logic were the correct logic, then either the continuum hypothesis would be true or it would not be true.

To capture this thought, they introduce the notion of a **hyperconvention** (inspired by Gibbard's (2003) notion of a hyperplan). Hyperconventions provide an interpretation of the modal language $\mathcal{L}$, including the logical connectives. They model the interpretation of a connective as an operation on sets of worlds—intuitively, an ordinary possible worlds proposition. For instance, a hyperconvention c will map $\neg$ to a function $c(\neg)$ from propositions to propositions. They then relativize truth to a world and a hyperconvention. So $\neg\phi$ is true at w according to c iff $w \in c(\neg)(X)$ where X is the proposition expressed by ϕ according to c.

Kocurek and Jerzak motivate this semantics by situating it in a more general philosophy of logic (what they call "logical expressivism") on which purely logical claims are expressions of commitments to logics. Expressions of logical commitments are meant to contrast with factual claims about "the one true logic" or descriptive claims about how speakers actually use logical vocabulary. Rather, their picture is that speakers have a range of factual beliefs as well as commitments to using language in a certain fashion. Such commitments are not automatically determined by the speaker's factual beliefs, but can be informed by them (like an intention or a plan).[16]

In what follows, I want to build off this idea to develop a semantics for hyperlogic. To be clear, we do not need to endorse Kocurek and Jerzak's logical expressivism to do this. We can, if we wish, interpret the hyperconvention parameter as a purely factual one (e.g., determined by the way speakers at a world use logical vocabulary). In employing the hyperconvention semantics, we need not take a stand on these controversial issues. The crucial idea, for our purposes, is to separate out the semantic contributions of the interpretation of the connectives from the world parameter so that the two can be shifted independently.

Let's define the notion of a hyperconvention more formally.[17]

---

[16] This view can be seen as a species of logical pluralism (Beall and Restall, 2006), though the latter is a broader category. Kocurek and Jerzak (2021: p. 17) argue that logics are effectively semantic conventions governing logical vocabulary. According to logical expressivism, adopting a logic is akin to adopting a language: just as there is no such thing as "the one true" semantic convention, so too, there is no such thing as "the one true" logic.

[17] Since Kocurek and Jerzak's (2021) semantics does not include propositional quantifiers or an entailment operator, we need to generalize their notion of a hyperconvention.

Definition 1 (*Hyperconvention*) Let π be a new symbol not in the signature of $\mathcal{H}$. A **hyperconvention** over a set W is a function c with domain $\{\pi, \neg, \Box, \Diamond, \wedge, \vee, \rightarrow, \rhd\} \cup$ Prop such that:

(i) $c(\pi) \subseteq \wp\, W$ (the **proposition space** of c, written as π_c)
(ii) $c(p) \in \pi_c$ for all $p \in$ Prop
(iii) $c(\bullet)\colon \pi_c \to \pi_c$ where $\bullet \in \{\neg, \Box, \Diamond\}$
(iv) $c(\circ)\colon \pi_c^2 \to \pi_c$ where $\circ \in \{\wedge, \vee, \rightarrow\}$
(v) $c(\rhd)\colon \pi_c^{<\omega} \times \pi_c \to \pi_c$.

For readability, we may write $\star_c$ (with infix notation) instead of $c(\star)$. Let $\mathbb{H}_W$ be the set of all hyperconventions over W.

A **convention** over W is a nonempty set of hyperconventions over W. Let $\mathbb{C}_W = \wp^+ \mathbb{H}_W$ be the set of conventions over W.

Notice I have not placed any constraints on the possible interpretations of $\rhd$. We don't require $\rhd$ obey any sort of structural rule (commutativity, contraction, etc.). We don't even require $\rhd$ to be factive (so that $\rhd_c X \subseteq X$) or noncontingent (so that $X_1, \ldots, X_n \rhd_c Y$ is always either W or $\emptyset$).[18] This is mainly for the sake neutrality, as I do not want to take a stand on what entailment relations count as "genuine" logics. Ultimately, we will let the models determine which interpretations of $\rhd$, or any of the other connectives, are to be considered.

Most logics can be represented by a convention, i.e., a set of hyperconventions. In fact, it can be shown that any (finitary, single-conclusion) logic over $\mathcal{L}$ can be represented by a convention assuming W is infinite (§A). One can often represent the most commonly discussed logics using a possible worlds semantics for that logic.

To illustrate, let's consider a convention representing an intuitionistic propositional logic (I'll ignore modal operators for now). Let $\mathcal{K} = \langle W, \leq \rangle$ be an intuitionistic Kripke frame. Let $C_\mathcal{K}$ be the set of hyperconventions c satisfying the constraints below:

$$\pi_c = \big\{ X \subseteq W \,\big|\, \forall w, v \in W\colon w \in X \,\&\, w \leq v \Rightarrow v \in X \big\}$$
$$\neg_c X = \big\{ w \in X \,\big|\, \forall v \geq w\colon v \notin X \big\}$$
$$X \wedge_c Y = X \cap Y$$
$$X \vee_c Y = X \cup Y$$
$$X \rightarrow_c Y = \big\{ w \in W \,\big|\, \forall v \geq w\colon v \in X \Rightarrow v \in Y \big\}$$
$$X_1, \ldots, X_n \rhd_c Y = \{ w \in W \mid X_1 \cap \cdots \cap X_n \subseteq Y \}.$$

Then $C_\mathcal{K}$ represents the logic of $\mathcal{K}$: any propositional ϕ is intuitionistically valid over $\mathcal{K}$ iff $(\rhd_c [\![\phi]\!]^c) = W$ for each $c \in C_\mathcal{K}$ (where $[\![\phi]\!]^c = \{v \mid \phi$ is true at v according to $c\}$; see Definition 5 in Sect. 5.3).

---

[18] One reason to allow $\rhd$ to be noncontingent is to model logics such as **K3** on which there are no logical validities. It would be natural to capture this fact about **K3** with the formula $\forall p \neg @_{k3}(\rhd p)$. But if entailment facts are necessary, then this formula will be falsifiable.

It is essential to this example that π_c is not the full powerset. Otherwise, $p \triangleright \neg\neg p$ would be falsifiable in $C_{\mathcal{K}}$—that is, $c(p) \nsubseteq \neg_c \neg_c c(p)$ for some $c \in C_{\mathcal{K}}$—and so $C_{\mathcal{K}}$ would not represent intuitionistic logic.[19] It is crucial to the intuitionistic Kripke semantics that truth is persistent under $\leq$: if w satisfies ϕ and $w \leq v$, then v satisfies ϕ. The "counterexamples" to double negation introduction involve sets that are not upward closed under $\leq$. Most nonclassical logics come equipped with a view about what counts as a proposition. In order to represent a logic, one must ensure the proposition space captures the relevant notion of a proposition for that logic.

5.2 Three notions of proposition

Again, the key idea behind the hyperconvention semantics is to evaluate truth relative to a world and a hyperconvention. That way, according-to operators can shift the interpretation of the connectives by shifting the hyperconvention parameter of points of evaluation or "indices". Thus, indices need to be defined as world-hyperconvention pairs.

Definition 2 (*Index*) Given a set of hyperconventions H over W, an **index** over H is a pair $\langle w, c \rangle$ where $w \in W$ and $c \in H$. Let $\mathbb{I}_H = W \times H$ be the set of indices over H. Where $A \subseteq \mathbb{I}_H$, let $A(c) := \{w \in W \mid \langle w, c \rangle \in A\}$.

Since truth is evaluated relative to worlds and hyperconventions, there are several notions of a "proposition" in this semantics, each of which play an important role. First, there is a coarse-grained notion of a proposition as a set of worlds. World propositions act as the interpretation of propositional variables relative to a hyperconvention; operations over world propositions act as the interpretation of the connectives relative to a hyperconvention. Second, there is a more fine-grained notion of a proposition as a set of indices. Index propositions act as the refined compositional semantic values of formulas assigned by our model.

There is also a third, intermediate notion of a proposition worth discussing: sets of indices A where, relative to any hyperconvention c, the set of worlds $A(c)$ is in the proposition space of c.

Definition 3 (*Visible proposition*) Given a set of hyperconventions H over W, a **visible proposition** over H is a set of indices $A \subseteq \mathbb{I}_W$ such that $A(c) \in \pi_c$ for all $c \in \mathbb{H}_W$. Let $\mathbb{P}_H$ be the set of visible propositions over H. I'll use $X, Y, \ldots$ for world propositions, $A, B, \ldots$ for index propositions, and $P, Q, \ldots$ for visible propositions.

Observe that if H is nonempty, then $\mathbb{P}_H$ is nonempty, since the proposition space of any hyperconvention is nonempty. In particular, $c(p)$ is always a member of π_c by

---

[19] Proof: let $w \in W$ be such that there is a $w' > w$ in W. Let $c(p) = \{w\}$. Then:

$$\neg_c \neg_c \{w\} = \{v \in W \mid \forall u \geq v \colon u \notin \neg_c \{w\}\}$$
$$= \{v \in W \mid \forall u \geq v \, \exists z \geq u \colon z \in \{w\}\}$$
$$= \{v \in W \mid \forall u \geq v \colon w \geq u\}.$$

But $w \notin \neg_c \neg_c \{w\}$, since $w' \geq w$ but $w \ngeq w'$. Thus, $\{w\} \nsubseteq \neg_c \neg_c \{w\}$.

clause (ii) of Definition 1. Thus, the index proposition P_p, where $P_p(c) = c(p)$ for all c, is visible.

Visible index propositions are the kind of proposition that the propositional quantifiers range over. We do not want quantifiers to range over all index propositions. For if they did, it would be too easy to refute a law of logic according to a nonclassical logic. To illustrate, recall the example from Sect. 5.1 of the convention $C_{\mathcal{K}}$ representing intuitionistic logic. We saw that if the proposition space of every $c \in C_{\mathcal{K}}$ were unrestricted, we could find counterexamples to double negation introduction ($p \rhd \neg \neg p$), which is intuitionistically valid. Similarly, if the domain of propositional quantification were to include every index proposition, we could find counterexamples to the *quantified* version of double negation introduction ($\forall p (p \rhd \neg \neg p)$).[20] By contrast, if we restrict the domain of propositional quantification to visible propositions, no such counterexample can be constructed.

5.3 Semantics

We are now ready to present the semantics more explicitly. The models of our semantics must specify (i) a set of states (or "worlds"), (ii) a domain of admissible conventions, (iii) a domain of admissible (visible) propositions, and (iv) a valuation function.[21]

Definition 4 (*Hypermodel*) A **hypermodel** is a tuple of the form $\mathcal{M} = \langle W, D_{\mathbb{C}}, D_{\mathbb{P}}, V \rangle$, where:

- $W \neq \emptyset$ is a **state space**
- $D_{\mathbb{C}} \subseteq \mathbb{C}_W$ is a **convention domain**; we also define $D_{\mathbb{H}} = \bigcup D_{\mathbb{C}}$ to be the **hyperconvention domain**
- $D_{\mathbb{P}} \subseteq \mathbb{P}_{D_{\mathbb{H}}}$ is a **proposition domain** such that:

  (i) for all $p \in \mathsf{Prop}$, $P_p \in D_{\mathbb{P}}$, where $P_p(c) = c(p)$
  (ii) for all $c \in D_{\mathbb{H}}$ and $X \in \pi_c$, there is a $P \in D_{\mathbb{P}}$ where $P(c) = X$

- V is a **valuation** such that:

  (i) $V(p) \in D_{\mathbb{P}}$
  (ii) $V(l) \in D_{\mathbb{C}}$
  (iii) $V(i) \in D_{\mathbb{C}} \cup \{\{c\} \mid c \in D_{\mathbb{H}}\}$

where x is a variable and v is a possible value for that variable, we write V_v^x for the valuation like V except that $V_v^x(x) = v$. We likewise write $\mathcal{M}_v^x$ for $\langle W, D_{\mathbb{C}}, D_{\mathbb{P}}, V_v^x \rangle$.

I have imposed two very weak requirements on the domain of propositional quantification, largely for technical convenience. First, the index proposition that picks out the interpretation of p at each hyperconvention must be included in the domain. This ensures that hyperconventions that assign different world propositions to the atom-

---

[20] Proof: just pick an index proposition whose restriction to some $c \in C_{\mathcal{K}}$ is $\{w\}$.

[21] While Kocurek and Jerzak (2021) model logics (which are the referents of interpretation terms) as hyperconventions, this proposal is too stringent for our purposes, as it would require logics to provide interpretations of every propositional variable. To avoid this concern, we can let the interpretation terms denote *conventions*, i.e., *sets* of hyperconventions (Kocurek and Jerzak, 2021, fn. 21).

ics are not indiscernible.[22] Second, each world proposition in a hyperconvention's proposition space must be picked out by some index proposition in the domain. This rules out "propositional impossibilia", i.e., world propositions that from the model's perspective do not exist.

Finally, here is the semantics:

Definition 5 (*Semantics*) Where $\star \in \{\neg, \wedge, \vee, \rightarrow, \Box, \Diamond, \rhd\}$:

$$\mathcal{M}, w, c \Vdash p \quad\Leftrightarrow\quad w \in V(p)(c)$$

$$\mathcal{M}, w, c \Vdash \iota \quad\Leftrightarrow\quad c \in V(\iota)$$

$$\mathcal{M}, w, c \Vdash \star(\phi_1, \ldots, \phi_n) \quad\Leftrightarrow\quad w \in c(\star)(\llbracket \phi_1 \rrbracket^{\mathcal{M},c}, \ldots, \llbracket \phi_n \rrbracket^{\mathcal{M},c})$$

$$\mathcal{M}, w, c \Vdash \forall p\, \phi \quad\Leftrightarrow\quad \text{for all } P \in D_\mathbb{P} : \mathcal{M}_P^p, w, c \Vdash \phi$$

$$\mathcal{M}, w, c \Vdash \exists p\, \phi \quad\Leftrightarrow\quad \text{for some } P \in D_\mathbb{P} : \mathcal{M}_P^p, w, c \Vdash \phi$$

$$\mathcal{M}, w, c \Vdash @_\iota \phi \quad\Leftrightarrow\quad \text{for all } c' \in V(\iota) : \mathcal{M}, w, c' \Vdash \phi$$

$$\mathcal{M}, w, c \Vdash \downarrow i.\phi \quad\Leftrightarrow\quad \mathcal{M}_{\{c\}}^i, w, c \Vdash \phi,$$

where $\llbracket \phi \rrbracket^{\mathcal{M},c} = \{v \in W \mid \mathcal{M}, v, c \Vdash \phi\}$.

Here is an informal statement of the semantic clauses. Propositional variables denote visible index propositions; so p is true at $\langle w, c \rangle$ iff w is a member of the world proposition assigned to p at c. Interpretation terms denote conventions, i.e., sets of hyperconventions; so ι is true at $\langle w, c \rangle$ iff c is a member of the convention denoted by ι, i.e., ι denotes a "correct" convention from c's perspective. Connectives denote intensional operations; so (e.g.) $\neg \phi$ is true at $\langle w, c \rangle$ iff (i) $\llbracket \phi \rrbracket^c$ denotes a world proposition that exists according to c, and (ii) the operation c assigns to $\neg$, when applied to $\llbracket \phi \rrbracket^c$, yields a world proposition that is true at w. Quantifiers range over interpretations of the propositional variables; so (e.g.) $\forall p\, \phi$ is true at $\langle w, c \rangle$ iff ϕ is true at $\langle w, c \rangle$ on any interpretation of p within the proposition domain. According-to operators quantify over the hyperconventions in a convention; so $@_\iota \phi$ is true at $\langle w, c \rangle$ iff ϕ is true at w on any maximally specific convention compatible with the convention denoted by ι. Finally, the binder is a device for keeping track of individual hyperconventions; so $\downarrow i.\phi$ is true at $\langle w, c \rangle$ iff ϕ is true at $\langle w, c \rangle$ when we reassign i to denote $\{c\}$.

Let me note one feature of the semantics that some may find questionable: iterated according-to operators are redundant in that $@_\iota @_\kappa \phi$ is semantically equivalent to $@_\kappa \phi$. Thus, all logics agree on what holds according to all other logics. Just speaking for myself, I do not have clear intuitions about whether this is correct. In general, it is unclear how to interpret a sentence like (18).

(18)   According to intuitionistic logic, the law of excluded middle holds according to classical logic.

---

[22] Note, however, that $V(p)$ need not be P_p; this is essential, since quantifiers need to be capable of reinterpreting atomics.

In one sense, (18) seems false: intuitionistic logic doesn't say anything about what laws hold according to classical logic. But in another sense, (18) seems true: according to intuitionistic logic, classical logic is incorrect precisely because the law of excluded middle holds according to it. In other words, if intuitionistic logic "says nothing" about what holds according to classical logic, then it "says nothing" about whether classical logic is correct.

I suspect there are simply two ways to hear (18). For simplicity, I've chosen to set this issue aside and treat iterated according-to operators as redundant. This is in line with how these operators standardly work in hybrid logic and it simplifies the formalism greatly. But I acknowledge one may want to generalize hyperlogic to allow for nonredundant iteration of according-to operators. There are ways of generalizing the framework to accomplish this, but I want to set this complication aside. I view this redundancy as an idealizing assumption that requires further investigation, rather than the final say on the logic of according-to operators.

5.4 Consequence

In the hyperconvention semantics, there are two notions of consequence we can define. There is a classical notion of consequence, i.e., truth-preservation relative to a classical interpretation of the connectives. There is also a universal notion of consequence, i.e., consequence no matter how we interpret the connectives.

To make this precise, I need to explain what a "classical" convention is. Throughout, I will use cl as a designated classical interpretation nominal.

Definition 6 (*Classical hyperconvention*) A hyperconvention c over W is **classical** if for all $X, Y \in \pi_c$:

$$\neg_c X = \overline{X} \qquad\qquad X \wedge_c Y = X \cap Y$$
$$\Box_c X = \{w \in W \mid X = W\} \qquad X \vee_c Y = X \cup Y$$
$$\Diamond_c X = \{w \in W \mid X \neq \emptyset\} \qquad X \rightarrow_c Y = \overline{X} \cup Y,$$

and for all $X_1, \ldots, X_n, Y \in \pi_c$:

$$(X_1, \ldots, X_n \rhd_c Y) := \{w \in W \mid X_1 \cap \cdots \cap X_n \subseteq Y\}.$$

A convention is **classical** if all of its member are classical. A hypermodel $\mathcal{M}$ is **classical** if $V(cl)$ is classical.

If c is classical, then the truth conditions for the connectives reduce to their classical ones and, moreover, $\rhd$ reduces to necessary implication: $(X_1, \ldots, X_n \rhd_c Y) = \Box_c((X_1 \wedge_c \cdots \wedge_c X_n) \rightarrow_c Y)$.

Definition 7 (*Consequence*) Where $\Gamma \subseteq \mathcal{H}$ and $\phi \in \mathcal{H}$:

- Γ **classically entails** ϕ, written $\Gamma \vDash \phi$, if for any classical hypermodel $\mathcal{M} = \langle W, D_{\mathbb{C}}, D_{\mathbb{P}}, V \rangle$, any $w \in W$, and any $c \in V(cl)$:

$$\mathcal{M}, w, c \Vdash \Gamma \;\Rightarrow\; \mathcal{M}, w, c \Vdash \phi.$$

- Γ **universally entails** ϕ, written $\Gamma \Vvdash \phi$, if for any classical hypermodel $\mathcal{M} = \langle W, D_{\mathbb{C}}, D_{\mathbb{P}}, V \rangle$, any $w \in W$, and any $c \in D_{\mathbb{H}}$:

$$\mathcal{M}, w, c \Vdash \Gamma \quad \Rightarrow \quad \mathcal{M}, w, c \Vdash \phi.$$

Note universal entailment is still restricted to classical hypermodels, even though it requires truth-preservation over any hyperconvention. This restriction is largely a matter of formal convenience, as it allows us to rigidly refer back to classical logic (or rather, a classical interpretation of the connectives) using *cl*. Universal entailment, however, still requires truth preservation over *any* hyperconvention, even nonclassical ones.

In Sect. 3, I noted that classical validity and universal validity can be defined in terms of one another. I now make this observation precise.

Theorem 8 (Embedding Validity). *Let* $\Gamma \subseteq \mathcal{H}$ *and* $\phi \in \mathcal{H}$. *Where* l *is an interpretation nominal, let* $@_l \Gamma = \{@_l \gamma \mid \gamma \in \Gamma\}$.

(a) Assume $l \neq cl$ *does not occur anywhere in* Γ *or in* ϕ. *Then* $\Gamma \Vvdash \phi$ *iff* $@_l \Gamma \vDash @_l \phi$.
(b) $\Gamma \vDash \phi$ *iff* $cl, \Gamma \Vvdash \phi$.

Proof The only interesting direction is the right-to-left direction of (a). Suppose $@_l \Gamma \vDash @_l \phi$ and $\mathcal{M}, w, c \Vdash \Gamma$. Let $\mathcal{M}_c^l = \langle W, D_{\mathbb{C}}, D_{\mathbb{P}}, V_c^l \rangle$, where V_c^l is exactly like V except that $V_c^l(l) = \{c\}$. By a simple induction on formulas, if χ does not contain l, then $\mathcal{M}, w, c \Vdash \chi$ iff $\mathcal{M}_c^l, w, c \Vdash \chi$. Hence, $\mathcal{M}_c^l, w, c \Vdash \Gamma$, and so $\mathcal{M}_c^l, w, c' \Vdash @_l \Gamma$ for each $c' \in V_c^l(cl)$. By supposition, $\mathcal{M}_c^l, w, c' \Vdash @_l \phi$, in which case $\mathcal{M}_c^l, w, c \Vdash \phi$. Therefore, $\mathcal{M}, w, c \Vdash \phi$. ∎

In Sect. 3, I said we could explain the relative goodness of patterns of reasoning with logic talk by appealing to a notion of consequence where the premises entail the conclusion "regardless of one's views about logic". Universal consequence can capture this notion of relative goodness. For instance, $@_\iota \phi$ and ι universally entail ϕ, which explains why inferences like (13) seem good. Thus, the logic of hyperlogic has the requisite features needed to fully resolve the regimentation problem.

5.5 Counterfactuals

Now let's look at how the hyperconvention semantics would handle counterfactuals. One simple way to do this is to combine the hyperconvention semantics in Sect. 5.3 with Kocurek and Jerzak's (2021) logical expressivist semantics for counterfactuals. On this approach, counterfactuals are allowed to shift hyperconventions, so that the antecedent and consequent may be evaluated relative to other logics.[23] Like with the standard selection semantics, we may assume that which hyperconventions the counterfactual may shift to will depend on the context (specifically, on which indices count as "closer" or "more similar" to the starting index in that context). Thus, even

---

[23] Kocurek and Jerzak argue elsewhere, following the work of Einheuser (2006), that even conditionals not about logic can shift the convention used to interpret material in their scope (Kocurek et al., 2020).

if two sentences are necessarily equivalent according to our actual conventions, the counterfactual may require taking us to a convention that distinguishes between those sentences.[24]

We can flesh this out formally as follows. Again, just for illustration, I will adopt a selection semantics for counterfactuals.

Definition 9 (*Selection hypermodel*) A **selection hypermodel** is a tuple $\mathcal{M} = \langle W, D_{\mathbb{C}}, D_{\mathbb{P}}, f, V \rangle$, where $\langle W, D_{\mathbb{C}}, D_{\mathbb{P}}, V \rangle$ is a hypermodel in the sense of Definition 4 and $f : \wp \, \mathbb{I}_{D_{\mathbb{H}}} \times \mathbb{I}_{D_{\mathbb{H}}} \to \wp \, \mathbb{I}_{D_{\mathbb{H}}}$ is a **selection function**.

Definition 10 (*Selection semantics*) Satisfaction is defined as in Definition 5 with the following additional clause:

$$\mathcal{M}, w, c \Vdash \phi \,\square\!\!\rightarrow\, \psi \;\; \Leftrightarrow \;\; f(\llbracket \phi \rrbracket^{\mathcal{M}}, w, c) \subseteq \llbracket \psi \rrbracket^{\mathcal{M}}$$

where $\llbracket \phi \rrbracket^{\mathcal{M}} = \big\{ \langle v, d \rangle \in \mathbb{I}_{D_{\mathbb{H}}} \mid \mathcal{M}, v, d \Vdash \phi \big\}$.

Simply put: $\phi \,\square\!\!\rightarrow\, \psi$ is true at $\langle w, c \rangle$ iff ψ is true at all the closest indices to $\langle w, c \rangle$ where ϕ is true. Again, we could place constraints on f (e.g., $f(A, w, c) \subseteq A$) as desired.

Kocurek and Jerzak (2021) prove that their logical expressivist semantics generates the same logic as that of the impossible worlds semantics: Γ entails ϕ in the impossible worlds semantics iff Γ classically entails ϕ in the logical expressivist semantics semantics. Since the hyperconvention semantics is an extension of the logical expressivist semantics, the same holds for the hyperconvention semantics over the quantifier-free and hybrid-free fragment (i.e., over the language $\mathcal{L}$ extended with $\square\!\!\rightarrow$). Thus, (classical) consequence in the two semantic theories coincide. In particular, both do equally well at solving the hyperintensionality problem, at least over this restricted language.

But there are two major differences between the semantic theories. First, in the hyperconvention semantics, truth is determined in a uniform, compositional manner: the semantic value of a complex formula is always a function of the semantic value of its parts. It's just that the "semantic value" of a formula is a set of world-hyperconvention pairs, rather than a set of worlds. By contrast, the semantic value of a formula in the impossible worlds semantics is a set of possible and impossible worlds. In general, two formulas with the same semantic value relative to a model may make different contributions to the semantic value of complex formulas. Thus, even if $\llbracket p \rrbracket^{\mathcal{I}} = \llbracket q \rrbracket^{\mathcal{I}}$, it does not follow that $\llbracket \neg p \rrbracket^{\mathcal{I}} = \llbracket \neg q \rrbracket^{\mathcal{I}}$. This is not necessarily a decisive objection to the impossible worlds semantics, but it is a theoretical cost.

Second, when we expand the language to include propositional quantifiers, the hyperconvention semantics seems to fare much better than the impossible worlds semantics. Earlier in Sect. 4, we saw that the impossible worlds semantics faced problems interpreting propositional quantifiers in part because impossible worlds do not contain a set of rules for determining truth. This led to the awkward result that

---

[24] Note, this does not mean that counterfactuals are "metalinguistic" in the sense of being *about* language or conventions. Rather, the semantics models hyperintensionality by allowing counterfactuals to *shift* the conventions used to interpret material in their scope. Compare: quantifiers shift variable assignments, but that doesn't mean they're "about" variable assignments.

either $\exists p((p \wedge \neg p) \square\!\!\rightarrow q)$ entails $\forall p((p \wedge \neg p) \square\!\!\rightarrow q)$, or that $\neg \Diamond \phi$ entails $\exists p(\phi \square\!\!\rightarrow (p \wedge \neg p))$.

The hyperconvention semantics overcomes this problem. Hyperconventions are essentially rules for interpreting complex formulas from their constituents. So quantifiers range not only over interpretations of p but also interpretations of complex formulas involving p. As a result, $\exists p((p \wedge \neg p) \square\!\!\rightarrow q)$ does not (classically or universally) entail $\forall p((p \wedge \neg p) \square\!\!\rightarrow q)$. But since quantifiers are restricted to visible propositions, they cannot quantify over every way of interpreting a formula. This is a restriction we saw was independently motivated even before we considered counterfactuals. As a result, $\neg \Diamond \phi$ does not (classically or universally) entail that $\exists p(\phi \square\!\!\rightarrow (p \wedge \neg p))$. Thus, the hyperconvention semantics enjoys a distinctive advantage over the impossible worlds approach.

With that said, the hyperconvention semantics does not entirely jettison the notion of an impossible world. In a way, world-hyperconvention pairs can be seen as a different model of what an impossible world is. On the standard picture, impossible worlds are ersatz entities, e.g., sets of formulas. In the hyperconvention semantics, impossible worlds are more like worlds "under an alternative description" (i.e., paired with an alternative hyperconvention). The latter determines the former—indeed, every set of $\mathcal{L}$-formulas is the truth set of some world-hyperconvention pair. But unlike the impossible worlds semantics, the hyperconvention semantics achieves this without appealings to separate truth conditions for possible and impossible cases. This allows us to capture the flexibility of impossible worlds as arbitrary sets of formulas while avoiding the challenges facing the impossible worlds semantics in Sect. 4.2.

6 Conclusion

We started with two challenges for developing a semantics for logic talk. First, there's a regimentation problem: how do we regiment logic talk in the object language so as to be compositionally embeddable? To this question, I proposed hyperlogic as a solution. I argued that it was more satisfactory as a language for regimenting logic talk than the brute-force atomic regimentation.

Second, there's a hyperintensionality problem: how do we interpret logic talk so that embedding expressions can discern classically equivalent sentences? To this question, I proposed the hyperconvention semantics as a solution. We saw that this semantics avoids the problems plaguing the more common impossible worlds semantics. Yet it does so in a way that still preserves the flexibility of impossible worlds.

In closing, I mention two ways in which hyperlogic can be fruitfully generalized that would be worth further investigation in future research. First, there's the question of how to extend hyperlogic with attitude verbs, such as a belief operator B. A natural proposal would be to add an accessibility relation R_B between indices to hypermodels and then interpret B as follows:[25]

---

[25] We could remove relativization to c and say R_B relates worlds to indices. Either approach would be fine for our purposes.

$$\mathcal{M}, w, c \Vdash \mathsf{B}\phi \quad \Leftrightarrow \quad \text{for all } \langle v, d \rangle \in R_{\mathsf{B}}(w, c): \mathcal{M}, v, d \Vdash \phi.$$

This could be useful for addressing the problem of logical omniscience for agents who are perfect reasoners but endorse a nonclassical logic. For instance, while $(\rhd \phi) \rightarrow \mathsf{B}\phi$ is not valid, $(\mathsf{B}\iota \wedge @_\iota (\rhd \phi)) \rightarrow \mathsf{B}\phi$ is valid (assuming $\rhd$ is factive and noncontingent).[26] In other words, if an agent accepts a (reasonable) logic, their beliefs are closed under that logic, though beliefs need not be closed under classical consequence generally.

However, this prediction may seem unwelcome if we want to model agents who are uncertain or mistaken about what holds according to a specific logic. As it stands, interpretation terms are rigid in the hyperconvention semantics, meaning their denotation does not shift. Thus, $@_\iota (\phi_1, \ldots, \phi_n \rhd \psi)$ entails $\mathsf{B} @_\iota (\phi_1, \ldots, \phi_n \rhd \psi)$, i.e., agents are always omniscient about what follows from what according to each logic. Similarly, $@_\iota (\phi_1, \ldots, \phi_n \rhd \psi)$ entails $\alpha \mathbin{\square\!\!\rightarrow} @_\iota (\phi_1, \ldots, \phi_n \rhd \psi)$ for any α, i.e., claims about what follows from what according to a specific logic are counterfactually trivial.

To avoid this prediction, we would need to allow interpretation nominals to be nonrigid, so that belief operators (and counterfactuals) can shift their interpretation. That way, even if an agent nominally accepts a logic, they may be uncertain or mistaken about what follows according to that logic. Such a revision could also be employed to generalize the semantics so that iterated according-to operators are not redundant.

Second, given that counterfactuals can shift the interpretation of $\neg$, $\wedge$, and so on, it would be natural to wonder whether counterfactuals can shift the interpretation of *counterfactuals* as well. Even if we accept that $\phi \mathbin{\square\!\!\rightarrow} \phi$ is valid, we may want to consider logics where it is not, and even counterfactually entertain its failure ('If $\phi \mathbin{\square\!\!\rightarrow} \phi$ weren't valid,…').

Furthermore, though hyperconventions constrain the domain of propositional quantifiers, the basic meaning of the quantifiers is held fixed across hyperconventions. But different logics often provide different semantics for the propositional quantifiers, too. Thus, one may want hyperconventions to control the interpretation of propositional quantifiers.

As it stands, we cannot simply have hyperconventions interpret counterfactuals and quantifiers as with the other connectives. The problem is that the meanings of these expressions depend not just on the hyperconvention of evaluation but other hyperconventions as well. Thus, on pain of set-theoretic circularity, we could not have $c(\square\!\!\rightarrow) \colon \mathbb{P}_{D_\mathbb{H}} \times \mathbb{P}_{D_\mathbb{H}} \to \mathbb{P}_{D_\mathbb{H}}$ since c itself occurs in $D_\mathbb{H}$, and so appears in $\mathbb{P}_{D_\mathbb{H}}$.

One strategy for dealing with this is to introduce a set of "markers", which act as pointers to operations on sets of indices. Then hyperconventions can map $\square\!\!\rightarrow$ to one of these markers, thereby indirectly specifying how to interpret $\square\!\!\rightarrow$. (A similar strategy can apply to propositional quantifiers.) More precisely, a hypermodel will now be a tuple of the form $\mathcal{M} = \langle W, D_\mathbb{C}, D_\mathbb{P}, M, F, V \rangle$, where M is some nonempty set of "markers" and F maps each $m \in M$ to an binary operation on sets of indices, i.e., $F(m) \colon \wp\,\mathbb{I}_{D_\mathbb{H}} \times \wp\,\mathbb{I}_{D_\mathbb{H}} \to \wp\,\mathbb{I}_{D_\mathbb{H}}$. We revise the definition of hyperconventions so that $c(\square\!\!\rightarrow) \in M$, and revise the clause for counterfactuals:

---

[26] If an agent accepts a logic that is not factive (so a validity might be false) or is contingent (so a validity need not be necessarily valid), we should not expect their beliefs to be closed under that logic anyway.

$$\mathcal{M}, w, c \Vdash \phi \mathrel{\square\!\!\rightarrow} \psi \quad \Leftrightarrow \quad \langle w, c \rangle \in F(c(\mathrel{\square\!\!\rightarrow}))(\llbracket \phi \rrbracket^{\mathcal{M}}, \llbracket \psi \rrbracket^{\mathcal{M}}).$$

A hyperconvention in this sense is classical if $F(c(\mathrel{\square\!\!\rightarrow}))$ is a selection function in the sense of Definition 9. Thus, in shifting the hyperconvention, we may also shift the interpretation of the counterfactual.

Appendix

A representability theorem

In this appendix, I establish that any finitary single-conclusion logic can be represented in the hyperconvention semantics. In fact, we can represent any logic using just a single hyperconvention. In addition, I show that this improves upon the semantics in Kocurek and Jerzak 2021, which can only represent a smaller set of logics.

First, we need to clarify what we mean by 'logic' and 'represent'.

Definition 11 (*Logic*) A **logic** over $\mathcal{L}$ is a set $\mathbf{L} \subseteq \mathcal{L}^{<\omega} \times \mathcal{L}$ of pairs of the form $\langle \langle \phi_1, \ldots, \phi_n \rangle, \psi \rangle$. (We allow the lefthand side to be the empty tuple $\diamondsuit$.) We may write $\phi_1, \ldots, \phi_n \Rightarrow_{\mathbf{L}} \psi$ in place of $\langle \langle \phi_1, \ldots, \phi_n \rangle, \psi \rangle \in \mathbf{L}$. We use $\vec{\phi}$ as shorthand for $\langle \phi_1, \ldots, \phi_n \rangle$, where the length is implicit (possibly zero).

Definition 12 (*Representability*) A logic $\mathbf{L}$ is **representable** in $\mathcal{M} = \langle W, D_{\mathbb{C}}, D_{\mathbb{P}}, V \rangle$ if there's a $c \in \mathbb{H}_W$ such that for all $\phi_1, \ldots, \phi_n, \psi \in \mathcal{L}$:

$$\phi_1, \ldots, \phi_n \Rightarrow_{\mathbf{L}} \psi \quad \Leftrightarrow \quad \forall w \in W \colon \mathcal{M}, w, c \Vdash (\phi_1, \ldots, \phi_n \rhd \psi).$$

Theorem 13 (Representation) *Every logic is representable in an infinite hypermodel (i.e., $|W| = \aleph_0$).*

Proof Let $f \colon \mathcal{L} \to W$ be an injection. Let $[\phi] = W - \{f(\phi)\}$. Define c as follows.

$$\pi_c = \{W - \{f(\phi)\} \mid \phi \in \mathcal{L}\} \cup \{W\}$$
$$c(p) = [p]$$
$$\bullet_c[\phi] = [\bullet\phi]$$
$$\bullet_c W = W$$
$$[\phi] \circ_c [\psi] = [\phi \circ \psi]$$
$$W \circ_c [\phi] = [\phi] \circ_c W = W \circ_c W = W$$
$$[\phi_1], \ldots, [\phi_n] \rhd_c [\psi] = \begin{cases} W & \text{if } \phi_1, \ldots, \phi_n \Rightarrow_{\mathbf{L}} \psi \\ [p_1] & \text{otherwise.} \end{cases}$$
$$\vec{X} \rhd_c W = W$$
$$\vec{X}, W, \vec{Y} \rhd_c Z = \vec{X}, \vec{Y} \rhd_c Z.$$

Let $\mathcal{M}$ be a hypermodel over W with c in its hyperconvention domain such that $V(p)(c) = [p]$ (e.g., we can let $V(p) = P_p$). By a simple induction, $[\![\phi]\!]^{\mathcal{M},c} = [\phi]$ for all $\phi \in \mathcal{L}$.

Thus, if $\phi_1, \ldots, \phi_n \Rightarrow_\mathbf{L} \psi$, then $([\phi_1], \ldots, [\phi_n] \triangleright_c [\psi]) = W$, so $\mathcal{M}, w, c \Vdash (\phi_1, \ldots, \phi_n \triangleright \psi)$. Conversely, if $\phi_1, \ldots, \phi_n \not\Rightarrow_\mathbf{L} \psi$, then $([\phi_1], \ldots, [\phi_n] \triangleright_c [\psi]) = [p_1]$, so $\mathcal{M}, f(p_1), c \not\Vdash (\phi_1, \ldots, \phi_n \triangleright \psi)$.

Theorem 13 contrasts with the situation for the semantics developed in Kocurek and Jerzak 2021. There, they do not introduce an entailment operator $\triangleright$ into the language, so there is no obvious way to represent logics in the usual sense. In a footnote, they suggest the following (rephrased in our notation): c represents $\mathbf{L}$ in $\mathcal{M}$ iff $\phi_1, \ldots, \phi_n \Rightarrow_\mathbf{L} \psi$ iff for all w, if $\mathcal{M}, w, c \Vdash \{\phi_1, \ldots, \phi_n\}$, then $\mathcal{M}, w, c \Vdash \psi$. I will now show that on this understanding of representability, a strictly smaller class of logics are representable by a hyperconvention.

Definition 14 (Properties of logics) A logic $\mathbf{L}$ is **intensional** if it satisfies the following properties:

- **Reflexive**: $\phi \Rightarrow_\mathbf{L} \phi$
- **Transitive**: if $\vec{\psi} \Rightarrow_\mathbf{L} \chi$ and $\vec{\phi} \Rightarrow_\mathbf{L} \psi_i$ for each ψ_i, then $\vec{\phi} \Rightarrow_\mathbf{L} \chi$
- **Commutative**: if $\vec{\phi}, \vec{\psi} \Rightarrow_\mathbf{L} \chi$, then $\vec{\psi}, \vec{\phi} \Rightarrow_\mathbf{L} \chi$
- **Monotonic**: if $\vec{\phi} \Rightarrow_\mathbf{L} \psi$, then $\vec{\phi}, \chi \Rightarrow_\mathbf{L} \psi$
- **Congruential**: it obeys replacement of logical equivalents:

    – for $\bullet \in \{\neg, \Box, \Diamond\}$: if $\phi \Leftrightarrow_\mathbf{L} \phi'$, then $\bullet\phi \Leftrightarrow_\mathbf{L} \bullet\phi'$
    – for $\circ \in \{\wedge, \vee, \rightarrow\}$: if $\phi \Leftrightarrow_\mathbf{L} \phi'$ and $\psi \Leftrightarrow_\mathbf{L} \psi'$, then $\phi \circ \psi \Leftrightarrow_\mathbf{L} \phi' \circ \psi'$.

Observe that reflexivity, commutativity, and monotonicity entail (in fact, given transitivity, are equivalent to) the following property:

- **Reiterative**: $\vec{\phi} \Rightarrow_\mathbf{L} \phi_i$ for each ϕ_i.

Definition 15 (*Intensional representability*) A logic $\mathbf{L}$ is **intensionally representable** in $\mathcal{M} = \langle W, D_\mathbb{C}, D_\mathbb{P}, V \rangle$ if there's a $c \in \mathbb{H}_W$ such that for all $\phi_1, \ldots, \phi_n, \psi \in \mathcal{L}$:

$$\phi_1, \ldots, \phi_n \Rightarrow_\mathbf{L} \psi \ \Leftrightarrow \ [\![\phi_1]\!]^{\mathcal{M},c} \cap \cdots \cap [\![\phi_n]\!]^{\mathcal{M},c} \subseteq [\![\psi]\!]^{\mathcal{M},c}.$$

Theorem 16 (Intensional Representation). *A logic $\mathbf{L}$ is intensionally representable in an infinite hypermodel $\mathcal{M}$ iff $\mathbf{L}$ is intensional. Regardless of $\mathcal{M}$'s size, $\mathbf{L}$ is intensionally representable in $\mathcal{M}$ iff $\mathbf{L}$ is intensional and the number of nonequivalent finite sequences of formulas is no greater than $|W|$.*

Proof The left-to-right direction is straightforward: one simply checks that necessary implication has all the features of an intensional logic.

For the right-to-left direction, we begin by defining the **finite theory space** of $\mathbf{L}$, which is a generalization of the Lindenbaum-Tarski algebra. Where $\langle \phi_1, \ldots, \phi_n \rangle \in \mathcal{L}^{<\omega}$, define:

$$[\langle \phi_1, \ldots, \phi_n \rangle]_\mathbf{L} = \{\psi \in \mathcal{L} \mid \phi_1, \ldots, \phi_n \Rightarrow_\mathbf{L} \psi\}.$$

The finite theory space of $\mathbf{L}$ is the order $\mathbb{F}_\mathbf{L} = \langle F_\mathbf{L}, \leq_\mathbf{L} \rangle$ where:

- $F_{\mathbf{L}} = \left\{ [\langle \phi_1, \ldots, \phi_n \rangle]_{\mathbf{L}} \mid \langle \phi_1, \ldots, \phi_n \rangle \in \mathcal{L}^{<\omega} \right\}$
- $[\vec{\phi}]_{\mathbf{L}} \leq_{\mathbf{L}} [\vec{\psi}]_{\mathbf{L}}$ iff $\vec{\phi} \Rightarrow_{\mathbf{L}} \psi_i$ for each $\psi_i \in \vec{\psi}$.

From now on, we'll leave the **L**-subscripts implicit. Recall that a meet-semilattice is a partial order where the meet (i.e., greatest lower bound) of any two elements exists.

Claim $\mathbb{F}$ is a meet-semilattice with a top element.

Proof Since **L** is reiterative, $\phi_1, \ldots, \phi_n \Rightarrow_{\mathbf{L}} \phi_i$ for $1 \leq i \leq n$. So $\leq$ is reflexive. The transitivity and antisymmetry of $\leq$ follow from the transitivity of **L**. The top element is simply $[\langle\rangle]$. As for the existence of meets, define the meet $\sqcap$ operation as follows:

$$[\vec{\phi}] \sqcap [\vec{\psi}] = [\vec{\phi}, \vec{\psi}].$$

That $[\vec{\phi}, \vec{\psi}] \leq [\vec{\phi}]$ and $[\vec{\phi}, \vec{\psi}] \leq [\vec{\psi}]$ follows from reiterativity. Moreover, if $[\vec{\chi}] \leq [\vec{\phi}]$ and $[\vec{\chi}] \leq [\vec{\psi}]$, then $[\vec{\chi}] \leq [\vec{\phi}, \vec{\psi}]$ by definition of $\leq$. So $\sqcap$ really is a meet operation.

It is a well-known fact of order-theory that any partial order can be order-embedded into its powerset algebra in a meet-preserving manner. Since $F_{\mathbf{L}}$ is at most the size of W, that means there's an injective map $f : F \to \wp W$ such that:

(i) $[\vec{\phi}] \leq [\vec{\psi}]$ iff $f([\vec{\phi}]) \subseteq f([\vec{\psi}])$
(ii) $f([\vec{\phi}, \vec{\psi}]) = f([\vec{\phi}]) \cap f([\vec{\psi}])$.

What's more, we can take $f([\langle\rangle]) = W$. For if $f([\langle\rangle]) \neq W$, define f^* to be exactly like f except $f^*([\langle\rangle]) = W$. It is easy to then verify that f^* is still an order-embedding that preserves finite meets.

Define c as follows (where $\bullet \in \{\neg, \Box, \Diamond\}$ and $\circ \in \{\wedge, \vee, \to\}$):

$$\pi_c = \{f([\phi]) \mid \phi \in \mathcal{L}\}$$
$$c(p) = f([p])$$
$$\bullet_c f([\phi]) = f([\bullet\phi])$$
$$(f([\phi]) \circ_c f([\psi])) = f([\phi \circ \psi]).$$

This is well-defined. For if $f([\phi]) = f([\psi])$, then $[\phi] = [\psi]$ by injectivity. Thus, $\phi \Leftrightarrow_{\mathbf{L}} \psi$. So by congruentiality, $\bullet\phi \Leftrightarrow_{\mathbf{L}} \bullet\psi$, which means $[\bullet\phi] = [\bullet\psi]$. Hence, $f([\bullet\phi]) = f([\bullet\psi])$. (And similarly for binary connectives.)

Throughout, let $\mathcal{M}$ be a hypermodel of the relevant size with c in its hyperconvention domain where $V(p)(c) = c(p)$. By induction:

Claim For any $\phi \in \mathcal{L}$, $[\![\phi]\!]^{\mathcal{M},c} = f([\phi])$.

Finally, we show that c intensionally represents **L** in $\mathcal{M}$.

$\Leftarrow$ Suppose $\phi_1, \ldots, \phi_n \not\Rightarrow_{\mathbf{L}} \psi$. Thus, $[\phi_1, \ldots, \phi_n] \not\leq [\psi]$. Since f is order-reflecting, $f([\phi_1, \ldots, \phi_n]) \not\subseteq f([\psi])$. So there is a $w \in f([\phi_1, \ldots, \phi_n]) - f([\psi])$. Now observe:

$$f([\phi_1, \ldots, \phi_n]) = f([\phi_1]) \cap \cdots \cap f([\phi_n])$$
$$= [\![\phi_1]\!]^{\mathcal{M},c} \cap \cdots \cap [\![\phi_n]\!]^{\mathcal{M},c}.$$

Hence, $\mathcal{M}, w, c \Vdash \phi_1, \ldots, \phi_n$ but $\mathcal{M}, w, c \nVdash \psi$.

$\Rightarrow$ Suppose $\phi_1, \ldots, \phi_n \Rightarrow_{\mathbf{L}} \psi$. Then $[\phi_1, \ldots, \phi_n] \leq [\psi]$. Since f is order-embedding:

$$f([\phi_1, \ldots, \phi_n]) = f([\phi_1]) \cap \cdots \cap f([\phi_n]) \subseteq f([\psi]), \text{ i.e.,}$$
$$[\![\phi_1]\!]^{\mathcal{M},c} \cap \cdots \cap [\![\phi_n]\!]^{\mathcal{M},c} \subseteq [\![\psi]\!]^{\mathcal{M},c}.$$

Therefore, c intensionally represents $\mathbf{L}$ in $\mathcal{M}$. □

References

Anderson, Alan R., & Belnap, Nuel D. (1975). *Entailment: The logic of relevance and neccessity*. Princeton University Press.

Areces, C., & ten Cate, B. (2006). Hybrid logics. In P. Blackburn, F. Wolter, & J. van Benthem (Eds.), *Handbook of modal logic* (pp. 821–868). Elsevier.

Bacon, A. (2020). Logical combinatorialism. *The Philosophical Review, 129*, 537–589.

Beall, J. C., & Restall, G. (2006). *Logical pluralism*. Oxford: Oxford University Press.

Berto, F., French, R., Priest, G., & Ripley, D. (2018). Williamson on counterpossibles. *Journal of Philosophical Logic, 47*, 693–713.

Berto, F., & Jago, M. (2019). *Impossible worlds*. Oxford University Press.

Berto, F., & Nolan, D. (2021). Hyperintensionality. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy*. Metaphysics Research Lab, Stanford University.

Braüner, T. (2006). Axioms for classical, intuitionistic, and paraconsistent hybrid logic. *Journal of Logic, Language, and Information, 15*, 179–194.

Braüner, T. (2011). Intuitionistic hybrid logic: Introduction and survey. *Information and Computation, 209*, 1437–1446.

Braüner, T. (2017). Hybrid Logic. In E. N. Zalta (ed.), *The stanford encyclopedia of philosophy*. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2017/entries/logic-hybrid/

Braüner, T., & de Paiva, V. (2006). Intuitionistic hybrid logic. *Journal of Applied Logic, 4*, 231–255.

Brogaard, B., & Salerno, J. (2013). Remarks on counterpossibles. *Synthese, 190*, 639–660.

Bull, R. A. (1969). On modal logic with propositional quantifiers. *The Journal of Symbolic Logic, 34*, 257–263.

Chadha, R., Macedonio, D., & Sassone, V. (2006). A hybrid intuitionistic logic: Semantics and decidability. *Journal of Logic and Computation, 16*, 27–59.

Di C. R. & Miller, D. (2019). Linear Logic. In E. N. Zalta (ed.), *The stanford encyclopedia of philosophy*. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/entries/logic-linear/

Dunn, J. M., & Restall, G. (2002). Relevance logics. In D. M. Gabbay & F. Guenthner (Eds.), *Handbook of philosophical logic* (Vol. 6, pp. 1–128). Springer.

Einheuser, I. (2006). Counterconventional conditionals. *Philosophical Studies, 127*, 459–482.

Fine, K. (1969). *For some proposition and so many possible worlds*. University of Warwick. Ph.D. thesis.

Fine, K. (1970). Propositional quantifiers in modal logic. *Theoria, 36*, 336–346.

Fine, K. (2012). Counterfactuals without possible worlds. *Journal of Philosophy, 109*, 221–246.

French, R., Girard, P., & Ripley, D. (2020). Classical counterpossibles. *The Review of Symbolic Logic* 1–20.

Fritz, P. (2020). Propositional quantification in bimodal S5. *Erkenntnis, 2*, 455–465.

Gibbard, A. (2003). *Thinking how to live*. Harvard University Press.

Girard, P., & Weber, Z. (2015). Bad worlds. *Thought, 4*, 93–101.

Holliday, W. H. (2017). A note on algebraic semantics for S5 with propositional quantifiers. *Notre Dame Journal of Formal Logic, 60*, 311–332.

Holliday, W. H., & Litak, T. (2018). One modal logic to rule them all? In G. Bezhanishvili, G. D'Agostino, G. Metcalfe, & T. Studer (Eds.), *Advances in modal logic* (Vol. 12, pp. 367–386). College Publications.

Jago, M. (2014). *The impossible*. Oxford University Press.

Kaminski, M., & Tiomkin, M. (1996). The expressive power of second-order propositional modal logic. *Notre Dame Journal of Formal Logic, 37*, 35–43.

Kaplan, D. (1970). S5 with quantifiable propositional variables. *Journal of Symbolic Logic, 35*, 355.

Kment, B. (2014). *Modality and explanatory reasoning*. Oxford University Press.

Kocurek, A. W. (2019). Hyperlogic: A system for talking about logics. In J. J. Schlöder, D. McHugh, and F. Roelofsen (Eds.), *Proceedings for the 22nd Amsterdam colloquium* (pp. 238–247).

Kocurek, A. W., & Jerzak, E. J. (2021). Counterlogicals as counterconventionals. *Journal of Philosophical Logic, 50*, 673–704.

Kocurek, A. W., Jerzak, E. J., & Rudolph, R. E. (2020). Against conventional wisdom. *Philosophers' Imprint 20*, 1–27.

Krakauer, B. (2012). *Counterpossibles*. University of Massachusetts, Amherst Ph.D. thesis.

Leitgeb, H. (2019). HYPE: A system of hyperintensional logic (with an application to semantic paradoxes). *Journal of Philosophical Logic, 48*, 305–405.

Lewis, D. K. (1973). *Counterfactuals*. Harvard University Press.

Mares, E. D. (1997). Who's afraid of impossible worlds? *Notre Dame Journal of Formal Logic, 38*, 516–526.

Mares, E. D. (2004). *Relevant logic: A philosophical interpretation*. Cambridge University Press.

Mares, E. D. (2020). Relevance logic. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy*. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/entries/logic-relevance/

Meyer, Robert K., & Routley, Richard. (1977). Extensional Reduction–I. *The Monist, 60*, 355–369.

Muñoz, P. (2020). *Behavioral attitudes in hyperintensional grammar* (pp. 1–41). (un published)

Muskens, Reinhard A. (1991). Hyperfine-grained meanings in classical logic. *Logique et analyse, 34*, 159–176.

Nolan, D. (1997). Impossible worlds: A modest approach. *Notre Dame Journal of Formal Logic, 38*, 535–572.

Nolan, D. (2018). Reflections on Routley's ultralogic program. *Australasian Journal of Logic, 15*, 407–430.

Routley, R. (2019). *Ultralogic as universal?* Springer.

Sandgren, A., & Tanaka, K. (2020). Two kinds of logical impossibility. *Noûs, 54*, 795–806.

Schaffer, J. (2016). Grounding in the image of causation. *Philosophical Studies, 173*, 49–100.

Stalnaker, R. C. (1968). A theory of conditionals. In N. Rescher (Ed.), *Studies in logical theory (American philosophical quarterly monographs 2)* (pp. 98–112). Basil Blackwell Publishers.

Standefer, S. (2020). Actual issues for relevant logics. *Ergo, 7*, 241–276.

Tanaka, K. (2018). Logically impossible worlds. *Australasian Journal of Logic, 15*, 489–497.

Vander Laan, D. A. (2004). In F. Jackson & G. Priest (Eds.), Counterpossibles and similarity. *Lewisian Themes*.

Weber, Z. (2014). Naïve validity. *The Philosophical Quarterly, 64*, 99–114.

Weber, Z., Badia, G., & Girard, P. (2016). What is an inconsistent truth table? *Australasian Journal of Philosophy, 94*, 533–548.

Williamson, T. (2009). Probability and danger. *The Amherst Lecture in Philosophy, 4*, 1–35.

Williamson, T. (2013). *Modal logic as metaphysics*. Oxford University Press.

Wilson, A. (2018). Grounding entails counterpossible non-triviality. *Philosophy and Phenomenological Research, 46*, 716–728.

Yanovich, I. (2015). Expressive power of "now" and "then" operators. *Journal of Logic, Language, and Information, 24*, 65–93.