The Logic of Action and Control*

March 31, 2023

Abstract

In this paper I propose and motivate a logic of the interdefined concepts of *making true* and *control*, understood as intensional propositional operators to be indexed to an agent. While bearing a resemblance to earlier logics in the tradition, the motivations, semantics, and object language theory differ on crucial points. Applying this logic to widespread formal theories of agency, I use it as a framework to argue against the ubiquitous assumption that the strongest actions or options available to a given agent must always be pairwise incompatible. The conclusion is that this assumption conflicts with failures of *higher order control* of agents over their degree or precision of control, failures exhibited by such imperfect agents as ourselves. I discuss models in this setting for understanding such imperfectly self-controlling agents. In an appendix, I prove several relevant results about the logic described, including soundness and completeness both for it and for certain natural extensions.

Keywords

action; decision theory; logic of action; modal logic

Contents

*I would like to thank Andrew Bacon for extensive discussion and feedback on earlier drafts of this paper, and an anonymous referee at this journal for pressing me on certain crucial points. I would also like to thank Harvey Lederman, Taylor Friesen, and Douglas Vaaler for helpful discussions on these topics.

Introduction

This paper proposes and motivates a logic of the concepts of *control* and *making so*; that is to say, a general theory of their interaction with the familiar operators of classical logic. The senses in which I intend these expressions is familiar from everyday use: it is the sense in which, by placing a pot of water over a lit stove, one (seemingly, under ordinary circumstances) makes it so that the water boils, and controls *whether* it boils. This is not, of course, a new topic in philosophical logic, but the avenues pursued in our discussion will differ in measured but important respects from previous approaches.

The first section clarifies the target concept and scope of the paper, as well as motivating an interest in it. The substantive section of the paper begins with an investigation of the basic principles of our operator; while it shares a number of motivations with extant treatments, the final theory is neither strictly stronger nor strictly weaker than its prevailing rivals, and the discussion in the previous section illuminates otherwise obscure aspects of old debates. The resulting logic, as proved in the appendix, has an appealing semantics for which it is sound and complete, even though the motivations given in the main body of the paper are entirely non-semantic in nature, and there are natural extensions of the logic we may consider corresponding to natural conditions on the frames of the semantics. Finally, I take up the logic's relation to the concept of an *action proposition* in formal models of agency, using it as a framework in which to argue against the partitionality of such propositions.

While I present the theory formally in the appendix, the bulk of the paper is informal and glosses over many technical worries. The technical machinery is intended to bolster and formulate arguments that are in spirit conceptual and intuitive rather than mathematical.

1 Motivations

1.1 Conceptual Preliminaries

Our point of departure is a pair of equivalences:

1. Necessarily, one controls whether p iff one either makes it so that p or makes it so that not p

2. Necessarily, one makes it so that p iff p and one controls whether p

Together, these render control and making true interdefinable, and we will treat formulations with one as equivalent to the corresponding formulations with the other. As philosophical theses, the equivalences have great *prima facie* plausibility, and I am inclined to endorse both. A full defence of (1) and (2) as substantive claims, however, would take us too far afield; this leaves us with the question of which between control and making true to take as the actual primitive under discussion, and with which to formalise the principles we will consider.

I opt for control and making true, respectively. For the latter, the non-"interrogative" nature of the construction will make our formulations less tedious in their syntax, more easily visualised in their semantics, and readily comparable in both respects to extant competitors. For the former, interrogative uses of *control* are simply much more common in everyday speech than propositional uses of *make*. This does not, of course, on its own demonstrate the concept it expresses to be of greater philosophical significance, nor is this paper an exercise in natural language semantics. But it does put us in a better position to assess the pre-theoretical plausibility of various principles we will consider, and to thus shed light on otherwise uncertain clashes of intuition.

In the ensuing discussion, therefore, we will treat the concept *X makes it so* as a property of propositions denoted semi-formally by a sentential operator $\Box$ (with sentences containing it to be ultimately understood, in accord with (1) and (2), as abbreviations of sentences about control). We will also carelessly conflate constructions like *α makes it so that X is φ*, *α makes it true that X is φ*, and *α makes X φ*; these are all to be analysed, again, in the natural way in terms of control. Thus, where p stands for "the water is boiling", $\Box p$ will stand for the sentence "*X* makes it so that the water is boiling" or, fundamentally, "The water is boiling and *X* controls whether it is".

This framing is somewhat committal. It rejects a view of agency in which *actions* as entities in their own right feature centrally for one treating it as a relation between agents and propositions, against certain prominent grains in action theory [1] [2]. This is, however, not intended to convey any fundamentality or ultimate perspicuity of the agency-as-operator perspective. It is consistent with the arguments of this paper that this general framing be provisional, or to be explained in other terms. What interests me is not the operator perspective itself, but what principles we should accept when taking it.

As a final point, nothing in this investigation hinges (or at least, is intended to hinge) on substantive assumptions about the scope of one's control, or the precise relevant sense of *making true* (and *controls*). It may be, for example, that one makes true such ”worldly” propositions as that the water in the pot boils; or it may be the only things an agent ever makes true are ”internal” facts about their will or mental state; or it may be that ”controls” and ”makes true” admit of stricter and weaker readings, on some of which an agent can be truthfully said to make true worldly propositions and on some of which she cannot. There are, I take it, general questions about the logic of these expressions independent of such ambiguities and substantive background views, and it is these more general questions I intend to investigate.

1.2 Further Significance

These coordinate concepts of control and making so are of wide-ranging philosophical significance. They are closely connected, for one, to the understanding of an *action* present in both all versions of causal decision theory and Jeffrey-style evidential decision theory [3] [4] [5]. Such versions of decision theory take the objects they recommend to a rational agent (given her preferences and prior doxastic state) to be *propositions* she is in a position to make true (as opposed to *lotteries* one is in a position to gamble on, as in [6] [7]). This is commonly taken to be an advantage of such theories: it renders these objects of recommendation more familiar, and unifies them with the objects of credence and desire. It does, however, raise the question of what is involved in making a proposition true, a question with both more substantive aspects in the theory of action proper and more abstract aspects, such as those discussed in this paper.

In less heavily technical areas, the concepts bear on debates in the study of free will between compatibilists and incompatibilists, and on debates over the scope of our moral responsibilities under the heading of ”moral luck” [8] [9] [10] [11] [12] [13]. Debates in the former case centre precisely on the relation of making true to nomological necessity, and debates in the latter case have centered around the contentious principle that one is morally responsible for a fact p only if one controls (or has control over) whether p, which if true will mean that the logic of making true straightforwardly limits the logic of moral responsibility. Getting clear on the logic of $\Box$, therefore, stands to illuminate a number of disparate areas.

2 Logical Principles

2.1 Normality

For our operator $\Box$ to be a *normal* one, the following must be true of it (taking some harmless liberties with use and mention):

Distributivity It must distribute across the material conditional, so that if one makes it so that, if p, q, then if one makes it so that p, one makes it so that q. $(\Box(p \to q) \to (\Box p \to \Box q))$

Necessitation If one can prove in classical logic, together with the aforementioned principle and this additional current rule of inference, that p, then one makes it the case that p (If $\vdash p$, $\vdash \Box p$)

Not only does common sense about what we can make so contradict these principles, they each fall afoul of it separately. To begin, consider that, in the presence of the Necessitation, Distributivity is equivalent to the following (sometimes called, respectively, M and C):

Weakening If one makes it so that both p and q, one makes it so that p and one makes it so that q

Agglomeration If one makes it so that p and one makes it so that q, one makes it so that both p and q

The first of these is open to obvious counterexamples. Suppose that I am handling a ball tethered to a pole by a robust 2m chain, which I freely place within 1m of the pole. Then I seem to have made true that (controlled whether) the ball is within 1m of the pole, but not to have controlled whether it is within 2m of the pole, contradicting Weakening. Note that this argument does not appeal to any overt, linguistic conjunction, disjunction, or quantification, thus sidestepping objections (such as those raised in [14] to similar arguments) that it trades on syntactically introduced ambiguities and presuppositions.

Necessitation is even more facially absurd. One line of argument would appeal to the common intuition that ordinary agents do not make true any facts of logic (see [15] [16]), in a possible worlds setting understood as the trivial proposition $\top$. But this thinking is vulnerable to methodological objections: it is a mainstay on formal theories of propositional attitudes that attitudes towards the trivial proposition should be seen as themselves degenerate and trivial instances of those attitudes, and our disinclination to report them explicable on merely pragmatic grounds [17] [18] [19]. The defender of Necessitation is thus free to extend similar reasoning to making true (see again [14]).

A better route appeals instead to intuitions about our control over *necessities* of varying strictness. There are some very restricted and weak species of necessity over which I exert control: I cannot (by reason of the ill-suited shoes I happen to be wearing) run a six minute mile, and yet I plausibly control whether I run a six minute mile, since the necessity involved is of a very weak kind. As the species of necessity increase in strictness, my intuitions of non-control grow increasingly stronger: I cannot (by reason of my overall build) run a four minute mile, cannot (by reason of human physiology) run a two minute mile, cannot (by reason of physical law) run a one nanosecond mile, and cannot (by conceptual necessity) run a zero second mile. In each case, I am more certain I do not control the relevant impossibility; failure to control $\top$ is a terminus in a progressively stronger series of intuitions, not an outlier treatable as an aberrant or degenerate case.

This is, obviously, a variant on the above argument against Weakening. It appeals directly to intuitions about what we do and do not control, rather than (as in [20]) a version of the principle of alternative possibilities, or any other contentious general theses about control [21] [22] [23]. Nor does it tie failure of control to any specific privileged species of necessity. These belong to a broader range of entanglements with modality our theory will seek to avoid.

2.2 A Basic Logic

While our target concept plainly cannot be analysed as a normal operator, it is illuminating to observe respects in which it does *not* violate normality. Unlike with many representational attitudes, for one, whenever p is logically equivalent to q it seems that one makes it so that p iff one makes it so that q: what one has made happen seems, as it were, insensitive to any differences of logically extraneous guise. Indeed, this seems true not just of logical equivalences, but equivalence under various narrower species of necessity. It seems to be a law of nature, for example, that a surface appears green (to ordinary humans in ordinary conditions) iff it reflects light of wavelengths 495-570nm, even though we can easily imagine the laws of nature requiring otherwise. And, for this reason, it seems that if I make it so that a surface reflects light of of wavelengths 495-570 nm (by painting it, say), I have thereby made it to appear green, and *vice versa*.

This suggests that the logic of action is, if not normal, at least *congruent*, i.e. it is closed under the following rule of inference:

RE If p and q are provably equivalent in our logic, that one makes it so that p and that one makes it so that q are provably equivalent (If $\vdash p \leftrightarrow q$, $\vdash \Box p \leftrightarrow \Box q$)

Of course, **RE** requires not only that propositions equivalent modulo classical logic are made true when the other is, but that this holds of of propositions equivalent modulo the *whole logic being proposed* (including the inference rule **RE**). Nevertheless, I think the theory's other principles are themselves suitably necessary to permit equivalence of what is made true holding them fixed. This assumption is ubiquitous among competitors for the reasons we have given.

As an obvious first entry in our list of axioms (indeed, one that follows from our definition (2) of *makes that* in terms of control), making true is factive; what one makes true *is* true:

Factivity If one makes it so that p, p ($\Box p \rightarrow p$)

Next, while Weakening on inspection turned out to be implausible, its sister principle Agglomeration seems perfectly good: if I make my bread into a circle and make it pink, I thereby make it into a pink circle. We will therefore add the axiom to our logic.

On the subject of Weakening, with **RE** in place there is an additional difficulty posed by the principle unmentioned in the previous section. Together with RE, it implies that, when one makes it so that p, for any arbitrary q one makes it so that either p or q; that is, when one has made any given disjunct of a given disjunction true, one has made true the disjunction itself. Experience and imagination are replete with counterexamples: for example, when making it so that a chained ball is on the left side of the room (by pushing it, say), I need not have made it so that the chained ball is somewhere or other in the room; the chain strips me of any input on that matter. In other cases, the additional disjunct may even be true, but still unrelated to my doings. Even though the sun is shining as I turn on an overhead light, for example, it does not seem true that I make it so that at least one of the sun or the overhead light shines.

In the first sort of counterexample, the agent determines one or another of some (not necessarily mutually exclusive) outcomes, where the range of options itself is fixed independent of the agent. In the second, among a range of such options, one of them is beyond the control of the agent, even if it still obtains. Such counterexamples cannot arise, however, when *both* disjuncts have been made true by an agent; in such a scenario, her control over the disjunction seems complete. I therefore propose the following axiom for our logic:

Modest Weakening If one makes it so that p and one makes it so that q, one makes it so that at least one of p or q is true $((\Box p \wedge \Box q) \to \Box(p \vee q))$

There is another broad class of cases where weakening seems licit. Suppose that I of my own volition choose to go walk, and moreover specifically to walk quickly to the store. Then it seems, not only that I must have controlled whether I walked or walked quickly to the store, but that I must have controlled whether I walked quickly *simpliciter*. While there are circumstances where it might seem reasonable to say I controlled whether I walked quickly to the store but deny that I controlled whether I walked quickly, the additional supposition that I controlled whether I walked at all seems to undermine this cotenability. This case is indicative of a broader trend: when I make true both p and some weaker q, I must also make true any r of intermediate strength. We can formulate this principle in our logic straightforwardly.

Convexity If one makes it so that one of p, q, or r is true and one makes it so that p is true, one makes it so that one of p or q is true $(\Box(p \vee q \vee r) \to \Box p \to \Box(p \vee q))$

The theory thus axiomatised, together with modus ponens and the theorems of classical propositional logic, turns out to have an elegant semantics (which we will call *sandwich semantics*), for which soundness and completeness can be proved. The semantics is a possible worlds one, with sentences interpreted by subsets of a domain of worlds, equipped with a (possibly partial) function $(S_{\mathsf{min}}, S_{\mathsf{max}})$ (for short, S) from the domain to pairs of its subsets, where $w \in S_{\mathsf{min}}(w) \subseteq S_{\mathsf{max}}(w)$. The domain and function together constitute a *sandwich frame*. The classical connectives, as usual, are interpreted by their set theoretic analogues, while our operator $\Box$ is interpreted so that, at any world w, the

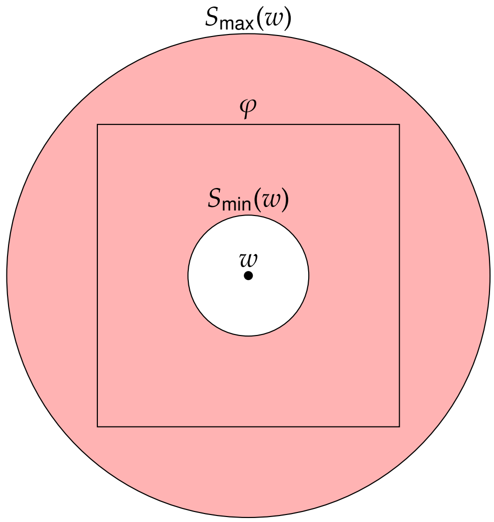

Figure 1: An illustration of the inner and outer circles at w with $w \in [\![\Box\varphi]\!]$ and the area "between" the two circles shaded

propositions (if any) of which $\Box$ is true at w are those sandwiched between (inclusively) $S_{\mathsf{min}}(w)$ and $S_{\mathsf{max}}(w)$, i.e. the X such that $S_{\mathsf{min}}(w) \subseteq X \subseteq S_{\mathsf{max}}(w)$.

One may picture a model of the logic as a surface, with each pointed associated with two concentric circles containing it, together with a function associating sentences of our language to regions of the surface. The circles around w (joined with their respective interiors) represent the strongest and weakest things the agent makes true at w, and the things one makes true are those subregions of the larger circle that entirely contain the smaller circle (see Figure 1).

Note that none of the principles here proffered *prima facie* require that one ever make anything true. This is confirmed by the semantics, which transparently allow for models with widespread or even universal failures of definedness for $(S_{\mathsf{min}}, S_{\mathsf{max}})$.

We call the logic generated by the axiom schemas and inference rules listed Conv, for "convexity logic"; the logic resulting from expanding that list of axiom schemas by some others we denote by appending the names of those additional schemas to the name Conv. The class of convexity logics is, clearly, a generalisation of the class of factive normal logics, and the class of sandwich frames a generalisation of the class of reflexive Kripke frames: any normal logic is equivalent to a convexity logic with the axiom $\Box\top$, and every reflexive Kripke frame is equivalent to a sandwich frame $\langle W, (S_{\mathsf{min}}, S_{\mathsf{max}})\rangle$ whose S_{max} is the constant function to W. Convexity logics are, as a slogan, just like normal logics with locally restricted domains of worlds.

We may thus think of Conv as standing to sandwich semantics frames as the logic KT stands to reflexive Kripke frames. It serves as a minimal framework for theorising about a factive convex operator, just as KT with reflexive Kripke frames serves as a minimal framework for theorising about a factive normal operator.[1]

2.3 Further Axioms

There are two important extensions to our logic which, while we will not adopt them, will become salient later.

Positive Introspection If one makes it so that p, one makes it so that one makes it so that p ($\Box p \to \Box\Box p$)

Relative Negative Introspection If one makes it so that p but does not make it so that q, one makes it so that both p and one does not make it so that q ($(\Box p \land \neg\Box q) \to \Box(p \land \neg\Box q)$)

The first is, of course, the familiar axiom 4, which in normal logics corresponds to transitive Kripke frames. In the sandwich semantics, it corresponds instead to the condition of S not *narrowing*; that is, for any frame $\langle W, (S_{\mathsf{min}}, S_{\mathsf{max}})\rangle$ on which 4 is valid, and for any $w \in W$, for no $v \in S_{\mathsf{min}}(w)$ does $S_{\mathsf{max}}(v)$ lack some $u \in S_{\mathsf{max}}(w)$ nor does $S_{\mathsf{min}}(v)$ contain some $u \notin S_{\mathsf{min}}(w)$ (see Figure 2). In line with the picture given above, this means that, when moving to a point within the smaller of the two concentric circles associated with w, the new circles only differ (if at all) by strictly *increasing* the region between the two circles; either the inner circle contracts, the outer circle expands, or both (or neither). Conv4 is sound and complete with respect to this class of frames.

Just as 4 prohibits "narrowing," so the latter (which we will call 5*) prohibits "widening:" when moving within some $S_{\mathsf{min}}(w)$ to another world (with defined circles), the outer circle gains no worlds, and the inner circle loses none. Unlike its analogue 5, however, it does not in the presence of Factivity entail 4; the convex logic Conv45* with both axioms, instead, corresponds to those sandwich frames where no movement within the inner circle of w changes either inner or outer circle (that is, prohibits both narrowing and widening).[2]

2.4 Conv and STIT

The question of the true logic of making so–sometimes called the logic of *action* or of *STIT* (seeing to it that)–is not virgin philosophical territory. Early entries in the investigation of this logic include [15] [25], and the debate was raised to

[1]Our models are, of course, a special case of neighbourhood or Scott-Montague models, and thus stand in an intermediate position between Kripke/relational and neighbourhood semantics for modal operators [24].

[2]Conv5* is equivalent to Conv45*, however, as long as we consider only sandwich models where, for all w with $S(w)$ defined, $S(w')$ is defined for all $w' \in S_{\mathsf{min}}(w)$ (Corollary A.14). That is, informally, 5* fails to entail 4 only under the perhaps strange circumstance that it might be consistent with all one makes true that one makes nothing true. The plausibility of such a possibility, and how best to approach it under the present framework, is an intriguing subject for future investigation.

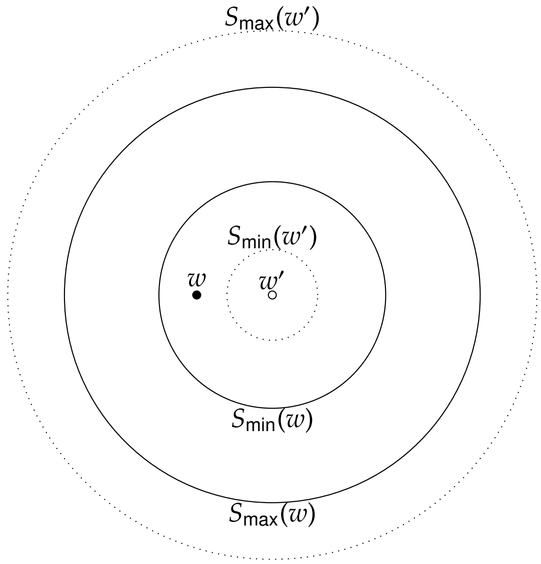

Figure 2: An illustration of the "expansion" permitted by 4 of the "sandwich" as moving from w to $w' \in S_{\mathsf{min}}(w)$

its current level of formal semantic sophistication by [26] (for a comprehensive historical overview see [27]). Subsequent formal research has been dominated by the tradition of [16], and it is therefore worth noting where the current discussion agrees with and diverges from the main motivations and commitments of this tradition. We will describe two classes of STIT model: the traditional *branching* models and more minimalist *atemporal* models. While the latter are simpler and more comparable to our own sandwich models, a review of the branching semantics will be helpful in conveying the philosophical motivations of the STIT project.

A *branching STIT frame* (with one agent) is a structure $\langle T, \leq, \mathrm{Ch}\rangle$, where T is some non-empty set, $\leq$ a partial order on T which branches only "upwards" and which has some common lower bound for any two elements, and Ch is a function from $m \in T$ to partitions on maximal chains (or "histories") running through m. Sentences are interpreted on such a frame not, as one might expect, as subsets of T, but as sets of pairs $\langle m, h\rangle$ where $m \in T$ and m is a history with $m \in h$ (an *historical pair*). A *branching STIT model* is, accordingly, a STIT frame together with a base interpretation function associating each atomic sentence with a set of historical pairs. Sometimes further structure is added to the underlying frame, such as a set of multiple agents or a contemporaneity relation on T, or further restrictions are imposed on the relation between the parameters, but for our purposes this minimal structure will suffice.[3]

The vocabulary of the language interpreted on such a frame minimally

[3]For extensive technical development of this framework see [20] [28] [29] [30] [31].

contains denumerably many atomic sentences, the classical connectives, and the unary modal operator [*dstit*] (the analogue to our $\Box$). The first two are interpreted as usual; an historical pair $\langle m, h\rangle$ will be in $[\![[dstit]\varphi]\!]$ just in case a) for all $h' \in \mathrm{Ch}(m)[h]$, $\langle m, h'\rangle \in [\![\varphi]\!]$, and b) for some $h'' \ni m$, $\langle m, h''\rangle \notin [\![\varphi]\!]$. This is not the only STIT operator studied, but it is both the best formally understood and the closest to ours in both formal and informal interpretation, and so we will restrict our attention to it.

Other operators sometimes included in the language are the tense operators P, F (for the past and future tenses, respectively) and the "historical necessity" operator $\blacksquare$. We have $\langle m, h\rangle \in [\![\mathsf{P}(\mathsf{F})\varphi]\!]$ iff $\langle m', h\rangle \in [\![\varphi]\!]$ for some $m' < (>)\, m$, and $\langle m, h\rangle \in [\![\blacksquare\varphi]\!]$ iff all $\langle m, h'\rangle \in [\![\varphi]\!]$.

From these technical remarks, much about the underlying philosophical picture should already be apparent. For STIT enthusiasts, time is a garden of forking paths: each moment has only one linear past, but an array of various possible future trajectories, of which only one will be taken. What is settled, or historically necessary, at a given point is what is bound to occur along any path thence is taken. At any given moment, these future trajectories are partitioned into the choices open to our agent, of which she chooses one. What she makes true are her choice and all of its consequences not already settled for her by her position in, and the overall structure of, time's branching garden path.[4]

While the branching time structure in the models is both traditional and well illustrative of the theory's guiding intuitions, atemporal axiomatisations and semantic structures are both available and formally well-understood [28] [32] [33]. In these Kripke-style models, historical pairs are replaced with structureless worlds as points of evaluation, which together form the domain W of a model. An *atemporal STIT frame* $\langle W, \mathrm{Ch}, H\rangle$ consists of a non-empty domain W and two equivalence relations $\mathrm{Ch} \subseteq H$, representing respectively one's choice at a world and historical necessity, while an *atemporal STIT model* is such a frame plus a base valuation from atomic sentences to subsets of W. Atomic sentences and classical connectives are interpreted as usual; w is in $[\![\blacksquare\varphi]\!]$ just when $[\![\varphi]\!] \supseteq H(w)$, w is in $[\![[dstit]\varphi]\!]$ just when $[\![\varphi]\!] \supseteq \mathrm{Ch}(w), \not\supseteq H(w)$. This validates the same logic for these two operators as the old semantics; it is just a branching STIT frame shorn of its branching.

There is a close resemblance between $\mathsf{Conv45}^*$ and the logic of [*dstit*]. The axiomatic theory of the latter includes all our own axioms besides Modest Weakening, which plays the least significant role for us. The semantical picture is similar, too: in both cases, the extension of the operator at an index i is uniquely determined by two sets $S_{\mathsf{min}}(i), S_{\mathsf{max}}(i)$ containing it, one a subset of the other, the difference being that in the sandwich semantics its extension at i is all those sets between $S_{\mathsf{min}}(i)$ and $S_{\mathsf{max}}(i)$, in STIT semantics all those greater than $S_{\mathsf{min}}(i)$ but not greater than $S_{\mathsf{max}}(i)$. And while the STIT semantics bakes in the assumption of Relative Negative Introspection (and Positive Introspection), these can be eliminated fairly easily by replacing the co-domain of $\mathrm{Ch}(m)$ with *reflexive*

[4] Advocates of STIT thus typically accept a certain form of indeterminism, in that they generally deny that all facts about the future are historically settled.

relations on the histories through *m* (as is effectively done in [14]).[5]

The entanglements between [*dstit*] and historical necessity pose serious problems. The effect of the second, "negative" clause for [*dstit*] is to invalidate WEAKENING in generality, but to only allow it to fail under specific circumstances: where one sees to it that *p*, one does not see to it that either *p* or *q* only if it is historically necessary that either *p* or *q*. Under the analysis of making true or seeing to it as controlled truths, this seems open to obvious counterexamples. Suppose that I deliberately and freely place a die on my table to have six facing upward. In an hour, when I am asleep and unable to affect it, my friend will roll another, chancy die on the same table, which as it happens will land on a six. Then I seem to have controlled whether my first die would have a six face up on the table tonight, but not whether *any* die would have a six facing up on the table tonight (since one was going to land independent of me). But this latter fact, as I am making the first fact true, is unsettled and historically contingent if anything is; Conv (or even Conv45*), by contrast, does not impose this untoward result. The semantic analysis here on offer is begotten of an overly restrictive view of what it is we cannot make true, a view permitting only one possible source (historical settledness) of such failures of control over weak propositions.

3 Actions and Incompossibility

3.1 Action Propositions and Partitionality

A common feature of formal theories of agency that treat the actions open to an agent as propositions is that these actions are treated as *pairwise incompossible*. This is true, for example, of evidential decision theories in the style of Jeffrey [3], of the various available versions of causal decision theory [5] [34] [4], and of the aforementioned STIT models.[6] Where propositions are interpreted as subsets of a domain of worlds, this means that the set of actions available to an agent forms a *partition* on some subset of the domain.[7] In this section I will try to raise doubts about this assumption of partitionality.

Before we can enter into any substantive discussion of partitionality, we need something most of its major extant expositions lack: a more regimented account of an agent's action propositions at a world. The definition we adopt will importantly take sides on an ambiguity in the literature, which it will be illuminating to examine before proceeding.

---

[5]Or, in the atemporal semantics, letting Ch be merely reflexive rather than equivalent.

[6]For recent philosophical discussion of the nature of action propositions (or "options") in decision theory, see [35] [36] [37] [38]. None of these writers contest the pairwise incompossibility of an agent's action propositions at a world.

[7]In causal and Jeffrey-style decision theory, the actions are often assumed to form a partition on the domain itself. This is feasible because such theories ignore distinctions between propositions whose difference receives measure 0 in the agent's credence function, allowing the theory to screen off worlds at which, intuitively, the agent performs no action, since it is assumed the agent assigns probability 1 to the proposition that he will perform some action or other.

Our definition of an action proposition is: necessarily, p is an action proposition for A iff it is practically possible p be the strongest proposition A makes true. We will also sometimes speak of *the* action proposition of an agent at a world w, where this is simply the strongest proposition it makes true at w; context should disambiguate the two where necessary. Indulging in the ideology of possible worlds, and letting S_A denote the sandwich function for A, this can be clarified as follows.

Definition 3.1. *p is* an *action proposition for A at w iff there is some v practically possible from w such that $p = S_{A\,\mathsf{min}}(v)$*

Definition 3.2. *p is* the *action proposition of A at w iff $p = S_{A\,\mathsf{min}}(w)$*

Practical possibility is intended as the same modality generally denoted in the literature by *in a position to* (as in, "S is in a position to make p true"), and strength is understood in a possible worlds setting as set containment. This captures the spirit, I believe, of one prominent strain in the literature.[8]

The machinery of sandwich semantics gives us a straightforward means of modelling the relation between practical possibility and actions. An ordinary sandwich frame, as described above, consists of a non-empty domain W equipped with a (possibly partial) function $(S_{\mathsf{min}}, S_{\mathsf{max}})$ taking a member w in W to a pair of subsets in W, one containing the other and each with w as a member. To represent practical possibility (whose dual, practical necessity, we more formally denote by $\boxdot$), we may add as a further parameter a reflexive relation R on W (equivalently representable as a function from W to its power set), to be interpreted standardly as in Kripke semantics for normal logics at least as strong as KT. Under the construal of actions as the possibly strongest propositions made true, for all $w \in W$, the actions at w will be those $V \subseteq W$ such that for some $v \in R(w)$, $S_{\mathsf{min}}(v) = V$. In pictorial terms, every world is associated with a set (including itself) of worlds practically possible from it, and its action propositions are the inner circles of that set; in Alex's case, we might picture them as open discs about his pen on the paper in each world.[9]

We will appeal, in the ensuing argument, to a further schematic principle relating making true to practical possibility:

Practical Strengthening If it is not practically possible that not p, and one makes true q, one makes true that p and q $((\boxdot p \wedge \Box q) \rightarrow \Box(p \wedge q))$

As an axiom schema, this corresponds semantically (check the appendix for details) to the constraint that $S_{\mathsf{min}}(w) \subseteq R(w)$.[10] We call the logic resulting from

[8] See the passage from Lewis below, and see also our discussion above of STIT, in which the role of practical necessity is played by historical settledness; in STIT frames, settledness also satisfies the constraint of Practical Strengthening stated below.

[9] A natural question at this juncture is whether partitionality of a model is definable in our object language. The answer is No(see Theorem A.19 in the appendix). For this reason, arguments about partitionality are ineliminably arguments about models conducted in the metalanguage, rather than arguments about control in and of itself conducted in the object language.

[10] We will not in the upcoming argument need to appeal to this constraint on S and R applying universally in our model, just for one designated $\alpha \in W$, allaying possible worries about its modal status.

joining the axioms and inference rules of Conv for $\Box$ with those of KT for $\boxdot$, plus Practical Strengthening, Conv+, and it is provably sound and complete with respect to just the class of models described (call them *supplemented sandwich models*).

Practical Strengthening has considerable intuitive appeal. Suppose, for example, that Riko makes her mother's whistle squeal (by blowing it, say), and that by dint of its resonance it is beyond her power for it to squeal at any tone besides $C^\sharp$; it seems to follow she has made it squeal in $C^\sharp$, as that is the only practical way for it to squeal at all. This is an instance of a broader pattern noted earlier: when p and q are necessarily (for some sufficiently strong brand of necessity) equivalent, one makes p true iff one makes q true. Practical Strengthening simply says that practical necessity is such a necessity.[11]

This does require for plausibility a constraint on the joint interpretation of making true and practical possibility: the weaker our action propositions, the more things must be practically possible, lest the latter unduly strengthen the former. But this constraint has an intuitive basis: we should not interpret the practical modality more restrictively than need be, and naïvely the weaker the strongest things we are in a position to make true, the easier it is to make them true, and so the less restricted practical possibility has to be to accommodate the practical possibility that we make them true.

A word here on our extended semantics' relation to STIT semantics.[12] Practical necessity (or its corresponding parameter R) plays much the same role in our theory as historical necessity (and its corresponding accessibility relation) does in the logic of STIT. Whereas it is a ubiquitous assumption in the STIT literature that this is an S5 modality (i.e. its accessibility relation is equivalent), however, our own theory has only stipulated that practical necessity is factive. Why this weakening?

There are a couple of reasons not to assume at this juncture quite so strong a logic for practical necessity. To begin, that logic has deep motivations within the traditional STIT theory that do not extend to ours. In the traditional "branching time" version of the STIT semantics, it will be remembered, the points of evaluation are not *moments* in a branching time structure but *historical pairs* $\langle m, h\rangle$ of such a moment plus a *history* (maximal chain) in the structure running through that moment. In such a traditional presentation, historical necessity does not even have its own distinctive model parameter: its accessibility relation is the one obtaining between $\langle m, h\rangle, \langle m', h'\rangle$ just in case $m = m'$, which of course must be an equivalence. This fits with a committal philosophical interpretation of the modality: for p to be historically necessary, at a moment m and given a history h, is for p to be true at m given any other history h', too; historical necessity (at a historical pair) is, as it were, truth (at a moment) *irrespective* of any particular history, just as metaphysical necessity is truth (at a time) irrespective of any particular world and eternality is truth (at a world) irrespective of the time. Even

[11] Suppose $\boxdot(p \leftrightarrow q) \rightarrow (\Box p \leftrightarrow \Box q)$ for all p, q. Suppose, further, $\boxdot p_0$. Then (by normality of $\boxdot$) $\boxdot(q \leftrightarrow (p_0 \wedge q))$, so by the first supposition if $\Box q$ then $\Box(p_0 \wedge q)$. Thus, if $\boxdot p_0$ and $\Box q$, then $\Box(p_0 \wedge q)$.

[12] Thanks to anonymous reviewer for raising this question.

in the study of atemporal STIT models, the influence lingers of this guiding vision. It is this metaphysical picture that inspires (indeed, when articulated formally in terms of branching time structures, *forces*) the assumption of S5 for the operator.

Our own theory does not come along with such a specific philosophical motivation. It is agnostic about what this practical necessity (or, dually, practical possibility) fundamentally amounts to, and instead makes only the relatively weak commitment that it satisfy Practical Strengthening and be factive (which seems to be part and parcel of practical possibility being a species of *possibility* in a more than merely technical sense).[13] As long as there is *some* relevant kind of possibility satisfying Practical Strengthening, our arguments will be of interest concerning this modality. And there is no guarantee that this modality will fit the picture described in the last paragraph.

This leads into the second reason: the desirability of weak assumptions. One reason for introducing talk of practical necessity at all in this paper, beyond just its intrinsic interest, is to prove that there are no partitional supplemented sandwich models validating certain natural assumptions about realistic cases of agency (see below). It is therefore beneficial to give the weakest such assumptions possible, as this allows us to prove a stronger result. This non-existence proof still goes through, *a fortiori*, if one imposes stronger constraints on the models (such as a logic of S5 rather than KT for practical necessity). Even if practical necessity *does* have the logic of S5, it is interesting to see that our anti-partitional results do not hinge on this fact about it.

Another ubiquitous line of thinking about action propositions and practical capacity, and its similarly ubiquitous conflation with the thought formalised in Definition 3.1, finds especially clear expression in [5]. In a revealing passage, Lewis defines the action partition for an agent like so:

> Suppose we have a partition of propositions that distinguish worlds where the agent acts differently. . . . Further, he can act at will so as to make any one of these propositions hold; but he cannot act at will so as to make any proposition hold that implies but is not implied by (is properly included in) a proposition in the partition. The partition gives the most detailed specifications of his present action over which he has control. Then this is the partition of the agents' alternative *options*. (emphasis retained)

There are two thoughts going on here. Evidently, there is ours: the action propositions are those the agent "can act at will so as to make. . . hold" with no stronger such propositions contained in them. But there is another, too; two worlds will be distinguished by this partition when the agent *acts differently* in them. Acting, I take it, should here be understood as making something true, and an agent will then "act differently" in w and v when it makes something

[13]Note that, as long as a modality satisfies Practical Strengthening, factivity comes along for free, conditional on the agent making some p true: if it makes p true and q is necessary, p and q both follow by Practical Strengthening and the factivity of making true.

true in w it does not make true in v, or *vice versa*. So understood, partitionality is immediate.[14]

These are not, from our perspective, the same proposal at all. For there is nothing in our basic formalism of sandwich models to compel partitionality when action propositions are understood in the first way, and yet it follows trivially from the second. Why, then, does Lewis slide between them so readily? And why, if the cells of his partition need not be controllable by the agent, are they recommended to the agent in his decision theory? Why propose "choices" one cannot choose?

The most natural response I can give on his (and, by extension, the tradition's) behalf, though he nowhere states it overtly, is that for him *we control what it is we make true*. There is (besides the tautologous interpretation) a way of reading this as saying the same as 4, which as we have seen merely requires that one's action proposition contract while moving within it and thus brings us no closer to our solution. The intended reading is that at w the agent controls (the big conjunction specifying) *exactly* which propositions it makes (and does *not* make) true at w. Given this, the equivalence of the two definitions follows immediately.

This assumption is in effect just to accept our logic Conv45*, for the constraint it imposes on a sandwich frame is the very same (see Fact A.16). Now accepting Conv45* is not required for partitionality, since as formulated partitionality unlike Conv45* places no constraints on S_{max}. But it is the best way I can see to motivate the view from the understanding of action propositions we have adopted, and as we will see our reasons for rejecting partitionality will hinge on questioning it. But before passing to these reasons, an excursus on the semantics.

3.2 The Meaning of S

One problem to present itself more clearly, with the notions of action proposition and practical possibility now in view, is the intuitive interpretation of our semantics' distinctive parameters, S_{min} and S_{max}. To what real analogues do the sets they select at an index correspond? And given this, what further constraints should we impose on the interaction between S and R in supplemented frames?[15]

One captivating proposal links all three parameters S_{min}, S_{max}, and R together with a *type of actions*. In such a framework, in addition to and more basic than the *(action) propositions* an agent can *make true*, there is a type α of actions (considered as *events*) an agent can *perform*. Indeed, at any given world w there

[14] There are hints at a similar conflation between enact*ing* and enact*ed* in other *loci classici* of action partitions, though less clearly stated. Thus Joyce can, in summarising and endorsing Jeffrey, identify actions "with propositions that a decision maker can make true or false as she pleases" even as both authors exclusively use sentences *describing an agent making something true* to denote action propositions [4]. I take the Lewis passage as my main focus of interpretation because of both the foundational role of the discussion containing it and its relevant explicitness.

[15] Thanks to an anonymous reviewer for pressing these issues.

is some unique action α_w the agent performs, and the proposition *that* an agent performs α_w (represented semantically by $\{w' : \alpha_{w'} = \alpha_w\}$) will be its action proposition at w; partitionality, notably, falls out of this proposal directly, given that = is an equivalence relation. Such a vision of action weds the Davidsonian belief in the primacy of events with the STIT presentation of action as making true, and is naturally captured by STIT logics directly involving act-names.[16]

This accounts for S_{min}; to get S_{max} we once again invoke practical possibility and its corresponding accessibility relation R. Clearly, at any given w, there will be some disjoint set $\mathcal{A}_w$ of action propositions available to the agent at w, corresponding to the members of α the agent can practically possibly perform at w. $S_{\mathsf{max}}(w)$ is the join of $\mathcal{A}_w$, representing the *scope* or *limits* of the agent's control at w; this guarantees it will always be a superset of the agent's action proposition, and if we assume additionally and plausibly that practical possibility is an equivalence (or even just transitive) relation we also guarantee that the frames will validate $\mathsf{Conv45}^*$, which as we saw above fits especially cleanly with partitionality.[17] Thus, we derive a neat hierarchy in the metaphysics of action, in which α underlies S_{min}, and S_{min} in turn with R underlies S_{max}, predicting the constraints desired.

Despite its tidiness, I reject this reading of the semantics. We will see in the ensuing section reason to reject partitionality altogether, but even more damningly it requires that ones S_{max} at a world be very strong, since it can include no practically impossible worlds. I seem, currently, to be making my knee rest by my shoulder, but this comes up against obvious counterexamples if we assume that practical possibility has at least the logic of $\mathsf{S4}$. After all, my knee could (in a very weak sense) have rested by my shoulder had Corbyn won the UK general election in 2019, which means S_{max} must include such worlds, even as none are now practically possible for me.[18] My joint conviction both that I have no power over the 2019 election and that I make my knee rest by my shoulder is stronger than any theoretical attractions of the above picture.

My preferred reading of the semantics rejects the original contention of action-uniqueness. If we do countenance the type α of event-actions, we should replace its corresponding *function* with a *relation* the agent bears at a given world to (possibly) *many* event-actions. This sits better with our understanding of

[16] See e.g. [39] [40]; note that the use of action labels is in these papers developed in the context of multi-agent and collective STIT logics, which introduce subtleties well beyond our scope.

[17] Suppose the agent performs α_w at w. Clearly, for $w' \in S_{\mathsf{min}}(w) = \{w' : \alpha_{w'} = \alpha_w\}$, $S_{\mathsf{min}}(w') = S_{\mathsf{min}}(w)$ by partitionality. Moreover, by transitivity of R and $S_{\mathsf{min}}(w) \subseteq R(w)$, there are no new $v \in R(w'), \notin R(w)$, so by construction of $S_{\mathsf{max}}(*)$ as $\bigcup \mathcal{A}_*$, $S_{\mathsf{max}}(w') \subseteq S_{\mathsf{max}}(w)$. Since $S_{\mathsf{min}}(w) = S_{\mathsf{min}}(w') \subseteq R(w')$, similarly by transitivity $S_{\mathsf{max}}(w') \supseteq S_{\mathsf{max}}(w)$. Thus, $S_{\mathsf{max}}(w') = S_{\mathsf{max}}(w)$ and $S(w) = S(w')$.

Note that this validates a constraint like Practical Strengthening for S_{max}. For a given w and $w' \in R(w)$, $S_{\mathsf{min}}(w') \subseteq R(w')$, so by transitivity of R we have $R(w) \supseteq S_{\mathsf{min}}(w')$ for all $w' \in R(w)$, meaning also $R(w) \supseteq \mathcal{A}_w$. So, $S_{\mathsf{max}}(w) \subseteq R(w)$; in fact, if it is practically impossible at w the agent perform *no* action, $S_{\mathsf{max}}(w) = R(w)$, since then $u \in S_{\mathsf{min}}(u) \subseteq \mathcal{A}_w$ for each $u \in R(w)$. This is, recall, the same structure exhibited by the STIT analogues Ch, H of $(S_{\mathsf{min}}, S_{\mathsf{max}})$.

[18] Indeed, I think no Corbyn-victorious world is now *practically possibly* practically possible for me (or, that it is *historically settled* that it is historically settled Corbyn lost), which would undercut any need for $\mathsf{S4}$ in this argument.

realistic agents, like you or me. At any given world w, there are several actions we will be performing at once (typing, sitting, moving ones eyes, etc.), and we will in turn *make true* that these actions *occur*. Given our earlier axioms for Conv, these enacted propositions (about the occurrence of event-actions) determine further enacted propositions (closing under conjunctions, disjunctions, and logical intermediates), which together constitute the *convex sublattice* (in the Boolean lattice of propositions, i.e. $(\mathcal{P}(W), \subseteq)$) of facts made true at w. A truly single-minded agent, who performed only one action whatsoever, would make true nothing beyond the fact of the action's occurrence.

This picture upends the hierarchy of the last. S_{min} and S_{max}, rather than playing any central role themselves, fall out as just the meet and join respectively of the "generating" action-occurrence propositions; this point is made vivid by observing that, as long as this "generating" set is infinite, the finitary closure rules corresponding to our axioms for Conv need not determine any such bound propositions, and indeed it is easy to see how our semantics might be generalised without technical loss by replacing the codomain of S with the (possibly unbounded) convex sublattices of $(\mathcal{P}(W), \subseteq)$ instead of the pairs $X_1 \subseteq X_2$ in $\mathcal{P}(W)^2$. Generalised infinitary versions of these rules, which would guarantee bounds on the sublattice at a given world, are indeed appealing independently, but again it is the closure rules giving rise to the bounds and not *vice versa*. This is a vision on which it is the axioms themselves explaining the value of the semantics, not fundamentally semantic intuitions underwriting the plausibility of the axioms.

3.3 A Problem Case

We can now proceed to the case against partitionality. As a warmup, here is a toy case that intuitively conflicts with the principle. Suppose that Alex, a man with typical human fine motor control and under ordinary circumstances, places his pen idly somewhere on a 300cm-wide square sheet of paper, leaving behind a spot of ink. Assuming for the example that the only propositions he is practically in a position to make true are those about where the pen falls on the sheet, partitionality is deeply counterintuitive here. There should be, given some minimal assumptions, a smallest region (set of points on the sheet) R of the paper such that Alex makes the pen fall centred somewhere in R. And for every different way Alex could have acted, there will be some other, counterfactual smallest region in which he makes the pen fall. These regions are, in effect, his action propositions in the toy case.

What partitionality (as restricted, again, to propositions about where the pen lands) requires is that these regions form a partition on (a subregion of) the sheet's surface. There is something deeply strange about this. Intuitively we would think his "action region" in a given practically possible world should contain a buffer of ϵ cm around the point where the pen actually falls; ordinary motor control is not good enough to forestall miniscule deviations from one's actual path in a certain direction. And yet this would require, under partitionality, that there be some points (namely those in a given action region less than ϵ

cm from its border) on the sheet Alex simply cannot touch with the pen.[19] What force would prevent him from so placing it? And even absent these margin of error considerations, whence comes this particular partition on the sheet? The whole proposal smacks of arbitrariness.

This is so far just a picture, not an argument. Alex makes many more facts true than just ones about the location of his pen on the paper, and nothing we have said precludes these stronger enacted propositions from ensuring his action propositions satisfy partitionality. Thus, for example, while the interior of the 1cm circle around a point x might be (in the world where he hits x) the smallest region in which he makes his pen fall, perhaps at each such world he *chooses* that it will fall inside the 1cm circle about the struck point (and thus *makes himself make* the pen fall there); thus, at two worlds x_1 and x_2 within 1cm of one another, while his action *regions* overlap, his action *propositions* at the two are disjoint, for one entails, and the other precludes, *that he makes the pen fall within 1cm of* x_1. This is particularly natural if we distinguish, as in the first proposal from the last section, his event-action at a world from the proposition(s) it determines; as, on this view, his event-action at a world is unique, his event-actions at x_1 and x_2 *must* be distinct (and his action propositions disjoint) if his action regions at both are not identical, since his action fixes what he makes true. It thus remains to convert the above picture into a tighter argument.

The tip of Alex's pen, we can stipulate, is centred on a point η near the middle of the sheet. In such a case, the following all seem true:

3. It is practically necessary that, if the pen falls within 500,000 μm of η on the sheet, Alex makes it appear somewhere on the sheet
$$\boxdot(q_{500,000} \rightarrow \Box p)$$
4. Alex makes the pen fall on the sheet less than 500,000 μm from η
$$\Box(p \wedge q_{500,000})$$
5. Alex does not make the pen fall on the sheet less than 1 μm from η [20]
$$\neg\Box(p \wedge q_1)$$

Plus the following family of premises indexed to $0 < i \leq 500,000$:

C$_i$ Alex does not make it so that he both makes the pen fall less than i μm from η on the sheet and does not make it fall less than $i - 1$ μm from η on the sheet[21]
$$\neg\Box(\Box(p \wedge q_i) \wedge \neg\Box(p \wedge q_{i-1}))$$

[19] This framing is, of course, partly adapted from the discussion of "luminosity" in [41]. We will not in the main argument, however, need to appeal to the kinds of margin of error principles central to that discussion.

[20] For the argument immediately at hand we could replace 1 with 0, leaving the uncontrolled proposition contradictory, though this (slightly) stronger concession will be dialectically useful in discussing the models of the next section.

[21] This assumption bears a close and nonaccidental resemblance to the gap principles known in the theory of vagueness due to the late Graff Fara [42] [43].

(4) and (5) require little comment. (5), in particular, simply follows from the description of the case, plus a lack of superhuman precision. (3) is justified by supposing that only Alex would put the pen on the sheet. The real work, clearly, is to be done by (C_i), which imposes a limit on Alex's *higher order control*, his control over *what* he controls.

It is a trivial consequence of (4) and (5) that there is some i such that Alex makes the pen fall less than i, but not less than $i-1$, μm from η; assuming that, for all $j < k$, if it falls less than j μm it falls less than k μm (which follows by Convexity if their disjunction is equivalent to the second disjunct), this i will indeed be *unique*. $\Box(p \wedge q_i)$ and $\neg\Box(p \wedge q_{i-1})$ then are a (the) limit to Alex's fine motor control over where the pen falls on the paper, up to micrometres from η. What (C_i) says is that the degree of Alex's fine motor control thus expressed is not *itself* something he controls, nor is it within his power to control it. This claim, that ordinary humans like Alex do not and cannot fix the precise limits of their own motor control in action, seems eminently plausible.

Note that (C_i), as an expression of higher order non-control, is (when conjoined with Alex's limit $\Box(p \wedge q_i) \wedge \neg\Box(p \wedge q_{i-1})$) simply a negated instance of 5^*. This provides a deeper insight into Conv5^*: it is the logic of agents with perfect higher-order control, and logics imposing 5^* (like that of [*dstit*]) can only describe such idealised agents.

These assumptions are collectively incompatible with partitionality:

Fact 3.3. *There is no supplemented sandwich model in which (3 - 5) and all (C_i) are true at a world w while the actions at w are pairwise disjoint.*[22]

A simple proof of this fact is given in the appendix, but the thrust behind it should be easy to see. Throughout the inner circle of w, p is made true, and thus when propositions stronger than p contain this inner circle (assuming the inner circle function is constant throughout the inner circle of w, which is forced by partitionality), they are made true throughout the circle. This will include both $\Box(p \wedge q_i)$ and $p \wedge \neg\Box(p \wedge q_{i-1})$, and so also their conjunction, which directly conflicts with (C_i).

There is an obvious objection to this argument. Many philosophers who appeal to the concept of an action proposition think that, in the relevant sense of *making true*, we make true only much weaker, more "internal" propositions, about (say) the state of our volition and not what appears on sheets of paper.[23] Thus the objector will not concede (3) or (4), undermining the argument.

[22] As to why the argument is formulated in the metalanguage in terms of models rather than directly in our object language itself, see fn. 9.

[23] Saith Joyce: "My inclination is to [. . .] use the term *act* narrowly to denote *pure, present exercises of the will*. Many things we ordinarily call acts do not count as such under this reading. Walking to work, for example, is not really an act one can choose to perform because, unfortunately, one cannot simply will it to be the case that one's legs function properly, that one will not be shot by a madman on the way out the front door, that a great chasm will not open in the earth to prevent one from reaching one's destination, or even that one's 'future self' will carry through on one's choice." See also Ford's discussion of "volitionalism" in [44] and the discussion in [45] of "tryings" in decision theory.

Model 1: Naïve representation of Alex

W	$[0,300]^2_{x,y} \subset \mathbb{R}^2$
R	$\{(w,v) : w, v \in W\}$
$S_{\mathsf{min}}(w)$	$\{v : \lvert w_x - v_x\rvert^2 + \lvert w_y - v_y\rvert^2 < 1\}$
$S_{\mathsf{max}}(w)$	W
$V(p\ ;\ q_i^w)$	$W\ ;\ \{v : \lvert w_x - v_x\rvert^2 + \lvert w_y - v_y\rvert^2 < i/10{,}000\}$

But there is a simple fix. The volitionalist objector should at least grant, for example, that Alex makes true that the pen *would* fall less than 500,000 μm from η, were the ordinary and limited causal dependence between the state of the paper and Alex's will to hold. With these intuitions in mind, letting o stand for *the ordinary and limited causal dependence obtains* and $>$ for the subjunctive conditional, replace (3), (4), (5), and (C$_i$) with

- $\boxdot((o > q_{500{,}000}) \rightarrow \Box(o > p))$
- $\Box(o > (p \wedge q_{500{,}000}))$
- $\neg\Box(o > (p \wedge q_1))$
- $\neg\Box(\Box(o > (p \wedge q_i)) \wedge \neg\Box(o > (p \wedge q_{i-1})))$

respectively. Under minimal assumptions about the subjunctive conditional $>$ (in particular that $o >$ is a normal operator), the argument will proceed just as cleanly as before.

3.4 Modelling Clumsy Agency

Of course, this on its own is only half of an argument against partitionality; to dispel worries about the assumptions going into it, we must provide an at least somewhat realistic supplemented model validating those assumptions.

Our initial discussion of Alex suggests one such. Have the domain of the model be the real coordinates on the sheet, the accessibility relation R the universal relation, any point's inner circle the points with 1cm of it, and the outer circle the whole sheet.

The model then interprets worlds as coordinates on the sheet of paper on which to place the pen, takes it to be practically possible to place the pen anywhere, and has it that Alex's action proposition at w is the disc of worlds less than 1cm from w while he makes it true everywhere that the pen falls somewhere.[24] V gives our propositional constants p and q_i^w (for $w \in W, i \in \mathbb{N}$) their natural interpretations, with q_i^w standing for the analogue with respect to w of q_i for η.

[24] The choice of 1cm is of course simply for convenience; feel free to replace it with some more plausible distance if desired.

Model 2: Adding a degree-of-control parameter

W	$([0,300]^2_{x,y} \times (0,300]_d) \subset \mathbb{R}^3$
R	$\{(w,v) : w, v \in W\}$
$S_{\mathsf{min}}(w)$	for some $a, b \in (0,300], a \leq w_d \leq b$: $\{v : \lvert w_x - v_x\rvert^2 + \lvert w_y - v_y\rvert^2 < 1, a \leq v_d \leq b\}$
$S_{\mathsf{max}}(w)$	W
$V(p\ ;\ q^w_i)$	$W\ ;\ \{v : \lvert w_x - v_x\rvert^2 + \lvert w_y - v_y\rvert^2 < i/10,000\}$

While this indeed validates all our assumptions, it suffers from serious problems of unrealism. For while it does validate ($C_{10,000}$) (the relevant such premise) at η, it affirms his perfect higher order control in a deeper sense. In this model, the disjunction

$$\mathbf{D}_{10,000} \quad \bigvee_{w \in W} (\Box(p \wedge q^w_{10,000}) \wedge \neg\Box(p \wedge q^w_{9,999}))$$

is true everywhere, and as $S_{\mathsf{max}}(w) = W$ for all w, Alex makes true this disjunction, effectively controlling his precise degree of motor control considered independently of where he actually places the pen (his *location-neutral* such degree of control). It is tempting to respond that, as a toy model, we should expect such artificialities, but this would be to selfishly excuse the very flaw we accused in the friends of partitionality. Denying ($C_{10,000}$) is bad; this is worse.

An initial impulse might be to add a third parameter d of *degree of control* into the indices in the domain, represented by $(0, 300] \subset \mathbb{R}$ as the radii of the action propositions for each degree. Thinking of the domain as a 300cm^3 cube with the point-thick top layer shaved off, at a given point one's action proposition would be the interior of a cylinder of radius d centred about its vertical axis (plus its top and bottom).

This would make good on the idea that Alex's is but one possible degree of motor control among many. But, however implemented, it introduces more problems than it solves.

For one, it guarantees by the same reasoning Alex's yet higher order control over his *control over his degrees of control itself*, control even *more* dubious than the higher order control the model was just revised to deny, thus simply kicking the problem up one level. We could, of course, introduce higher levels of control infinitely or indefinitely, but even barring the obvious problem of his control over his *infinitely* higher order control, this method of amending the model has the perverse quality of making yet more unreasonable claims of Alex at each finitary step. It abandons the frying pan by way of an infinitely nested worsening series of fires.

Second, implementing it usefully requires abandoning the very intuitions supporting (C_i) in the first place. Our first suspicion about Alex is that his practical possibilities are limited to ones with his current degree of control, that he has no higher control over this degree, and that these facts are mutually reinforcing rather than in tension. But here they are placed precisely in *conflict*.

Model 3: Adding a paper-avoiding possibility

W	$[0,300]^2_{x,y} + 1$
R	$\{(w,v) : w, v \in W\}$
$S(1)$	$(\{1\}, \{1\})$
$S_{\mathsf{min}}(w \in [0,300]^2)$	$\{v : \lvert w_x - v_x\rvert^2 + \lvert w_y - v_y\rvert^2 < 1\}$
$S_{\mathsf{max}}(w \in [0,300]^2))$	$[0,300]^2 \subset \mathbb{R}^2$
$V(p\ ;\ q_i^w)$	$[0,300]^2 \subset \mathbb{R}^2\ ;\ \{v : \lvert w_x - v_x\rvert^2 + \lvert w_y - v_y\rvert^2 < i/10,000\}$

If Alex's practical possibilities all share his actual degree of motor control (that is, wRv only if $w_d = v_d$), then by Practical Strengthening, Convexity, and $S_{\mathsf{max}}(w) = W$, Alex controls his exact degree of control (in sandwich speak, $S_{\mathsf{min}}(w) \subseteq \{v : v_d = w_d\} \subseteq S_{\mathsf{max}}(w)$; pictorially, his cylindrical action proposition would be flattened to a disc). A less toyish model will of course require multiple degrees of control available at a given coordinate of the page, but this crude introduction of them does more harm than good.

A better, and simpler, approach instead starts by adding an additional world representing the possibility of simply not placing the pen on the sheet at all. We take crucial advantage of the convex non-normality of our operator: this world's singleton is its own inner and outer circle, while for the rest we have $S_{\mathsf{max}}(w) = [0,300]^2$ and S_{min} as before. When hitting the sheet, that is the least he controls; when not hitting it, the same. R is, again, universal.

We still have, of course, Alex controlling ($\mathrm{D}_{10,000}$) as long as he hits the sheet, but in this setting we should take pains to distinguish ($\mathrm{D}_{10,000}$) from the conditional claim (for some appropriate kind of conditional, be it material or counterfactual or strict) that *if* he hits the sheet, ($\mathrm{D}_{10,000}$) holds. In our earlier model, these were both simply the trivial proposition, but here only the latter is trivial, and only the former is made true (a distinction made possible by the pecularities of our logic for $\Box$).

It is this conditional, I think, that represents in this setting the location-neutral degree of Alex's motor control considered as a *fixed capacity* of his, the sense in which his controlling it strikes us as absurd. He of course controls the degree of control he exercises in placing the pen on the sheet (in the sense of making true ($\mathrm{D}_{10,000}$)), but this is simply because for an agent with his particular limited motor capacities to place a pen on the sheet is just to place it while exercising exactly those limited capacities; for a helplessly clumsy thrower, in the same vein, to toss a ball just is to toss it clumsily, and indeed to toss it with their exact kind of clumsiness. Over his background musculature and neurology, over the dispositional sensitivity of his pen to mild pressure and sudden movements, Alex thereby need exercise no control at all.

If the claim that Alex controls his exercised degree of control down to the micrometre still bothers you, we can introduce still more relevant verisimilitude by rejecting its constancy across the sheet. We could instead have, say, the inner circle of w be all points within a random $f(w)$ between 1cm and 1.5cm. $f(w)$

Model 4: Degrees of control varying by location

W	$[0, 300]^2_{x,y} + 1$
R	$\{(w, v) : w, v \in W\}$
$S(1)$	$(\{1\}, \{1\})$
$S_{\mathsf{min}}(w \in [0, 300]^2)$	$\{v : \lvert w_x - v_x \rvert^2 + \lvert w_y - v_y \rvert^2 < f(w)\}$ for $f : [0, 300]^2 \subset \mathbb{R}^2 \to [1, 1.5] \subset \mathbb{R}$
$S_{\mathsf{max}}(w \in [0, 300]^2))$	$[0, 300]^2 \subset \mathbb{R}^2$
$V(p\ ;\ q^w_i)$	$[0, 300]^2 \subset \mathbb{R}^2\ ;\ \{v : \lvert w_x - v_x \rvert^2 + \lvert w_y - v_y \rvert^2 < i/10,000\}$

represents the motor control Alex *would* exercise at w specifically.[25]

This respects the intuition that the degree of motor control Alex would exhibit in placing his pen down somewhere on the paper is, to some extent, itself variable; we in our bodily motions like placing down a pen leave up to chance, not just where the pen lands, but how much control we exert over this. This leaves, of course, him controlling the more varied disjunction

$$\mathbf{D}'_{10,000} \quad \bigvee_{w \in W} (\Box(p \wedge q^w_{10000f(w)}) \wedge \neg\Box(p \wedge q^w_{10000f(w)-1}))$$

but there is something less strikingly impressive about this feat. He controls his (location-neutral, exercised) control down to the micrometre, but not to a *consistent* micrometre.

Do these remarks undermine the plausibility of (C_i)? No. For what renders unthreatening Alex's control over ($D_{10,000}$) is its practical equivalence with an obviously easy proposition for him to make true, namely that he hits the paper at all. For him to make true the exact boundary of his control in hitting η (that is, for the relevant instance of (C_i) to be false), however, would correspondingly be for him in hitting η to possibly control down to a micrometre's variation *where exactly he actually hits the paper*, which has little pre-theoretic intuitive plausibility (whence our conviction in (5)).

Model 3 straightforwardly satisfies for each candidate for η all our earlier assumptions. For Model 4, for an evenly distributed finite set X of points within $S_{\mathsf{min}}(w)$ the probability that, for all $x \in X$, $S_{\mathsf{min}}(x) \subseteq S_{\mathsf{min}}(w)$ approaches zero as $|X| \to \infty$ given natural assumptions about the probability distribution on the values of $f(w)$, so that almost certainly our assumptions from the last section are true almost everywhere.

[25] This assumes a restricted version of counterfactual excluded middle. For detractors who take it the counterfactual degree of control exercised at a given coordinate is indeterminate, add to the domain for each relevant counterfactual possibility a copy w of that coordinate with the appropriate $S_{\mathsf{min}}(w)$.

Conclusion

In the foregoing I have attempted to justify, both on its own terms and against its most prominent rivals, a basic modal logic of control and making so. It has, I argue, not merely an independent plausibility, but a flexibility and naturalness–semantic as well as syntactic–suited for framing fundamental questions about the concept it formalises. As a proposal it is open-ended: it is offered only as a lower limit on the strength of the "true" logic of the operator, including its interactions with other modalities of note. It is my hope that as a starting point it can aid in resolving outstanding questions, and illuminating new such questions of interest, concerning the phenomenon it seeks to formalise.

A Appendix

We here prove relevant results for the logics discussed above. Proofs are largely adapted from analogous ones in [46] and [24].

A.1 The Logic Conv

Our language $\mathcal{L}$ (or equivalently, its set of wff's) is as usual that generated by a countable set At of sentential variables $p, q, r, \ldots$ and

$$\mathsf{At} \mid \neg\varphi \mid \varphi \vee \psi \mid \varphi \wedge \psi \mid \varphi \rightarrow \psi \mid \Box\varphi$$

Our logic itself, $\mathsf{Conv} \subset \mathcal{L}$, i.e. the φ such that $\vdash_{\mathsf{Conv}} \varphi$, is the set of wff's generated by all instances of the following axiom schemas

All classical tautologies

T $\Box\varphi \rightarrow \varphi$

M $(\Box\varphi \wedge \Box\psi) \rightarrow \Box(\varphi \wedge \psi)$

O $(\Box\varphi \wedge \Box\psi) \rightarrow \Box(\varphi \vee \psi)$

Conv $\Box(\varphi \vee \psi \vee \chi) \rightarrow \Box\varphi \rightarrow \Box(\varphi \vee \psi)$

with the following inference rules

MP if $\vdash_{\mathsf{Conv}} \varphi$ and $\vdash_{\mathsf{Conv}} \varphi \rightarrow \psi$, $\vdash_{\mathsf{Conv}} \psi$

RE if $\vdash_{\mathsf{Conv}} \psi \leftrightarrow \phi$, $\vdash_{\mathsf{Conv}} \Box\psi \leftrightarrow \Box\phi$

A model $\mathcal{M}$ of our logic is a triple $\langle W, V, (S_{\mathsf{min}}, S_{\mathsf{max}})\rangle$, with W some nonempty set, $V : \mathsf{At} \rightarrow \mathcal{P}(W)$, and the possibly partial $(S_{\mathsf{min}}, S_{\mathsf{max}}) : W \hookrightarrow \mathcal{P}(W) \times \mathcal{P}(W)$ (or just S) such that (where defined) $\{w\} \subseteq S_{\mathsf{min}}(w) \subseteq S_{\mathsf{max}}(w)$. The clauses of the interpretation function $[\![\ ]\!]$ for the atoms and Boolean clauses are as usual, while $[\![\Box\varphi]\!] = \{w \in W : S_{\mathsf{min}}(w) \subseteq [\![\varphi]\!] \subseteq S_{\mathsf{max}}(w)\}$. (Note that $w \notin [\![\Box\varphi]\!]$ when

S is undefined on w, and that as components of a single function from W into the cartesian square of $\mathcal{P}(W)$ S_{min} and S_{max} are each defined iff the other is.) φ is *valid* in a model $\mathcal{M}$, i.e. $\vDash_{\mathcal{M}} \varphi$, iff $[\![\varphi]\!]_{\mathcal{M}} = W_{\mathcal{M}}$, and is valid simpliciter, i.e. $\vDash_{\mathsf{Conv}} \varphi$, iff for all models $\mathcal{M}$, $\vDash_{\mathcal{M}} \varphi$.

We now prove a lemma that will be useful in proving completeness.

Lemma A.1. *If* $\vdash_{\mathsf{Conv}} (\varphi \wedge \psi) \to \chi$ *and* $\vdash_{\mathsf{Conv}} (\neg\varphi \wedge \neg\psi) \to \neg\chi$, $\vdash_{\mathsf{Conv}} (\Box\varphi \wedge \Box\psi) \to \Box\chi$.

Proof. Suppose the antecedent. Then by classical reasoning, we have $\vdash_{\mathsf{Conv}} (\varphi \wedge \psi) \to \chi$ and $\vdash_{\mathsf{Conv}} \chi \to (\varphi \vee \psi)$. Now assume $\Box\varphi \wedge \Box\psi$. By M this gives us $\Box(\varphi \wedge \psi)$, and by O $\Box(\varphi \vee \psi)$. But then by **Conv** and **RE** we have $\Box\chi$, and discharging we have $\vdash_{\mathsf{Conv}} (\Box\varphi \wedge \Box\psi) \to \Box\chi$. □

Corollary A.2. *For any finite* $\{\varphi_1, \dots, \varphi_n, \psi\}$, *if* $\vdash_{\mathsf{Conv}} (\varphi_1 \wedge \dots \wedge \varphi_n) \to \psi$ *and* $\vdash_{\mathsf{Conv}} (\neg\varphi_1 \wedge \dots \wedge \neg\varphi_n) \to \neg\psi$, $\vdash_{\mathsf{Conv}} (\Box\varphi_1 \wedge \dots \wedge \Box\varphi_n) \to \Box\psi$.

A.2 Soundness

It is easy to see that our semantics is sound.

Theorem A.3. *If* $\vdash_{\mathsf{Conv}} \varphi$, $\vDash_{\mathsf{Conv}} \varphi$

Proof. We omit as routine the proofs of the validity of classical tautologies and **MP**. The validity of T is straightforward from the stipulation that (when defined) $\{w\} \subseteq S_{\mathsf{min}}(w)$. The validity of M is straightforward from the fact that for $X, Y \supseteq S_{\mathsf{min}}(w)$, $X \cap Y \supseteq S_{\mathsf{min}}(w)$. The validity of O is straightforward from the fact that for $X, Y \subseteq S_{\mathsf{max}}(w)$, $X \cup Y \subseteq S_{\mathsf{max}}(w)$. The validity of **Conv** is straightforward from the fact that, for any X, Y, Z, $X \subseteq (X \cup Y) \subseteq (X \cup Y \cup Z)$. By induction, for any φ, ψ provably equivalent and for which our theorem holds, in all $\mathcal{M}$, $[\![\varphi]\!] = [\![\psi]\!]$, validating **RE** by the intensionality of the clause for $\Box$. □

A.3 Completeness

Let a *consistent set* Γ be a set of wff's in $\mathcal{L}$ such that for no $\varphi_1 \dots \varphi_n \in \mathcal{L}$ (for finite n) do we have $\vdash_{\mathsf{Conv}} \neg(\varphi_1 \wedge \dots \wedge \varphi_n)$. Let a *maximal consistent set* Γ be a consistent set such that for all $\varphi \in \mathcal{L}$ either $\varphi \in \Gamma$ or $\neg\varphi \in \Gamma$. The proof that for any consistent set Γ there is a maximal consistent set Γ' such that $\Gamma \subseteq \Gamma'$ is as usual. We denote the set of maximal consistent sets in $\mathcal{P}(\mathcal{L})$ as $\mathcal{L}_{\mathsf{M}}$.

We define the *canonical model* $\mathcal{M}_{\mathsf{canon}}$, as per the above, as the triple $\langle W, V, (S_{\mathsf{min}}, S_{\mathsf{max}})\rangle$ such that

- $W_{\mathcal{M}_{\mathsf{canon}}} = \mathcal{L}_{\mathsf{M}}$
- For $p \in \mathsf{At}$, $V_{\mathcal{M}_{\mathsf{canon}}}(p) = \{w \in W_{\mathcal{M}_{\mathsf{canon}}} : p \in w\}$
- $S_{\mathsf{min}\mathcal{M}_{\mathsf{canon}}}(w') = \{w \in W_{\mathcal{M}_{\mathsf{canon}}} : \text{for all } \Box\varphi \in w', \varphi \in w\}$ unless no $\Box\varphi \in w'$, in which case undefined

- $S_{\mathsf{max}\mathcal{M}_{\mathsf{canon}}}(w') = \{w \in W_{\mathcal{M}_{\mathsf{canon}}} : \text{for some } \Box\varphi \in w', \varphi \in w\}$ unless no $\Box\varphi \in w'$, in which case undefined

Lemma A.4. *Let Γ be some consistent set of wff's of the form $\Box\varphi$, $\Gamma_{-\Box}$ be the set obtained by exchanging φ for each $\Box\varphi \in \Gamma$, and $\neg\Gamma_{-\Box}$ the set obtained by negating each member of $\Gamma_{-\Box}$. Then if, for arbitrary $\neg\Box\psi$, we have $\Gamma \cup \{\neg\Box\psi\}$ consistent, we have either $\Gamma_{-\Box} \cup \{\neg\psi\}$ consistent or $\neg\Gamma_{-\Box} \cup \{\psi\}$ consistent.*

Proof. Suppose that $\Gamma \cup \{\neg\Box\psi\}$ is but neither $\Gamma_{-\Box} \cup \{\neg\psi\}$ nor $\neg\Gamma_{-\Box} \cup \{\psi\}$ also is consistent. Then there must be some finite $\{\gamma_1, \ldots, \gamma_n\} \subseteq \Gamma_{-\Box}$ where $\vdash_{\mathsf{Conv}} \neg(\gamma_1 \wedge \ldots \wedge \gamma_n \wedge \neg\psi)$ and some (not necessarily disjoint) $\{\gamma_{n+1}, \ldots, \gamma_m\} \subseteq \Gamma_{-\Box}$ where $\vdash_{\mathsf{Conv}} \neg(\neg\gamma_{n+1} \wedge \ldots \wedge \neg\gamma_m \wedge \psi)$. Indeed, by the monotonocity of inconsistency, we may simplify by saying we have both $\vdash_{\mathsf{Conv}} \neg(\gamma_1 \wedge \ldots \wedge \gamma_m \wedge \neg\psi)$ and $\vdash_{\mathsf{Conv}} \neg(\neg\gamma_1 \wedge \ldots \wedge \neg\gamma_m \wedge \psi)$.

By classical reasoning and the first conjunct, we have $\vdash_{\mathsf{Conv}} (\gamma_1 \wedge \ldots \wedge \gamma_m) \rightarrow \psi$. By classical reasoning and the second conjunct, we have $\vdash_{\mathsf{Conv}} (\neg\gamma_1 \wedge \ldots \wedge \neg\gamma_m) \rightarrow \neg\psi$. But since by supposition $\Gamma \cup \{\neg\Box\psi\}$ is consistent, we have $\nvdash_{\mathsf{Conv}} (\Box\gamma_1 \wedge \ldots \wedge \Box\gamma_m) \rightarrow \Box\psi$. But by Corollary A.2, this is impossible. □

Theorem A.5. $w \in [\![\varphi]\!]_{\mathcal{M}_{\mathsf{canon}}}$ *iff* $\varphi \in w$

Proof. The proof proceeds by induction. It is straightforward to see that it holds in the base step for all $p \in \mathsf{At}$, as the interpretation function $[\![]\!]$ in that case simply reduces to the assignment function V, which in turn by construction reduces in $\mathcal{M}_{\mathsf{canon}}$ to $\in$. We omit as routine the induction steps for the Boolean connectives, leaving only our operator $\Box$.

We first prove right to left. Suppose $\Box\varphi \in w$. Now by hypothesis we have $w' \in [\![\varphi]\!]_{\mathcal{M}_{\mathsf{canon}}}$ iff $\varphi \in w'$; by construction we then have for all $w' \in S_{\mathsf{min}\mathcal{M}_{\mathsf{canon}}}(w)$ that $\varphi \in w'$, and thus $[\![\varphi]\!]_{\mathcal{M}_{\mathsf{canon}}} \supseteq S_{\mathsf{min}\mathcal{M}_{\mathsf{canon}}}(w)$. Moreover, by construction of $S_{\mathsf{max}\mathcal{M}_{\mathsf{canon}}}$ and by the inductive hypothesis, $[\![\varphi]\!]_{\mathcal{M}_{\mathsf{canon}}} \subseteq S_{\mathsf{max}\mathcal{M}_{\mathsf{canon}}}(w)$. So by construction of $[\![]\!]$ we have $w \in [\![\Box\varphi]\!]_{\mathcal{M}_{\mathsf{canon}}}$.

Now suppose $\Box\varphi \notin w$, i.e. (by maximality) $\neg\Box\varphi \in w$. Either $S_{\mathcal{M}_{\mathsf{canon}}}(w)$ is defined or not. If not, then by stipulation $w \notin [\![\Box\varphi]\!]_{\mathcal{M}_{\mathsf{canon}}}$. Now let Γ be $\{\Box\psi : \Box\psi \in w\}$; then, if $S_{\mathcal{M}_{\mathsf{canon}}}(w)$ is defined, by Lemma A.4 and consistency of $w \supset \Gamma \cup \{\neg\Box\varphi\}$ we have either $\Gamma_{-\Box} \cup \{\neg\varphi\}$ consistent or $\neg\Gamma_{-\Box} \cup \{\varphi\}$ consistent. But then (as the domain $\mathcal{L}_{\mathsf{M}}$ includes all maximal consistent sets, and all consistent sets can be extended to some $v \in \mathcal{L}_{\mathsf{M}}$) there must exist either some $w' \supset \Gamma_{-\Box} \cup \{\neg\varphi\}$ or $w'' \supset \neg\Gamma_{-\Box} \cup \{\varphi\}$. If the former, then by the inductive hypothesis, the consistency of w', and construction of $S_{\mathsf{min}\mathcal{M}_{\mathsf{canon}}}$ we have $w' \in (S_{\mathsf{min}\mathcal{M}_{\mathsf{canon}}}(w) \setminus [\![\varphi]\!])$, and thus $w \notin [\![\Box\varphi]\!]_{\mathcal{M}_{\mathsf{canon}}}$. If the latter, then by the inductive hypothesis, the consistency of w'', and construction of $S_{\mathsf{max}\mathcal{M}_{\mathsf{canon}}}$ we have $w'' \in ([\![\varphi]\!] \setminus S_{\mathsf{max}\mathcal{M}_{\mathsf{canon}}}(w))$, and thus $w \notin [\![\Box\varphi]\!]_{\mathcal{M}_{\mathsf{canon}}}$. □

Corollary A.6. $\vDash_{M_{\mathsf{canon}}} \varphi$ *iff* $\vdash_{\mathsf{Conv}} \varphi$

Proof. Suppose $\vdash_{\mathsf{Conv}} \varphi$. Then we have $\{\neg\varphi\}$ inconsistent, and thus by maximality of $w \in W_{\mathcal{M}_{\mathsf{canon}}}$ each such w must contain φ, which by the above means $\vDash_{M_{\mathsf{canon}}} \varphi$.

Suppose $\nvdash_{\mathsf{Conv}} \varphi$. Then we have $\{\neg\varphi\}$ consistent, meaning some $w \in W_{\mathcal{M}_{\mathsf{canon}}} = \mathcal{L}_{\mathsf{M}}$ must extend it, so by the consistency of $w \in W_{\mathcal{M}_{\mathsf{canon}}}$ this w must not contain φ, which by the above means $\nvDash_{M_{\mathsf{canon}}} \varphi$. □

By familiar reasoning, proving completeness for our logic now only requires, in light of Corollary A.6, that the frame $\langle W_{\mathcal{M}_{\mathsf{canon}}}, (S_{\mathsf{min}\mathcal{M}_{\mathsf{canon}}}, S_{\mathsf{max}\mathcal{M}_{\mathsf{canon}}})\rangle$ satisfies the requirements for a model frame given in the first section. (That it satisfies the requirements for its base valuation function $V_{\mathcal{M}_{\mathsf{canon}}}$ is immediate.)

Corollary A.7. *If* $\vDash_{\mathsf{Conv}} \varphi$, $\vdash_{\mathsf{Conv}} \varphi$

Proof. The conditions for a frame are just that, for all w with $S(w)$ defined, a) $w \in S_{\mathsf{min}}(w)$ and b) $S_{\mathsf{min}}(w) \subseteq S_{\mathsf{max}}(w)$.

$S_{\mathcal{M}_{\mathsf{canon}}}(w)$, by construction, is defined only for w with some $\Box\varphi \in w$. Let w be such a world. Now by T and maximality, for all $\Box\varphi \in w$, also $\varphi \in w$, and so by construction of $S_{\mathsf{min}\mathcal{M}_{\mathsf{canon}}}(w)$ we have $w \in S_{\mathsf{min}\mathcal{M}_{\mathsf{canon}}}(w)$.

Since $S_{\mathcal{M}_{\mathsf{canon}}}(w)$ is defined only when there exists such a $\Box\varphi \in w$, the set of $w' \in W_{\mathcal{M}_{\mathsf{canon}}}$ with *some* $\psi \in w'$ such that $\Box\psi \in w$ extends the set of those containing all such ψ, i.e. $S_{\mathsf{min}\mathcal{M}_{\mathsf{canon}}}(w) \subseteq S_{\mathsf{max}.\mathcal{M}_{\mathsf{canon}}}(w)$. □

Corollary A.8. $\vDash_{\mathsf{Conv}} \varphi$ *iff* $\vdash_{\mathsf{Conv}} \varphi$

A.4 The Logics Conv4, Conv5*, and Conv45*

The logics $\mathsf{Conv4}, \mathsf{Conv5}^*, \mathsf{Conv45}^* \subset \mathcal{L}$ are those generated by the inference rules and axiom schemas of Conv (*mutatis mutandis*) plus each of the following and both, respectively.

4 $\Box\varphi \rightarrow \Box\Box\varphi$

5* $(\Box\varphi \wedge \neg\Box\psi) \rightarrow \Box(\varphi \wedge \neg\Box\psi)$

The models of Conv4 and Conv5* are just like those of Conv, with the added stipulation that, for the first, $S_{\mathsf{min}}(w') \subseteq S_{\mathsf{min}}(w)$ and $S_{\mathsf{max}}(w') \supseteq S_{\mathsf{max}}(w)$ for all $w' \in S_{\mathsf{min}}(w)$ where $S(w)$ is defined; for the second, $S_{\mathsf{min}}(w') \supseteq S_{\mathsf{min}}(w)$ and $S_{\mathsf{max}}(w') \subseteq S_{\mathsf{max}}(w)$ for all $w' \in S_{\mathsf{min}}(w)$ where $S(w), S(w')$ are defined. Their canonical models are constructed just like that of Conv, save for maximal consistency in the respective logics replacing maximal $\vdash_{\mathsf{Conv}}$-consistency.

Soundness is routine.

Theorem A.9. *If* $\vdash_{\mathsf{Conv4}} \varphi$, $\vDash_{\mathsf{Conv4}} \varphi$

Proof. To extend our soundness proof for Conv, observe first that $w \in [\![\Box\Box\varphi]\!]$ iff for each $w' \in S_{\mathsf{min}}(w)$, $w' \in [\![\Box\varphi]\!]$, and for no $w'' \notin S_{\mathsf{max}}(w)$ is $w'' \in [\![\Box\varphi]\!]$. Now clearly, if $w \in [\![\Box\varphi]\!]$, by the at most restriction of $S_{\mathsf{min}}(w')$ and at most expansion of $S_{\mathsf{max}}(w')$ for $w' \in S_{\mathsf{min}}(w)$, $S_{\mathsf{min}}(w') \subseteq [\![\varphi]\!] \subseteq S_{\mathsf{max}}(w')$ (i.e. $w' \in [\![\Box\varphi]\!]$). Moreover, if $w \in [\![\Box\varphi]\!]$, $S_{\mathsf{min}}(w) \subseteq [\![\varphi]\!] \subseteq S_{\mathsf{max}}(w)$, so for no $w'' \notin S_{\mathsf{max}}(w)$ do we have $w'' \in [\![\varphi]\!]$, much less $w'' \in [\![\Box\varphi]\!]$. So we have $w \in [\![\Box\Box\varphi]\!]$ if $w \in [\![\Box\varphi]\!]$, i.e. we have 4 valid. □

Theorem A.10. *If* $\vdash_{\mathsf{Conv5^*}} \varphi$, $\vDash_{\mathsf{Conv5^*}} \varphi$

Proof. To see that $\mathsf{5}^*$ is valid, we will show that, whenever $w \in [\![\Box\varphi \wedge \neg\Box\psi]\!]$, $w \in [\![\Box(\varphi \wedge \neg\Box\psi)]\!]$. Consider that $w \in [\![\Box(\varphi \wedge \neg\Box\psi)]\!]$ iff for all $w' \in S_{\mathsf{min}}$, $w' \in [\![\varphi]\!]$ but not $w' \in [\![\Box\psi]\!]$ (i.e. not $S_{\mathsf{min}}(w') \subseteq [\![\psi]\!] \subseteq S_{\mathsf{max}}(w')$), and for all $w'' \notin S_{\mathsf{max}}(w)$, either $w'' \notin [\![\varphi]\!]$ or $w'' \in [\![\Box\psi]\!]$. But since for such $w' \in S_{\mathsf{min}}$, $S_{\mathsf{min}}(w') \supseteq S_{\mathsf{min}}(w)$ and $S_{\mathsf{max}}(w') \subseteq S_{\mathsf{max}}(w)$ (where $S(w')$ is defined), if we have $S_{\mathsf{min}}(w) \subseteq [\![\varphi]\!] \subseteq S_{\mathsf{max}}(w)$ but not $S_{\mathsf{min}}(w) \subseteq [\![\psi]\!] \subseteq S_{\mathsf{max}}(w)$ (i.e. if $w \in [\![\Box\varphi \wedge \neg\Box\psi]\!]$), we have the latter also for (defined) $S(w')$ (so that $w' \notin [\![\Box\psi]\!]$) and we have $w' \in [\![\varphi]\!]$ trivially, as $w' \in S_{\mathsf{min}}(w) \subseteq [\![\varphi]\!]$. (Where S is not defined, $w' \notin [\![\Box\psi]\!]$ trivially.) And, since if $w \in [\![\Box\varphi \wedge \neg\Box\psi]\!]$ we have $[\![\varphi]\!] \subseteq S_{\mathsf{max}}(w)$, for no such $w'' \notin S_{\mathsf{max}}$ do we have $w'' \in [\![\varphi]\!]$. □

Completeness is similarly routine.

Theorem A.11. *If* $\vDash_{\mathsf{Conv4}} \varphi$, $\vdash_{\mathsf{Conv4}} \varphi$

Proof. By our earlier result, it will suffice to show, where $\mathcal{M}'_{\mathsf{canon}}$ is the canonical model of $\mathsf{Conv4}$, that for $w \in W_{\mathcal{M}'_{\mathsf{canon}}}$ with $S_{\mathcal{M}'_{\mathsf{canon}}}(w)$ defined, $S_{\mathsf{min}\mathcal{M}'_{\mathsf{canon}}}(w') \subseteq S_{\mathsf{min}\mathcal{M}'_{\mathsf{canon}}}(w)$ and $S_{\mathsf{max}\mathcal{M}'_{\mathsf{canon}}}(w') \supseteq S_{\mathsf{max}\mathcal{M}'_{\mathsf{canon}}}(w)$ for all $w' \in S_{\mathsf{min}\mathcal{M}'_{\mathsf{canon}}}(w)$.

Suppose there is some $w' \in S_{\mathsf{min}\mathcal{M}'_{\mathsf{canon}}}(w)$ and $v \in S_{\mathsf{min}\mathcal{M}'_{\mathsf{canon}}}(w'), \notin S_{\mathsf{min}\mathcal{M}'_{\mathsf{canon}}}(w)$ (i.e. $S_{\mathsf{min}\mathcal{M}'_{\mathsf{canon}}}(w') \nsubseteq S_{\mathsf{min}\mathcal{M}'_{\mathsf{canon}}}(w)$). Then by construction, there is some $\Box\varphi \in w, \notin w'$. But by $\mathsf{4}$ this is impossible, since for any $\Box\varphi \in w$, $\Box\Box\varphi \in w$, and so $\Box\varphi \in w'$. (For this same reason, $S_{\mathcal{M}'_{\mathsf{canon}}}(w')$ is always defined when $S_{\mathcal{M}'_{\mathsf{canon}}}(w)$ is defined and $w' \in S_{\mathsf{min}\mathcal{M}'_{\mathsf{canon}}}(w)$.)

Suppose next there is some $w' \in S_{\mathsf{min}\mathcal{M}'_{\mathsf{canon}}}(w)$ and $v \in S_{\mathsf{max}\mathcal{M}'_{\mathsf{canon}}}(w), \notin S_{\mathsf{min}\mathcal{M}'_{\mathsf{canon}}}(w')$ (i.e. $S_{\mathsf{max}\mathcal{M}'_{\mathsf{canon}}}(w') \nsupseteq S_{\mathsf{max}\mathcal{M}'_{\mathsf{canon}}}(w)$). Then again by construction, there is some $\Box\varphi \in w, \notin w'$, which we have just seen to be impossible. □

Theorem A.12. *If* $\vDash_{\mathsf{Conv5^*}} \varphi$, $\vdash_{\mathsf{Conv5^*}} \varphi$

Proof. Again it will suffice to prove that for $w \in W_{\mathcal{M}'_{\mathsf{canon}}}$ with $S_{\mathcal{M}'_{\mathsf{canon}}}(w)$ defined, $S_{\mathsf{min}\mathcal{M}'_{\mathsf{canon}}}(w') \supseteq S_{\mathsf{min}\mathcal{M}'_{\mathsf{canon}}}(w)$ and $S_{\mathsf{max}\mathcal{M}'_{\mathsf{canon}}}(w') \subseteq S_{\mathsf{max}\mathcal{M}'_{\mathsf{canon}}}(w)$ for all $w' \in S_{\mathsf{min}\mathcal{M}'_{\mathsf{canon}}}(w)$ with some $\Box\psi \in w'$ (where $\mathcal{M}'_{\mathsf{canon}}$ is now the canonical model of $\mathsf{Conv5}^*$).

Suppose there is some $w' \in S_{\mathsf{min}\mathcal{M}'_{\mathsf{canon}}}(w)$ (with $S_{\mathcal{M}'_{\mathsf{canon}}}(w')$ defined) and $v \in S_{\mathsf{min}\mathcal{M}'_{\mathsf{canon}}}(w), \notin S_{\mathsf{min}\mathcal{M}'_{\mathsf{canon}}}(w')$ (i.e. $S_{\mathsf{min}\mathcal{M}'_{\mathsf{canon}}}(w') \nsupseteq S_{\mathsf{min}\mathcal{M}'_{\mathsf{canon}}}(w)$). Then by construction (and definedness of $S_{\mathcal{M}'_{\mathsf{canon}}}(w')$), there is some $\Box\varphi \in w', \notin w$. Now since $S_{\mathcal{M}'_{\mathsf{canon}}}(w)$ is defined, there exists $\Box\psi \in w$, with $\psi \neq \varphi$. We then by maximality have $\Box\psi \wedge \neg\Box\varphi \in w$. But then we have by $\mathsf{5}^*$ that $\Box(\psi \wedge \neg\Box\varphi) \in w$, and so by construction and maximality $\psi \wedge \neg\Box\varphi, \neg\Box\varphi \in w'$, which is by consistency of w' impossible.

Suppose next there is some $w' \in S_{\mathsf{min}\mathcal{M}'_{\mathsf{canon}}}(w)$ and $v \in S_{\mathsf{max}\mathcal{M}'_{\mathsf{canon}}}(w'), \notin S_{\mathsf{min}\mathcal{M}'_{\mathsf{canon}}}(w)$ (i.e. $S_{\mathsf{max}\mathcal{M}'_{\mathsf{canon}}}(w') \nsubseteq S_{\mathsf{max}\mathcal{M}'_{\mathsf{canon}}}(w)$). Then again by construction, there is some $\Box\varphi \in w', \notin w$, which we have just seen to be impossible. □

Corollary A.13. Conv4 *and* Conv5* *are independent.*

Proof. For a model of Conv5* invalidating 4, let $W = \{0, 1\}$, $S_{\mathsf{min}}(0) = S_{\mathsf{max}}(0) = W$, $S(1)$ be undefined. For a model of Conv4 invalidating 5*, let $W = \{0, 1\}$, $S_{\mathsf{min}}(0) = S_{\mathsf{max}}(0) = S_{\mathsf{max}}(1) = W$, $S_{\mathsf{min}}(1) = \{1\}$. For one satisfying both let $W = \{0\}$, $S_{\mathsf{min}}(0) = S_{\mathsf{max}}(0) = W$. □

Corollary A.14. Conv45* *is sound and complete on the class of frames where for any w with $S(w)$ defined, $S(v) = S(u)$ where $v, u \in S_{\mathsf{min}}(w)$.*

Proof. This is straightforward from the preceding proofs. □

Corollary A.15. *Among models in which, for all w with $S(w)$ defined, $S(w')$ is defined for all $w' \in S_{\mathsf{min}}(w)$, the* Conv5* *models are the* Conv45* *models.*

Proof. Take any such model of Conv5* and any w with $S(w)$ defined. For any $w' \in S_{\mathsf{min}}(w)$, $S_{\mathsf{min}}(w')$ (which we know to be defined) must contain w, as $w \in S_{\mathsf{min}}(w) \subseteq S_{\mathsf{min}}(w')$. So, by the defining property of Conv5* models, $S_{\mathsf{min}}(w) \subseteq S_{\mathsf{min}}(w')$ (as $w' \in S_{\mathsf{min}}(w)$) and $S_{\mathsf{min}}(w') \subseteq S_{\mathsf{min}}(w)$ (as $w \in S_{\mathsf{min}}(w')$), i.e. $S_{\mathsf{min}}(w) = S_{\mathsf{min}}(w')$; similarly, $S_{\mathsf{max}}(w) \supseteq S_{\mathsf{max}}(w')$ and $S_{\mathsf{max}}(w') \supseteq S_{\mathsf{max}}(w)$, i.e. $S_{\mathsf{max}}(w) = S_{\mathsf{min}}(w')$. Thus, for all $w' \in S_{\mathsf{min}}(w)$, $S(w) = S(w')$.

The reverse direction is trivial. □

We can now prove the correspondence between Conv45* and the principle attributed to Lewis in Section 3.1.

Fact A.16. *Take an arbitrary sandwich frame $\langle W, S\rangle$. Let $\mathcal{N}(p \subseteq W)$ be $\{w \in W : S_{\mathsf{min}}(w) \subseteq p \subseteq S_{\mathsf{max}}(w)\}$. Let O_w be the intersection of all $p \ni w$ such that for some $q \subseteq W$ either $p = \mathcal{N}(q)$ or $p = W \setminus \mathcal{N}(q)$.*

If for all w (with $S(w)$ defined), $S_{\mathsf{min}}(w) \subseteq O_w \subseteq S_{\mathsf{max}}(w)$, then for all v and $u \in S_{\mathsf{min}}(v)$, $S(v) = S(u)$, and vice versa.

Proof. **Left to right:** Clearly, given the antecedent, for all w (with $S(w)$ defined), $O_w = S_{\mathsf{min}}(w)$, as by definition $O_w \subseteq \mathcal{N}(S_{\mathsf{min}}(w)) \subseteq S_{\mathsf{min}}(w)$. It is now easy to see that for any $v \in O_w$, neither is there $u \in O_v, \notin O_w$ nor $u' \in O_w, \notin O_v$. For if the former, then $\mathcal{N}(S_{\mathsf{min}}(w)) \ni w, \not\ni v$, which is prohibited by construction of O as long as $v \in O_w = S_{\mathsf{min}}(w)$; and if the latter, either $\mathcal{N}(S_{\mathsf{min}}(v)) \ni v, \not\ni w$ or no $\mathcal{N}(p) \ni v$ at all, both of which are prohibited by construction of O as long as $v \in O_w = S_{\mathsf{min}}(w)$, since $w \notin \mathcal{N}(p)$ iff $w \in (W \setminus \mathcal{N}(p))$. Therefore $O_w = O_v$, and thus $S_{\mathsf{min}}(w) = S_{\mathsf{min}}(v)$.

Moreover, whenever $O_w = S_{\mathsf{min}}(w) = O_v = S_{\mathsf{min}}(v)$, $S_{\mathsf{max}}(w) = S_{\mathsf{max}}(v)$, as otherwise w, v will be distinguished by some p in that $w \in, v \notin \mathcal{N}(p)$ (or *vice versa*). But again by definition of O, this is impossible. So $S(w) = S(v)$.

Right to left: Suppose for all w (with $S(w)$ defined) and $v \in S_{\mathsf{min}}(w)$, $S(w) = S(v)$. Then $w \in \mathcal{N}(p)$ iff $v \in \mathcal{N}(p)$, so $O_w = O_v$. So by $v \in O_v$ for all $v \in S_{\mathsf{min}}(w)$, $S_{\mathsf{min}}(w) \subseteq O_w$. But $O_w \subseteq \mathcal{N}(S_{\mathsf{max}}(w)) \subseteq S_{\mathsf{max}}(w)$. So $S_{\mathsf{min}}(w) \subseteq O_w \subseteq S_{\mathsf{max}}(w)$. □

A.5 The Logic Conv+

Let the language $\mathcal{L}_{\boxdot}$ be $\mathcal{L}$ closed under the introduction of formulas $\ulcorner \boxdot\varphi \urcorner$. Let the logic $\mathsf{Conv+} \subset \mathcal{L}_{\boxdot}$ be that generated by the axiom schemas and inference rules of Conv (*mutatis mutandis*), together with the additional axiom schemas

$\mathsf{K}_{\boxdot}$ $\boxdot(\varphi \to \psi) \to (\boxdot\varphi \to \boxdot\psi)$

$\mathsf{T}_{\boxdot}$ $\boxdot\varphi \to \varphi$

PS $(\boxdot\varphi \wedge \Box\psi) \to \Box(\varphi \wedge \psi)$

and the additional inference rule

Nec if $\vdash_{\mathsf{Conv+}} \varphi$, $\vdash_{\mathsf{Conv+}} \boxdot\varphi$

A model $\mathcal{M}^{+}$ of Conv+ is just like a model of Conv, together with an extra parameter $R : W \to \mathcal{P}(W)$; R obeys the further constraints that for all $w \in W$, $S_{\mathsf{min}}(w) \subseteq R(w)$ (where defined), and $w \in R(w)$. As usual, $[\![\boxdot\varphi]\!]_{\mathcal{M}^{+}} = \{w \in W : R(w) \subseteq [\![\varphi]\!]_{\mathcal{M}^{+}}\}$. The canonical model $\mathcal{M}^{+}_{\mathsf{canon}}$ of Conv+ is constructed as for Conv, with the obvious extensions for $R_{\mathcal{M}^{+}_{\mathsf{canon}}}$ (particularly that $w' \in R_{\mathcal{M}^{+}_{\mathsf{canon}}}(w)$ iff $\varphi \in w'$ for all $\Box\varphi \in w$).

It is easy to see the logic is sound.

Theorem A.17. *If* $\vdash_{\mathsf{Conv+}} \varphi$, $\vDash_{\mathsf{Conv+}} \varphi$

Proof. The only distinctive case to show is for PS. By the stated constraint on R and S_{min}, whenever $w \in [\![\boxdot\varphi]\!], \in [\![\Box\psi]\!]$, $S_{\mathsf{min}}(w) \subseteq [\![\psi]\!] \subseteq S_{\mathsf{max}}(w)$ (by construction of $[\![\Box\psi]\!]$) and $S_{\mathsf{min}}(w) \subseteq [\![\varphi]\!]$ (by the above constraint), and thus $S_{\mathsf{min}}(w) \subseteq [\![\varphi]\!] \cap [\![\psi]\!] = [\![\varphi \wedge \psi]\!] \subseteq [\![\psi]\!] \subseteq S_{\mathsf{max}}(w)$, and so by construction of $[\![\Box(\varphi \wedge \psi)]\!]$, $w \in [\![\Box(\varphi \wedge \psi)]\!]$. Whence the conclusion follows trivially. □

The bulk of the completeness proof can rely on those given above.

Theorem A.18. *If* $\vDash_{\mathsf{Conv+}} \varphi$, $\vdash_{\mathsf{Conv+}} \varphi$

Proof. It is easy to check that the proofs for Theorem A.5 and Corollary A.6 can be extended to proofs of their analogues for Conv+. Thus all that is required is a proof that our canonical model $\mathcal{M}^{+}_{\mathsf{canon}}$ satisfies the criteria for a model.

For the constraints on R by itself the proofs are well known. We thus need only demonstrate that, when defined, $S_{\mathsf{min}\mathcal{M}^{+}_{\mathsf{canon}}}(w) \subseteq R_{\mathcal{M}^{+}_{\mathsf{canon}}}(w)$. Suppose there is some $w' \in S_{\mathsf{min}\mathcal{M}^{+}_{\mathsf{canon}}}(w), \notin R_{\mathcal{M}^{+}_{\mathsf{canon}}}(w)$ (where these are defined). Then by construction there are $\varphi \in w', \psi \notin w'$ (that is, by maximality, $\varphi, \neg\psi \in w'$) such that $\Box\varphi, \boxdot\psi \in w$. But then we do not have (by consistency of w') $\varphi \wedge \psi \in w'$, and so by construction of $S_{\mathsf{min}\mathcal{M}^{+}_{\mathsf{canon}}}$ we do not have $\Box(\varphi \wedge \psi) \in w$, which contradicts the hypothesis given PS (and maximality). □

Say that a model is *partitional* just in case, for all w, for all $v, v' \in R(w)$, $S_{\mathsf{min}}(v)$ and $S_{\mathsf{min}}(v')$ are (when defined) either identical or pairwise disjoint.

Theorem A.19. *No formulas define the partitional models.*

Proof. It will suffice to show that, for two models $\mathcal{M}, \mathcal{M}'$ with $\mathcal{M}$ partitional but not $\mathcal{M}'$, the φ valid on $\mathcal{M}$ are just those valid on $\mathcal{M}'$.

Let $W_{\mathcal{M}} = \{w_0, w_1\}$; $S_{\mathsf{min}\mathcal{M}}(w_0) = S_{\mathsf{max}\mathcal{M}}(w_0) = \{w_0\}$; $S_{\mathcal{M}}(w_1)$ undefined; and for all atomic p, $V_{\mathcal{M}}(p) = \{w_0\}$. Let $W_{\mathcal{M}'} = \{v_0, v_1, v_2\}$; $S_{\mathsf{min}\mathcal{M}'}(v_0) = S_{\mathsf{max}\mathcal{M}'}(v_0) = \{v_0, v_1\}$, $S_{\mathcal{M}'}(v_1) = (\{v_1\}, \{v_0, v_1\})$; $S_{\mathcal{M}'}(v_2)$ undefined; and for all p, $V_{\mathcal{M}'}(p) = \{v_0, v_1\}$. For both, R is the universal relation (the constant function to W). Clearly, $\mathcal{M}$ but not $\mathcal{M}'$ is partitional, so it remains to show $\vDash_{\mathcal{M}} \varphi$ iff $\vDash_{\mathcal{M}'} \varphi$.

First we prove by induction on complexity that $w_1 \in [\![\varphi]\!]$ iff $v_2 \in [\![\varphi]\!]$. For the atomic base case, this is given. For the Boolean connectives, this follows by classicality of the logic. Moreover, as $S(w_1)$ and $S(v_2)$ are undefined, the step for $\Box\varphi$ is likewise trivial.

Next, $w_0 \in [\![\varphi]\!]$ iff $v_1 \in [\![\varphi]\!]$. Again, the base step and Boolean inductive step are trivial. For $\Box\varphi$, if $w_0 \notin [\![\varphi]\!]$ and $v_1 \notin [\![\varphi]\!]$, it follows by T that $v_1 \notin [\![\Box\varphi]\!]$ and $w_0 \notin [\![\Box\varphi]\!]$; if $w_0 \in [\![\varphi]\!]$ and $v_1 \in [\![\varphi]\!]$, either $v_2 \in [\![\varphi]\!]$ and (thus, by the above) $w_1 \in [\![\varphi]\!]$ or not. If so, $v_1 \notin [\![\Box\varphi]\!]$ and $w_0 \notin [\![\Box\varphi]\!]$, since $v_2 \notin S_{\mathsf{max}}(v_1)$ and $w_1 \notin S_{\mathsf{max}}(w_0)$. If not, then $S_{\mathsf{min}}(v_1) \subseteq [\![\varphi]\!] \subseteq S_{\mathsf{max}}(v_1)$ (meaning $v_1 \in [\![\Box\varphi]\!]$) and $S_{\mathsf{min}}(w_0) = [\![\varphi]\!] = S_{\mathsf{max}}(w_0)$ (meaning $w_0 \in [\![\Box\varphi]\!]$).

Finally, $w_0 \in [\![\varphi]\!]$ iff $v_0 \in [\![\varphi]\!]$. Base and Boolean steps are trivial. For $\Box\varphi$, by the inductive hypothesis we have $w_0 \in [\![\varphi]\!]$ iff $v_0 \in [\![\varphi]\!]$. As (by T) the result is trivial if $v_0 \notin [\![\varphi]\!]$ and $w_0 \notin [\![\varphi]\!]$, suppose $v_0 \in [\![\varphi]\!]$ and $w_0 \in [\![\varphi]\!]$. Again, either $v_2 \in [\![\varphi]\!]$ (and thus $w_1 \in [\![\varphi]\!]$) or not. In the former case, the conclusion follows as before. Otherwise, by the preceding paragraph, we have $w_0 \in [\![\varphi]\!]$ and $v_1 \in [\![\varphi]\!]$, so $v_0 \in [\![\varphi]\!]$ and $v_1 \in [\![\varphi]\!]$, and thus, as $S_{\mathsf{min}}(v_0) = S_{\mathsf{max}}(v_0) = \{v_0, v_1\}$ and $v_2 \notin [\![\varphi]\!]$, $v_0 \in [\![\Box\varphi]\!]$. But (since we are assuming $w_1 \notin [\![\varphi]\!]$), as $S_{\mathsf{min}}(w_0) = S_{\mathsf{max}}(w_0) = \{w_0\} = [\![\varphi]\!]$, also $w_0 \in [\![\Box\varphi]\!]$.

So the $\varphi \in \mathcal{L}$ true at w_0 are just those true at v_0, v_1, and the $\varphi \in \mathcal{L}$ true at w_1 are just those true at v_2 (extending this to $\varphi \in \mathcal{L}_{\Box}$ is straightforward). The desired result–$\vDash_{\mathcal{M}} \varphi$ iff $\vDash_{\mathcal{M}'} \varphi$–is now immediate. □

We now prove Fact 3.3.

Proof. Suppose our model is partitional and that (3 - 5) and (C_i) are true (at $w \in W$), where i is the least number such that $\Box(p \wedge q_i)$ and $\neg\Box(p \wedge q_{i-1})$ are true. As (by (4)) $S_{\mathsf{min}}(w) \subseteq [\![q_{500{,}000}]\!]$ and $S_{\mathsf{min}}(w) \subseteq R(w)$, so (by (3) and PS) for all $w' \in S_{\mathsf{min}}(w)$, $S_{\mathsf{min}}(w')$ is defined and $S_{\mathsf{min}}(w') \subseteq [\![p]\!] \subseteq S_{\mathsf{max}}(w')$. By partitionality, $S_{\mathsf{min}}(w) = S_{\mathsf{min}}(w')$. We thus have for all $\Box(\varphi \wedge p)$ true at w (i.e. such that $S_{\mathsf{min}}(w) \subseteq [\![\varphi]\!] \cap [\![p]\!] \subseteq [\![p]\!] \subseteq S_{\mathsf{max}}(w)$, with S_{min} defined and fixed throughout $S_{\mathsf{min}}(w)$) that $[\![\Box(\varphi \wedge p)]\!] \supseteq S_{\mathsf{min}}(w)$, and thus $S_{\mathsf{min}}(w) \subseteq [\![\Box(p \wedge q_i)]\!] \subseteq [\![p]\!] \subseteq S_{\mathsf{max}}(w)$, and therefore we have $\Box\Box(p \wedge q_i)$ true (at w).

Since we also have $\neg\Box(p \wedge q_{i-1})$ true (at w), we by similar reasoning have $\Box(p \wedge q_{i-1})$ true at no $w' \in S_{\mathsf{min}}(w)$, and thus $S_{\mathsf{min}}(w) \subseteq [\![p]\!] \cap [\![\neg\Box(p \wedge q_{i-1})]\!] \subseteq [\![p]\!] \subseteq S_{\mathsf{max}}(w)$, and thus $\Box(p \wedge \neg\Box(p \wedge q_{i-1}))$ true (at w).

We by these and M therefore have $\Box(\Box(p \wedge q_i) \wedge p \wedge \neg\Box(p \wedge q_{i-1}))$, or equivalently $\Box(\Box(p \wedge q_i) \wedge \neg\Box(p \wedge q_{i-1}))$, true (at w), which contradicts (C_i). □

References

[1] D. Davidson, "The logical form of action sentences," in *The Logic of Decision and Action* (N. Rescher, ed.), pp. 81–95, University of Pittsburgh Press, 1967.

[2] J. Horty and E. Pacuit, "Action types in stit semantics," *Review of Symbolic Logic*, vol. 10, no. 4, pp. 617–637, 2017.

[3] R. Jeffrey, *The Logic of Decision*. New York: McGraw-Hill, 1965.

[4] J. Joyce, *The Foundations of Causal Decision Theory*. New York: Cambridge University Press, 1999.

[5] D. Lewis, "Causal decision theory," *Australasian Journal of Philosophy*, 1981.

[6] L. Savage, *The Foundations of Statistics*. New York: John Wiley and Sons, 1954.

[7] J. von Neumann and O. Morgenstern, *Theory of games and economic behavior*. Princeton University Press, 1947.

[8] D. Lewis, "Are we free to break the laws?," *Theoria*, vol. 47, no. 3, pp. 113–21, 1981.

[9] P. V. Inwagen, *An Essay on Free Will*. Oxford University Press, 1983.

[10] T. Nagel, *Mortal Questions*. Cambridge University Press, 1979.

[11] B. Williams, *Moral Luck*. Cambridge University Press, 1981.

[12] D. Pritchard, "Duncan pritchard, epistemic luck," *Theoria*, vol. 73, no. 2, pp. 173–178, 2007.

[13] R. J. Hartman, *In Defense of Moral Luck: Why Luck Often Affects Praiseworthiness and Blameworthiness*. Routledge, 2017.

[14] B. F. Chellas, "Time and modality in the logic of agency," *Studia Logica*, vol. 51, no. 3-4, pp. 485–517, 1992.

[15] A. Kenny, "Human abilities and dynamic modalities," in *Essays on Explanation and Understanding: Studies in the Foundations of Humanities and Social Science*, D. Reidel Publishing Company, 1976.

[16] N. Belnap and M. Perloff, "Seeing to it that: A canonical form for agentives," *Theoria*, vol. 54, no. 3, pp. 175–199, 1988.

[17] R. Stalnaker, *Inquiry*. Cambridge University Press, 1984.

[18] D. K. Lewis, *On the Plurality of Worlds*. Blackwell, 1986.

[19] R. Stalnaker, "On logics of knowledge and belief," *Philosophical Studies*, vol. 128, no. 1, pp. 169–199, 2006.

[20] J. F. Horty and N. Belnap, "The deliberative stit: A study of action, omission, ability, and obligation," *Journal of Philosophical Logic*, vol. 24, no. 6, pp. 583–644, 1995.

[21] H. G. Frankfurt, "Alternate possibilities and moral responsibility," *Journal of Philosophy*, vol. 66, no. 23, p. 829, 1969.

[22] J. M. Fischer and M. Ravizza, *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge University Press, 1998.

[23] K. Vihvelin, *Causes, Laws, and Free Will: Why Determinism Doesn't Matter*. Oup Usa, 2013.

[24] E. Pacuit, *Neighborhood Semantics for Modal Logic*. Short Textbooks in Logic, Springer, 2017.

[25] G. H. von Wright, *Norm and Action*. New York: Humanities, 1963.

[26] B. F. Chellas, *The Logical Form of Imperatives*. PhD thesis, Stanford University, 1969.

[27] K. Segerberg, "Getting started: Beginnings in the logic of action," *Studia Logica*, vol. 51, pp. 347–378, 1992.

[28] M. P. Nuel Belnap and M. Xu, *Facing the Future: Agents and Choices in Our Indeterminist World*. Oxford University Press on Demand, 2001.

[29] M. Xu, "On the basic logic of stit with a single agent," *Journal of Symbolic Logic*, vol. 60, no. 2, pp. 459–483, 1995.

[30] M. Xu, *An Investigation in the Logics of Seeing-to-It-That*. PhD thesis, University of Pittsburgh, 1996.

[31] M. Xu, "Axioms for deliberative stit," *Journal of Philosophical Logic*, vol. 27, no. 5, pp. 505–552, 1998.

[32] A. H. Philippe Balbiani and N. Troquard, "Alternative axiomatics and complexity for deliberative stit theories," *Journal of Philosophical Logic*, vol. 37, no. 4, pp. 387–406, 2008.

[33] E. Lorini and F. Schwarzentruber, "A logic for reasoning about counterfactual emotions," *Artificial Intelligence*, vol. 37, no. 3-4, pp. 814–847, 2011.

[34] A. Gibbard and W. Harper, *Foundations and Applications of Decision Theory*, ch. Counterfactuals and Two Kinds of Expected Utility. No. 13a in University of Western Ontario Series in Philosophy of Science, Dordrecht: D. Reidel, 1978.

[35] J. L. Pollock, "Rational choice and action omnipotence," *Philosophical Review*, vol. 111, no. 1, pp. 1–23, 2002.

[36] B. Hedden, "Options and the subjective ought," *Philosophical Studies*, vol. 158, no. 2, pp. 343–360, 2012.

[37] J. Koon, "Options must be external," *Philosophical Studies*, vol. 177, no. 5, pp. 1175–1189, 2020.

[38] W. Schwarz, "Objects of choice," *Mind*, 2023.

[39] E. Lorini, D. Longin, and E. Mayor, "A logical analysis of responsibility attribution: emotions, individuals and collectives," *Journal of Logic and Computation*, vol. 24, no. 6, pp. 1313–1339, 2011.

[40] J. Boudou and E. Lorini, "Concurrent game structures for temporal stit logic," in *Proceedings of the Seventeenth International Conference on Autonomous Agents and MultiAgent Systems*, ACM Press, 2018.

[41] T. Williamson, *Knowledge and Its Limits*. Cambridge University Press, 2001.

[42] D. G. Fara, "Gap principles, penumbral consequence, and infinitely higher-order vagueness," in *New Essays on the Semantics of Paradox* (J. C. Beall, ed.), Oxford University Press, 2003.

[43] A. Bacon, *Vagueness and thought*. Manuscript, forthcoming.

[44] A. Ford, "The province of human agency," *Noûs*, vol. 52, no. 3, pp. 697–720, 2018.

[45] H. Lederman and B. Holguín, "Trying without fail." Unpublished manuscript, 2023.

[46] M. J. Cresswell and G. E. Hughes, *A New Introduction to Modal Logic*. Routledge, 1996.