Introspective Unawareness and Observable Choice*

Evan Piermont†

November 29, 2016

Abstract

This paper considers a framework in which the decision maker's (DM) knowledge and awareness are explicitly modeled, as is her ability to reason about her own (un)awareness. The DM has a ranking over consumption alternatives that is informed by her epistemic state (i.e., what she knows and what she is aware of). The main result is a characterization, via observable choice, of *introspective unawareness* –a DM who is both unaware of some information and aware she is unaware. In static environments, or when the DM is blind to her own ignorance, the presence of unawareness does not produce any observable choice patterns. However, under dynamic introspective unawareness, the DM will be unwilling to commit to making future choices, even when given the flexibility to write a contingent plan that executes a choice conditional on the realization of uncertain events. This is a behavior that cannot be explained by uncertainty alone (i.e., without appealing to unawareness). I show, in a simple strategic environment, this behavior can lead to the Pareto optimality of incomplete contracts.

JEL Classification: D01, D810, D830.
Keywords: Unawareness, Uncertainty, Epistemic Preferences, Dynamic Choice.

*I thank Eddie Dekel, Simon Grant, Edi Karni, Peter Klibanoff, Bart Lipman, Daniel Martin, Sofia Moroni, Tymofiy Mylovanov, Teddy Seidenfeld, Marciano Siniscalchi, Asher Wolinsky, the conference participants at RUD 2016 and LOFT 2016, and of course, my advisors Luca Rigotti and Roee Teper for their helpful and insightful comments and suggestions.
†Contact Information: Email: ehp5@pitt.edu. Telephone: (412) 648 7039. University of Pittsburgh, Department of Economics.

1 Introduction

There is a marked difference between being *unaware* of one's preferences and not *knowing* (also referred to as being *uncertain* about) one's preferences. While certainty and uncertainty (about some piece of information, φ) together constitute awareness of φ, unawareness describes total ignorance –a complete lack of perception of, or ability to reason directly about φ.[1] This paper explores the behavioral (i.e., observable) identification between unawareness and uncertainty and contemplates the type of data required to make such a separation. Due to the consideration of observability, the primary interest is in the decision maker's preference (hypothetically embodied by choice data), and how patterns in preference change in response to the structure of awareness. I will show that unawareness produces distinct patterns, and so, attempting to model unawareness with uncertainty, regardless of how complex, will fail. As an example of when such issues may arise and how they might alter predictions, I consider a simple contracting environment. When unawareness is taken into account, players can have an incentive to conceal mutually beneficial information, leading to the optimality of incomplete contracts.

To highlight the distinction between uncertainty and unawareness, consider a decision maker (DM) who will buy a new smartphone in six months. She will have three options at the time of purchase: x, y, and z. She might not know which phone she would most like to purchase six months from now. This uncertainty could arise because she does not know the technical specifications of the phones, their price, etc., and her true preference depends on the realization of these variables. Contrast this to the case where the DM has never heard of phone z. Here, she is unaware of z, and so naturally, of any preferences there regarding. The key aspect, if a decision maker is unaware of a piece of information (the existence of phone z), she is unable to make any choice based directly on this information.

More subtle, but just as fundamental, is our acknowledgement of our own unawareness. Indeed, most people would readily admit to the possibility that they cannot conceive of all future technologies or trends, or exhaustively list the set of possible occurrences for the upcoming week. This recognition of unawareness is important because it suggests the things a DM is unaware of may play an indirect role in her decision making, even if they cannot be directly acted upon. Central to the analysis, then, is the DM who is (1) unaware, (2) aware she is unaware, and (3) unaware about what she is unaware. A DM in such an epistemic state is referred to as *introspectively unaware*. By contrast, a DM who does not satisfy condition (2) would be referred to as *naively unaware*. In the presence of introspective unawareness, the DM above might envision a world in which she prefers *something* other than x and y. Of course, she cannot know this *something* is z, since that would require she is aware of it.

Notice, under either uncertainty or introspective unawareness, the DM has a natural inclination to delay making her choice (i.e., if she cannot start using the phone for six months, she might as well wait until then to choose). However, the motivation for delay is different. Under uncertainty, she would like to wait so as to make a decision based on the realization of the relevant variables (the technical specs, price, etc.). Under (introspective) unawareness, she would like to wait in case she becomes aware of something better than whatever she would have chosen today. Notice also, if the DM is unaware she is unaware, she has no reason to delay; she does not consider the possibility she becomes aware of new information.

Now, imagine the DM is going to get her brother to purchase the phone on her behalf, and has to instruct him *today* about which phone to purchase in six months. If the DM is either uncertain or introspectively

[1]This was first noted by Modica and Rustichini (1994).

unaware of her preference, it will not be optimal for her to specify any single phone. In the case of uncertainty, however, she could leave detailed instructions that would carry out her optimal choice: in the event the technical specs are (s_x, s_y, s_z), and the prices are $(\$x, \$y, \$z)$, purchase phone x, in the event ... etc. A commitment to consume (in the future) a particular alternative given the state of affairs is referred to as a *contingent plan*. Since her optimal decision depends on the realization of some variables,[2] it is enough for her to specify a contingent plan that depends on these variables. Contrast this to the case in which the DM is introspectively unaware. No plan, at least no articulable one,[3] can carry out her optimal decision (this vignette is captured formally in Example 3). This is because she would need to contract on events that are described by information she is unaware of—to include such information in a contingent plan would require she is aware of it.

The main result of this paper shows the behavioral condition for introspective unawareness is a strict preference for delaying choices rather than committing to a contingent plan, even when every possible articulable plan is offered. In particular:

- ($\star$1) *When the decision maker is fully aware, she is always willing to commit to some articulable contingent plan.* Intuitively, the DM's language is rich enough that she can contract on the resolution of any relevant uncertainty, effectively imitating whatever her dynamic behavior would have been.
- ($\star$2) *Without full awareness, the DM might find every articulable contingent plan unacceptable.* A strict preference for delay (relative to all articulable contingent plans) is possible only if the DM is not fully aware –this behavior is an indication of unawareness. Intuitively, if DM believes she may become aware of new alternatives, she understands her future self may have options better than any she could currently articulate in a contingent plan.
- ($\star$3) *If the DM does finds every articulable contingent plan unacceptable, she must be introspectively unaware.* Intuitively, the DM must be aware enough to come to the conclusion that waiting might afford new possibilities. When she is ignorant of her own unawareness, she cannot consider this possibility.

So a preference for delay that cannot be appeased by the appeal to contingent planning is the behavioral indication –in an exact sense– of introspective unawareness. The intuition is exactly as in the above example: the DM's language is not rich enough to specify the optimal contingent plan (unawareness), but is rich enough that she knows this fact (awareness of unawareness).

Directly incorporating unawareness into a decision theoretic model introduces subtleties that need to be dealt with judiciously. First, one must take care to ensure the process of eliciting preferences from a DM does not affect her preferences. While asking a DM to rank risky prospects should not affect her risk preference, asking her to contemplate objects of which she was formerly unaware certainty would affect her awareness (for a longer discussion on this topic, see Section 1.1). Second, the *type* of unawareness considered (i.e., naive or introspective, object-based or state-based, etc.) must be rich enough to produce observable patterns, even when keeping in mind the pervious concern. Finally, Modica and Rustichini (1994); Dekel *et al.* (1998) show that within the context of state space models, simply assuming that the DM is unaware of

[2]Variables of which she is aware, of course.

[3]A contingent plan –a function from events to outcomes– is *articulable* if the DM is aware of its constituent parts, its domain and image.

certain states (while retaining desirable properties of knowledge) is insufficient: the DM will either be fully aware or fully unaware.

To overcome these obstacles, I develop a logical framework that directly incorporates the DM's preference, as well as her knowledge and awareness thereof. This ensures that my notion of awareness is well founded and rigorous, and allows me to directly verify that the elicitation method does not require the DM to contemplate objects she herself could not have articulated. Finally, this construction indicates that contingent plans are precisely the type of data required to observe unawareness. Giving the DM either less flexibility (for example, deterministic choice of a single element) or more flexibility (for example, choice over incomplete contingent plans) fails to separate unawareness and uncertainty. Although, in the end, choices will be observable, this more general framework will provide the tools to analyze the *epistemic conditions* (i.e., knowledge and awareness) that generate the observable patterns in choice data.

I begin with an epistemic modal logic, based on set of logical statements that include a formal description of the DM's preference, and adapted from Fagin and Halpern (1988), Halpern and Rêgo (2009) and Board and Chung (2011). Each *state of the world* is defined by a set of statements which are true: statements that include how the DM ranks objects (for example, "x is preferred to y"), what the DM *implicitly knows* (for example "the DM implicitly knows 'x is preferred to y' ") and what the DM *is aware of* (for example "the DM is aware of the statement 'x is preferred to y' "). The intersection of implicit knowledge and awareness is *explicit knowledge*. Implicit knowledge can be thought of as idealized knowledge –what the DM would know if she was fully aware and logically omnipotent. In contrast, explicit knowledge can be thought of as working knowledge, subject to cognitive limitations. Then, on top of this logic, I build a decision theory. This allows me to speak of a preference as true (irrespective of the DM's epistemic state), implicitly known (if the DM's preference is the logical consequence of things she knows), and explicitly known (if the DM implicitly knows her preference and is aware of it). If the DM is fully aware, her implicitly and explicitly known preferences coincide. The characterization of unawareness arises from the contrast between the structure of these different preferences.

Failing to account for the effect of unawareness can distort predictions. Specifically, ex-ante solution concepts in dynamic environments.[4] However, including the formal machinery needed to deal appropriately with these concerns can impose a large cost in simple models, and so, should not be done needlessly. This paper, therefore, provides a test for the presence for introspective unawareness, so that economists might better understand in which context unawareness is present and where it can be safely ignored.

To exemplify how introspective unawareness can alter behavior in economic settings, I examine a highly stylized strategic environment in which a (fully aware) principal offers a take-it-or-leave-it contract to a (introspectively unaware) agent. I show, when constrained to offer complete contracts, the principal might have an incentive to conceal Pareto improving information. Intuitively, this is because the introspectively unaware agent is unable to properly anticipate the value of the actions she is unaware of but might become aware of (i.e., her value to not committing). By unveiling new actions by including them in his contract, the principal alters the agent's epistemic state –her perceived value of delay– potentially increasing her aversion to commitment. When unconstrained, the principal can overcome this by leaving the contract incomplete. I show, because this behavior is motivated by the agent's fuzzy perceived value to delay, it arises only in the

[4]For more concrete examples, the divergence from standard predictions in applications to game theory (Feinberg, 2012), contract theory (Filiz-Ozbay, 2012), information economics (Heifetz *et al.*, 2013), and mechanism design (Auster, 2013).

presence of introspective unawareness.

In addition, the epistemic preference framework relates to the decision theoretic literature on subjective state spaces, which began with Kreps (1979). I show that the Krepsian paradigm, epitomized by DM who ranks *menus* of objects, and whose ranking respects set inclusion (so that larger menus are preferred to their subsets), can be faithfully captured by a special case of the models studied in this paper. Such behavior arises without the need to appeal to unawareness. Intuitively, the flexibility that arises by appealing to menus is, just as with contingent plans, bounded by the DM's articulation. Hence, introspective unawareness produces the same unwillingness to commit in this domain.

In particular, this characterization is of interest in relation to models of subjective learning. To identify what the DM believes she might learn, axiomatizations generally include the requirement that any dynamic choice behavior is indifferent to some contingent plan –in essence, assuming the existence of acceptable, and articulable, plans.[5] As such, the results of this paper mandate that a theory of subjective learning under unawareness cannot be built on the same machinery. Put differently, current models of subjective learning necessarily reduce all learning to resolution of uncertainty rather than from the arrival unanticipated information.

1.1 Observability and Unawareness

Most decision theoretic models begin with the declaration of some set, X, over which the DM's preferences are described by a binary relation, $\succcurlyeq \subseteq X \times X$. Under the revealed preference interpretation, the modeler precludes from the outset the possibility of alternatives of which the decision maker is unaware. Indeed, if the modeler were to ask a person on the street, or a subject in the lab, to choose between phone x and phone y, it is unreasonable to believe, at the time of her answer, she is unaware of either x or y. The very act of asking forces the DM's hand. When moving to the realm of contingent plans –functions from events to consumption– this problem not only persists, but is compounded. Now, the modeler is precluding unawareness both of the consumption space and the set of contingencies. Finally, when trying to identify introspection, the problem becomes even more precarious. Any question regarding even the *existence* of unforeseen objects has the potential to change the DM's epistemic state.

The decision theoretic literature has posited many different behavioral markers for unawareness: incomplete preferences, preference for flexibility Kreps (1979), reverse Bayesianism Karni and Vierø (2016), unmeasurable states Kochov (2015); Minardi and Savochkin (2015); Grant and Quiggin (2014), etc.[6] However, these papers, formulated from a revealed preference approach, must make restrictive (ex-ante) assumptions on the DM's epistemic state. For example, Kochov (2015) and Minardi and Savochkin (2015) assume the modeler has a strictly more complete view of the world (is aware of more) than the agents within the model. This means it is impossible to detect unawareness in agents who are more aware than the modeler, severely narrowing the contexts in which these models can be applied. On the other hand, Karni and Vierø (2016) and Grant and Quiggin (2014) allow the DM's unawareness to be largely unrestricted. They do, however, assume that the DM is introspectively unaware, requiring her to rank objects explicitly containing *surprising* outcomes.

[5]The literature on learning has been principally interested in the case where the DM entertains a subjective state space, and identification regards the set of events in this state space the DM believes she might learn. Dillenberger *et al.* (2014, 2015); Piermont *et al.* (2015) have considered constructions that exactly correspond to contingent plans in this paper. While Ergin and Sarver (2010) and Riella (2013) do not directly construct such plans, the interpretation of both papers concerns a DM who constructs a contingent plan *after* observing the menu they receive.

[6]For a detailed account of these papers' and how they relate to this one, see the literature review in Section 8.

This paper asks the precedent question as to how such epistemic states might be identified. Such issues are dealt with by relaxing what is meant by the revealed preference approach. Instead of providing the decision maker with a set and asking her to indicate her preferences thereover, the modeler asks the DM to provide both the set of alternatives and the list of preference restrictions between them. That is to say, the DM provides a set of statements such as "x is preferred to y," where x and y are object of which she herself conceives. While this is clearly inadequate for some practical purposes, the complexity of the DM's task is no greater than in the standard model, and it does not beg the question of awareness.

A crucial aspect to the identification of unawareness contained in this paper is that it never requires the DM to contemplate objects she herself could not have conceived. It suffices for the modeler to consider the DM's preference over the set of objects she herself reported. The main contribution of this paper is, therefore, the assertion of a framework that characterizes introspective unawareness from choices regarding only information of which the DM is aware.

1.2 Organization

The structure of the paper is as follows. Section 2 provides a discussion and overview of the results, skirting the technical details. Section 3 introduces the logical underpinnings of the decision theory and expounds on the choice patterns based on implicitly known preference. Section 4 formally introduces awareness structures and explicit knowledge. The main results are contained in Section 5, which introduces contingent plans and the notions of implicit and explicit acceptability. Section 6 show the connection to subjective state space models and a preference for flexibility. Section 7 explores a simple strategic contracting game. A survey of the relevant literature can be found in Section 8. Additional results and proofs omitted from the text are contained in the appendix.

2 An Overview of Results

This section contains a discussion of the methods and results, without diving into technical details, to serve both as a motivation and a roadmap for the remainder of the paper. First, I show a purely decision theoretic account of static unawareness produces no observably different predictions from the standard model. This is an incentive to directly model the knowledge and awareness of the DM and allow for introspection, which will ultimately constitute a necessary condition for the identification of unawareness.

2.1 A (Far Too) Simple Model of Decision Making Under Unawareness.

Outside of the limits of real cognition, let X represent the objective set of all possible alternatives in all possible worlds for all possible decision makers. Assume the modeler asks the DM to report her personal set of preference restrictions (a set of order pairs of objects the DM conceived of). Then, the reported preference restrictions are a subset of $X \times X$. If the DM is unaware of some of the elements of X then she cannot include them in her report.

So, let $\succcurlyeq \subseteq X \times X$ be the decision maker's personal set of preference restrictions.[7] The highest standard of rationality would dictate that $\succcurlyeq$ is a transitive, complete, and reflexive relation. If we allow for unawareness, however, what properties should $\succcurlyeq$ possess? Transitivity has no reason to be discarded; if the decision maker's true preference is transitive and she is aware $x \succcurlyeq y$ and $y \succcurlyeq z$, then she has all the constituent parts to deduce that $x \succcurlyeq z$.

[7]For any relation $\succcurlyeq$, let $\sim$ an $\succ$ denote the symmetric and antisymmetric components.

[Trv] *For all* $x, y, z \in X$*,* $x \succcurlyeq y$ *and* $y \succcurlyeq z$ *then* $x \succcurlyeq z$*.*

Completeness, however, is clearly too strong; so too reflexivity. If the DM is unaware of x, she cannot report $x \succcurlyeq x$, even if it is true. However, if the DM is aware that $x \succcurlyeq y$ and that her preferences are complete, she can deduce that $x \succcurlyeq x$ and $y \succcurlyeq y$. Likewise if she is aware of both $x \succcurlyeq x$ and $y \succcurlyeq y$, she can deduce that there must be some preference between x and y. This is the idea behind local completeness:

[LCmp] *For all* $x, y \in X$*, '*$x \succcurlyeq y$ *or* $y \succcurlyeq x$*' if and only if '*$x \succcurlyeq x$ *and* $y \succcurlyeq y$*'.*

Trv and LCmp provide enough structure to provide a basic, but mostly pointless, result. There exists some $\mathcal{A} \subseteq X$ of objects the DM is *aware of*, and a utility function $U : \mathcal{A} \to \mathbb{R}$ that represents $\succcurlyeq$.[8] If $x \notin \mathcal{A}$, then $U(x)$ is undefined and there is no $y \in X$ such that $x \succcurlyeq y$ or $y \succcurlyeq x$: the DM is unaware of x. In addition, LCmp can be weakened to local reflexivity, to allow the decision maker to have incomplete preferences, even over objects she is aware of:

[LRfx] *For all* $x, y \in X$*, if* $x \succcurlyeq y$ *then '*$x \succcurlyeq x$ *and* $y \succcurlyeq y$*'.*

Trv and LRfx provide the analog to discarding completeness (while retaining reflexivity) in the standard model (Ok, 2002). There exists some $\mathcal{A} \subseteq X$ and a set of utility functions $U_k : \mathcal{A} \to \mathbb{R}$ such that $x \succcurlyeq y$ if and only if $U_k(x) \geqslant U_k(y)$ for all k.

These results show, in static contexts, unawareness has only trivial behavioral implications. Is there any economic content to considering X rather than $\mathcal{A}$? The behavior is *locally* identical –the patters of observable choice will be the same but for the inclusion or exclusion of particular elements. As alluded to in the pervious section, to see the full effects of unawareness, dynamic environments are needed. Of course, even there, if the DM is unaware she is unaware then she ignores the consequences of her limited understanding entirely; she acts as in the static case outlined above. In light of this, my focus is the introspectively unaware decision maker in a dynamic setting.

2.2 Epistemic Preferences

The cornerstone of this paper is the construction of *epistemic preferences* –the set of preference restrictions reported by the DM under varying epistemic constraints of knowledge and awareness. The construction begins with a set of logical formulae regarding a consumption set, X, and $T + 1$ time periods, indexed by $t = 0 \ldots T$.[9] These formulae include statements regarding the DM's preference over X: for example "$x \succcurlyeq y$" states that the DM prefers x to y. In addition, these formulae include statements about which formulae the DM knows and which she is aware of. That is, for each formula φ, there are corresponding formulae "$K_t\varphi$" (the DM implicitly knows at time t that the statement φ is true), and "$A_t\varphi$" (the DM is aware at time t of the statement φ).

I then consider models of decision making that provide a semantic interpretation to formulae. Models are described by a set of states of the world, S. Each state, $s \in S$, is characterized by a set of statements (i.e., logical formulae) that are true in that state, including statements that assure the DM has well defined preference relation, $\succcurlyeq_s$, over X that is complete and transitive (where $x \succcurlyeq_s y$ if and only if "$x \succcurlyeq y$" is a true statement in state s).

[8]Later, it is assumed that the set of consumable objects is denumerable. Here, however, if X is uncountable, the relevant continuity assumptions must also be made.

[9]The reuse of notation is deliberate. If the decision theoretic model presented above is founded on a larger epistemic model, then X (and $\mathcal{A}$) defined there will (almost) correspond to the same objects as denoted in the epistemic model.

The truth of $K_t\varphi$ depends on which states the DM believes are possible. In each state of the world, the DM entertains a set of states she believes contains the true state. This is represented by an *accessibility relation*, $R_t(s) \subseteq S$. $R_t(s)$ is the set of states the decision maker considers possible, at time t, when the true state is s. If $R_t(s)$ is a singleton, the DM *knows* the state. The DM knows φ if φ is true in every states she considers at time t.

As such, the DM's uncertainty regarding the state induces uncertainty regarding her true preference. The DM *implicitly knows* "x is preferred to y" at time t, denoted by $\succcurlyeq_{K_t,s}$, if it is true in every state she considers possible at time t. That is, $x \succcurlyeq_{K_t,s} y$ if and only if $x \succcurlyeq_{s'} y$ for all $s' \in R_t(s)$ if and only if $K_t(x \succcurlyeq y)$ is true in state s.

I show (in Remark 3.3 and Example 1) these implicit preferences are reflexive and transitive, but in general incomplete, and provide a representation. Being a preorder, it is representable by a multi-utility model (Ok, 2002; Evren and Ok, 2011). In particular, the representation is by the canonical set of utilities: the set of functionals that represent $\succcurlyeq_{s'}$ for $s' \in R_t(s)$. This result provides justification of the criticism that completeness is not, in general, normatively desirable Aumann (1962); Eliaz and Ok (2006). Although the idealized preferences are complete, a perfectly rational DM may still be unable to compare among some alternatives (i.e., exhibit indecisiveness) as a consequence of her uncertainty. Interpreting observable choice as the outlet of implicit knowledge attaches a clear epistemic interpretation to multi-utility models; incompleteness arises as the byproduct of the DM's lack of certainty regarding her true preference (which may be complete).

2.3 Awareness and Preference

I accommodate awareness via *awareness structures*. In each state of the world, s, and each time, t, the decision maker is aware only of some subset of all possible statements. The set is $\mathcal{A}_t(s)$. Then, in analogy to implicitly known preferences I define *explicitly* known preferences: $x \succcurlyeq_{E_t,s} y$ is and only if $x \succcurlyeq_{K_t,s} y$ and $(x \succcurlyeq y) \in \mathcal{A}_t(s)$. I show, when the awareness sets are suitably behaved, explicitly known preferences are transitive and locally reflexive, but not necessarily locally complete (Proposition 4.3), the later holding if the true presences are implicitly known (Remark 6.2). These results can be seen as placing bounds on the type of environment needed to uncover unawareness: in static models, unawareness provides no observably distinguishable patterns in choice data. In addition, it provides a good robustness check that the conception of static unawareness from a decision theoretic vantage point (embodied by the discussion in Section 2.1) is in fact being captured by the purely logical awareness structures.

Both dynamics and introspection are required to cleanly separate uncertainty from unawareness. It is the final requirement, that the DM can reason about her own awareness, which renders the formal modeling of knowledge and awareness, strictly necessary. To see how introspection can be captured, consider the following. $A_t(x \succcurlyeq x)$ is the statement that the DM is aware that $x \succcurlyeq x$ at time t. Thus if such a statement is true is state s, then in that state, the DM is aware of the existence of x at time t. As such,

$$\neg A_0(a \succcurlyeq a) \wedge A_1(a \succcurlyeq a)$$

is the statement that the DM is unaware of x at time 0 but will become aware of x by time 1. Notice that to assume that DM was aware of such a statement (at time 0) would be nonsensical, as it would assume she is both unaware of x and aware she is unaware of it. Nonetheless, such inconstancies are circumvented by appealing to quantification. Letting a be a variable that ranges over all objects (even ones the DM is unaware of), the statement

$$A_0(\exists a(\neg A_0(a \succcurlyeq a) \wedge A_1(a \succcurlyeq a))), \tag{2.1}$$

captures introspection. So, if (2.1) is true in state s, then in that state the DM is aware of the statement there exists some consumable object she currently is unaware of, but will be aware of at time 1. Observe, in the examples above, x and y referred to specific elements, while here a is a variable bound by the existential quantifier. Hence (2.1) posits the existence of an object that will come into the DM's awareness without specifying what the object is. My definition of *introspective unawareness* is a slightly stronger condition than (2.1), requiring not only that the DM is aware of such statements but also knows it is possible such statements are true.

2.4 Contingent Planning

Section 5, containing the main characterization result, considers a domain of contingent plans. A *contingent plan*, c_t, is a function from a set of statements, Λ, into consumption. That is, if a particular statement, $\varphi \in \Lambda$ is true, the decision maker commits to consume $c_t(\varphi) \in X$ at time t. Λ is chosen such that in every state exactly one element is true.

I do not consider a ranking over contingent plans directly, but rather, impose an epistemic restriction on when a DM would be willing to commit to a contingent plan (to commit to consuming, in time t, the alternative corresponding to the true state). The DM is willing to commit to a contingent plan, c_t, if its outcomes are no worse than her optimal choice had she waited until time t. This is characterized formally by the notion that a contingent plan c_t is *acceptable* (either implicitly or explicitly). If the DM implicitly knows at time 0 she will not implicitly know at time t any object that is strictly better than $c_t(\varphi)$ whenever φ is true, then c_t is *implicitly acceptable*. In other words, the DM knows at the time of accepting the plan that in the event x is the specified consumption, she will not know any better alternative than x, even if she waited until time t. If the same condition holds when implicit knowledge conditions are replaced with explicit ones, c_t is *explicitly acceptable*. A contingent plan is *articulable* if the DM is aware of its constituent parts (image and domain).

Using these definitions, the characterization of introspective unawareness is the conjunction of several results. First, I provide mild conditions such that there always exists an implicitly acceptable contingent plan (Theorem 5.3, corresponding to (★1)). With full awareness, the DM is willing to commit: her language is rich enough that she can, by using the appropriate contingent plan, mimic whatever she would have done had she waited. In contrast, I provide a well-behaved and intuitive model (Example 3) in which no articulable contingent plan is explicitly acceptable (corresponding to (★2)). This indicates that unawareness can produce distinct patterns in choice behavior. Theorems 5.5 and 5.6 place bounds on the level of awareness that will allow such behavior. Theorem 5.5 states, if the DM's language is rich enough to articulate an acceptable contingent plan, then she finds this contingent plan explicitly acceptable. As such, trivial forms of unawareness (i.e., being unaware of preference-irrelevant information) does not generate a preference for delay. Finally, Theorem 5.6 shows, if the DM is unwilling to commit to any articulable contingent plan, her unawareness must be introspective (corresponding to (★3)). The DM must explicitly know it is possible she is unaware or else she cannot anticipate the discovery of options better than what she could articulate today. Importantly, the modeler only ever requires the DM to report a preference over articulable contingent plans –the DM never ranks objects she does not understand.

3 Logical Foundations: Preferential Logic

This section outlines the formal construction of the logic used in this paper. First Section 3.1 provides the syntax for well defined formulae. That is, a purely mechanical account of which strings of characters will be *well defined*. Then Section 3.2 endows well defined formulae with meaning by providing a semantic interpretation. This interpretation is the standard *possible worlds semantics* adapted to consider preferential statements. Finally, section 3.3, considers an axiomatization (a method of deriving new true statements from old ones) corresponding to the semantic models.

3.1 A Syntactic Language: $\mathcal{L}(\mathcal{X})$

Preferential choices will be described directly by an epistemic logic. To this end, for each $n \geqslant 1$, define a countable set of n place predicates denoted by $\alpha, \beta, \gamma, \ldots$. Assume the existence of a countably infinite set of variables denoted by $\mathcal{X} = a, b, c \ldots$. Then, any n place predicate followed by n variables is a well formed *atomic* formula. That is, if α is a 2 place predicate, then αab is a well formed atomic formula, with the interpretation that a and b stand in the α relation to one another. For example, if α is "greater than", then αab states that a is greater than b. Also assume the existence of a distinguished predicate, $\succcurlyeq$, representing weak preference (where $(a \succcurlyeq b)$ is used rather than $(\succcurlyeq ab)$). Take note that variables are placeholders, and, until endowed with an interpretation, do not refer to any specific object.

Define the set of well formed formulae recursively: for any well formed formulae, φ and ψ, $\neg\varphi$, $\varphi \wedge \psi$ and $K_t\varphi$ are also well formed (for $t = 0 \ldots T$). K_t represents implicit knowledge at time t, the interpretation of which will be standard: the DM implicitly knows φ (at time t) if φ is true in every state of affairs she considers possible (at time t).[10] In addition, this language allows for universal quantification, $\forall$. So for any well formed φ, $\forall a\varphi$ is well formed, where a any individual variable. The resulting language is $\mathcal{L}(\mathcal{X})$.

Taking the standard shorthand, $\varphi \vee \psi$ is short for $\neg(\neg\varphi \wedge \neg\psi)$, $\varphi \implies \psi$ is short for $\neg\varphi \vee \psi$, and $\exists a\varphi$ is short for $\neg\forall a\neg\varphi$. In addition, let P_t denote $\neg K_t\neg$, with the intended interpretation of $P_t\varphi$ as the DM considers φ possible; she does not know it is not the case. Lastly, to expedite notation, for any φ let $\nabla_t\varphi$ denote $((\varphi \implies K_t\varphi) \wedge (\neg\varphi \implies K_t\neg\varphi))$ with the intended interpretation, the DM knows the truth value of φ at time t.

Per usual, an occurrence of a variable a is *free* in a formula φ if a is not under the scope of a quantifier, and is *bound* otherwise. A formula with no free occurrences is called a *sentence*. Two formulae φ and ψ are called *bound alphabetic variants* of one another if φ and ψ differ only because where φ has well formed sub-formulae of the form $\forall a\zeta$ where ψ has $\forall b\eta$ and ζ has free occurrences of a in exactly the same places as η has free occurrences of b.

If φ is a formula, then $\varphi[a/b]$ denotes the formula created by replacing all (and possibly no) free occurrences of a with b. Because this can change the interpretation of the formula in unintended ways, (in particular, if there was a free a in φ that corresponds to a bound b in $\varphi[a/b]$) I introduce the following notation: $\varphi[[a/b]]$ denotes the formula created first by taking a bound alphabetic variant of φ with no bound occurrences of b, and then replacing every free a with b.

3.2 Semantics

I will work with a fixed domain model. This means the variables (and hence quantification) range over the same domain in every possible world. A word should be said on this, as there is considerable philosophical

[10] The reason this type of knowledge is qualified as implicit, is in contrast to explicit knowledge that requires the DM also be aware of φ.

debate regarding constant domains. On one hand, it simplifies matters considerably to assume the same objects *hypothetically* exist in each possible world. One the other, the very intention that possible worlds be distinct means they might be defined by different objects. Here, I take the view that different possible worlds are defined by the different relation between objects (that may or may not exists, but always hypothetically exist). This view coincides with the main emphasis of looking at different worlds primarily as embodiments of different preferences. The DM conceives of possible worlds in which she entertains different preferences (where these preferences may be the result of different knowledge or awareness) – this is perfectly possible without the introduction of varying domains.

So, to begin, let X denote a domain of the individual variables, the class of all possible values a variable might take. Elements of X are referred to using $x, y, z \ldots$. Notice, this is same notation as used for the domain of consumption alternatives in previous sections. This is intentional, as I will interpret X as being a domain of consumable objects. Let an *assignment* be a function from the set of individual variables into X. If μ and μ' are both assignments such that differ only in the object assigned to a then they are referred to as a-variants, and related by $\mu \sim_a \mu'$.

Then, for a given language, $\mathcal{L}(\mathcal{X})$, each DM is characterized by the tuple $M = \langle S, X, \mathcal{V}, \{R_t\}_{t\leqslant T}, \{\succcurlyeq_s\}_{s\in S}\rangle$. M is referred to as a model (or, a model of decision making). $S = \{s, s', \ldots\}$ is a set of states of the world. $\{R_t\}_{t\leqslant T}$ is a time indexed family of accessibility relations on S; as is standard, the interpretation of $R_t(s) = \{s'|sRs'\}$ is the states the DM considers possible when the true state is s. Truth values will be assigned by $\mathcal{V}$, a function that assigns to each n place predicate and state of the world s, a class of n-tuples from X. The intended interpretation is, for some $\langle X, \mathcal{V}\rangle$, if $(x_1 \ldots x_n) \in \mathcal{V}(\alpha, s)$, with $x_1 \ldots x_n \in X$, then $\alpha x_1 \ldots x_n$ is true in that model in state s.

Let $\mathcal{M}(\mathcal{X})$ be the class of all models based on $\mathcal{L}(\mathcal{X})$, and for a given domain, X, let $\mathcal{M}^X$ denote the subclass of models based on X. Then a DM, M, is represented semantically via the operator $\models$, recursively, as

$$
\begin{aligned}
&(M, s) \models_\mu \alpha a_1 \ldots a_n && \text{iff } (\mu(a_1) \ldots \mu(a_n)) \in \mathcal{V}(\alpha, s), \text{ for any atomic formula (except } \succcurlyeq), \\
&(M, s) \models_\mu (a \succcurlyeq b) && \text{iff } \mu(a) \succcurlyeq_s \mu(b), \\
&(M, s) \models_\mu \neg\varphi && \text{iff not } (M, s) \models_\mu \varphi, \\
&(M, s) \models_\mu (\varphi \wedge \psi) && \text{iff } (M, s) \models_\mu \varphi \text{ and } (M, s) \models_\mu \psi, \\
&(M, s) \models_\mu K_t\varphi && \text{iff for all } s' \in R_t(s),\ (M, s') \models_\mu \varphi, \\
&(M, s) \models_\mu \forall a\varphi && \text{iff for } \mu' \sim_a \mu',\ (M, s) \models_{\mu'} \varphi.
\end{aligned}
$$

A formula φ is *satisfiable* if there exists a M, and a state thereof, s, and an interpretation μ, such that $(M, s) \models_\mu \varphi$. If a $(M, s) \models_\mu \varphi$ for every assignment μ, write $(M, s) \models \varphi$. Given a DM, M, φ is *valid* in M, denoted as $M \models \varphi$, if $(M, s) \models \varphi$ for all s. Likewise, for some class of DMs, $\mathcal{N}$, φ is *valid* in $\mathcal{N}$, denoted as $\mathcal{N} \models \varphi$, if $N \models \varphi$ for all $N \in \mathcal{N}$. Finally, φ is *valid* (i.e., without qualification) if $M \models \varphi$, for all models M.

3.3 Axioms

Consider the following axiom schemata (and inference rules) regarding the language $\mathcal{L}(\mathcal{X})$:

[Prop] *All substitution instances of valid formulae in propositional logic.*

[K] $(K_t\varphi \wedge K_t(\varphi \implies \psi)) \implies K_t\psi.$

[1∀] $\forall a\varphi \implies \varphi[[a/b]]$

[Barcan] $K_t\forall a\varphi \implies \forall a K_t\varphi$

[MP] *From* φ *and* $(\varphi \implies \psi)$ *infer* ψ.

[GenK] *From* φ *infer* $K_t\varphi$.

[Gen∀] *From* $\varphi \implies \psi$ *infer* $\varphi \implies \forall a\psi$*, provided* a *is not free in* φ.

Denote $\boldsymbol{K_T} = (\text{Prop} \cup \text{K} \cup \text{MP} \cup \text{GenK})$ and $\forall\boldsymbol{K_T} = \boldsymbol{K_T} \cup (1\forall \cup \text{Barcan} \cup \text{Gen}\forall)$. It is well known, $\forall\boldsymbol{K_T}$ is a sound and complete axiomatization of the first order language $\mathcal{L}(\mathcal{X})$ with respect to the $\mathcal{M}(\mathcal{X})$ (for example see Chapters 13 and 14 of Hughes and Cresswell (1996)).[11] That is, the above axiom system exactly captures the semantic structure imposed above.

Further axioms can impose structure on the DMs knowledge, and hence, semantically, on the family of accessibility relations. Consider the following:

[D] $K_t\varphi \implies P_t\varphi.$

[T] $K_t\varphi \implies \varphi.$

[4] $K_t\varphi \implies K_tK_t\varphi.$

[5] $P_t\varphi \implies K_tP_t\varphi.$

It is well known, in the presence of $\forall\boldsymbol{K_T}$, T, 4, and 5 correspond to the class of models where $\{R_t\}_{t\leqslant T}$ is a family whose members are reflexive, transitive, and Euclidean,[12] respectively (see Fagin *et al.* (2003) for the propositional case and Hughes and Cresswell (1996) for a first order treatment). Of note is the system $\boldsymbol{S5} = (\forall\boldsymbol{K_T} \cup \text{T} \cup 4 \cup 5)$, corresponding to the class of models where $\{R_t\}_{t\leqslant T}$ is a family whose members are equivalence relations, and therefore, partition the state space.

3.4 Implicit Preferences

Just as we can axiomatize the structure of knowledge, so to can we provide the structure to preferences. Consider the following basic axioms:

[Cmp] $\forall a\forall b\big(\neg(a \succcurlyeq b) \implies (b \succcurlyeq a)\big).$

[Trv] $\forall a\forall b\forall c\big((a \succcurlyeq b) \wedge (b \succcurlyeq c) \implies (a \succcurlyeq c)\big).$

[11] Given an axiom system $\boldsymbol{AX}$, and a language $\mathcal{L}$, we say that the formula $\varphi \in \mathcal{L}$ is a *theorem* of $\boldsymbol{AX}$ is it is an axiom of $\boldsymbol{AX}$ or derivable from previous theorems using rules of inference contained in $\boldsymbol{AX}$. Further, $\boldsymbol{AX}$ is said to be *sound*, for the language $\mathcal{L}$ with respect to a class of structures $\mathcal{N}$ if every theorem of $\boldsymbol{AX}$ is valid in $\mathcal{N}$. Conversely, $\boldsymbol{AX}$ is said to be *complete*, for the language $\mathcal{L}$ with respect to a class of structures $\mathcal{N}$ if every valid formula in $\mathcal{N}$ is a theorem of $\boldsymbol{AX}$.

[12] Recall, a relation is *Euclidian* if xRy and xRz imply yRz.

Denote $\boldsymbol{P} = (\text{Cmp} \cup \text{Trv})$. In a similar spirit to the above results, Cmp, Trv correspond to models where $\{\succcurlyeq_s\}_{s\in S}$ is a family whose members are complete and transitive, respectively. Formally, denote by the superscripts, r, t, e, the classes of DMs such that $\{R_t\}_{t\leqslant T}$ are reflexive, transitive, and Euclidean, and by cmp, trv the class of DMs such that $\{\succcurlyeq_s\}_{s\in S}$ that are complete, and transitive, respectively. For example, $\mathcal{M}^{r,e,cmp,trv}(\mathcal{X})$ is the class of DM for which R_t is reflexive and Euclidean for all t, and $\succcurlyeq_s$ is complete and transitive for all $s \in S$.

Proposition 3.1. *Let $\mathcal{C}$ be a possibly empty subset of $\{\text{T}, 4, 5, \text{Cmp}, \text{Trv}\}$, and let C be the corresponding subset of $\{r, e, t, cmp, trv\}$. Then $\forall \boldsymbol{K_T} \cup \mathcal{C}$ is a sound and complete axiomatization of $\mathcal{L}(\mathcal{X})$ with respect to $\mathcal{M}^C(\mathcal{X})$.*

Proof. In appendix C. ∎

Preferential axioms play the role of traditionally decision theoretic restrictions (i.e., completeness, transitivity, etc); any (satisfiable) theory including these restrictions will have a model of decision making adhering to the corresponding decision theoretic framework. The importance, therefore, of this framework is that it provides us a language to make a clean distinction between non-modal preference and its epistemic counterpart and to analyze the interplay there between. Specifically, the distinction between some elementary (read, true) preference and the preference the decision maker *knows* or is *aware of.*

The discrepancy between the DM's "true" preferences and her implicitly known preferences (and later, in the presence of unawareness, her explicitly known preferences), can be made formal. To do this, define, in addition to $\succcurlyeq_s$, implicitly known preference, $\succcurlyeq_{K_t,s}$.

Definition. *Let M be a model of decision making. Define the implicit preference relation, as*

$$\succcurlyeq_{K_t,s} = \bigcap_{s' \in R_t(s)} \succcurlyeq_{s'} \tag{3.1}$$

The interpretation is as follows: If the true state is s, then the DM is endowed with the preference $\succcurlyeq_s$. However, she may not know this is her preference. In particular, if $(M, s) \models_\mu (a \succcurlyeq b) \wedge \neg K_t(a \succcurlyeq b)$, then in state s at time t, the DM prefers $\mu(a)$ to $\mu(b)$, but does not (implicitly) know her preference. For the DM to make a report $\mu(a) \succcurlyeq \mu(y)$, or take an action on the basis of said preference, she must, as a prerequisite, know she prefers $\mu(a)$ to $\mu(b)$.

Remark 3.2. *$x \succcurlyeq_{K_t,s} y$ if and only if $(M, s) \models_\mu K_t(a \succcurlyeq b)$ holds for any μ such that $\mu(a) = x$ and $\mu(b) = y$.*

If M models a theory that includes preferential axioms, then the DM knows these restrictions. For example, if $M \in \mathcal{M}^{cmp,trv}(\mathcal{X})$, then the DM knows her preference are complete and transitive (since every instance of Cmp and Trv hold at every state, as implied by GenK). This observation implies that while the DM may not know her true preference, $(M, s) \models_\mu \neg K_t(a \succcurlyeq b) \wedge \neg K_t(b \succcurlyeq a)$, she nonetheless knows the structure of her true preference, $(M, s) \models_\mu K_t((a \succcurlyeq b) \vee (b \succcurlyeq a))$. As a result, structural ignorance, where the DM does not know the structure of her preferences, cannot be captured here. This behavior can be captured once awareness is introduced.

When the decision maker's true preferences are a weak order, her implicit preferences are a preorder (i.e., reflexive and transitive). While the implicit preferences inherit reflexivity and transitivity, the same can not be said about completeness. Even if the DM knows her preferences are complete, she might not know what

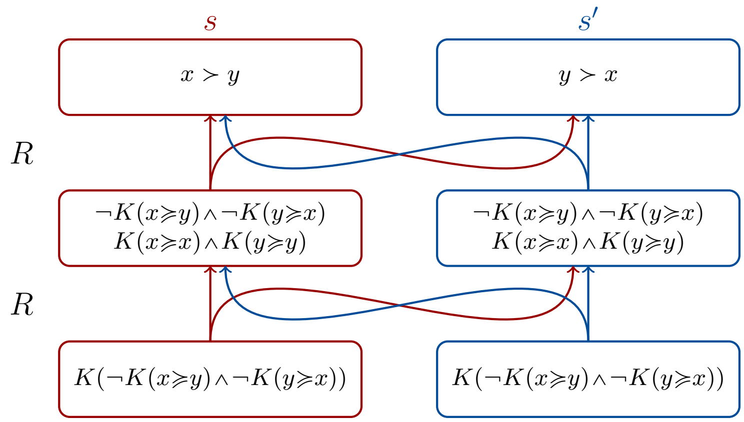

Figure 1: A visual representation of Example 1. The top line provides the truth value of relevant atomic statements; the second line, first order knowledge; the third, second order knowledge. Assessable states are linked by arrows. A statement is implicitly known if it is true in every state linked by an arrow.

her preference is. The framework takes a complete preference relation and returns an incomplete preference relation by nature of introducing uncertainty about the true state of affairs. Indeed, consider the following example.

Example 1. *There is a single time period. There are two states of the world, $S = \{s, s'\}$ and two elements that can be consumed, $X = \{x, y\}$. The preference relations in each state are given by $\succcurlyeq_s = \{(x,x),(x,y),(y,y)\}$ and $\succcurlyeq_{s'} = \{(x,x),(y,x),(y,y)\}$ and the accessibility relation is the trivial $R = S^2$. So, $\succcurlyeq_{K,s} = \succcurlyeq_{K,s'} = \{(x,x),(y,y)\}$, which are not complete. So, $M \models_\mu K((a \succcurlyeq b) \vee (a \succcurlyeq b)) \wedge \neg K(a \succcurlyeq b) \wedge \neg K(a \succcurlyeq b)$, for any μ such that $\mu(a) = x$ and $\mu(b) = y$. Notice, the DM also satisfies negative introspection (i.e., 5), so she knows she does not know her preference: $M \models_\mu K\big(\neg K(a \succcurlyeq b) \wedge \neg K(a \succcurlyeq b)\big)$.*

It is well known, in finite domains, incomplete preferences can be represented by family of utility functions (Ok, 2002; Evren and Ok, 2011). Multi-utility representations identify a set of utility functions, $\mathcal{U}$, such that x is preferred to y if and only if $u(x) \geqslant u(y)$ for all $u \in \mathcal{U}$. By nature of producing an incomplete preference relation, this framework also allows such a representation. From the conceptual analogy between the representation and the set $\mathcal{U}$ and the implicitly known preferences and the set $R(s)$, it is clear this framework instills the natural interpretation to the set of utility functionals.

Remark 3.3. *Let M be a model of $\forall \boldsymbol{K_T} \cup \boldsymbol{P}$. For each $s \in S$, let,*

$$\mathcal{U}_{s,t} = \{u_{s'} : X \to \mathbb{R} \mid u_{s'} \text{ represents } \succcurlyeq_{s'} \text{ and } s' \in R_t(s)\}. \tag{3.2}$$

Then, for every $s \in S$ and $t \leqslant T$,

1. *$\succcurlyeq_{K_t,s}$ is reflexive and transitive, and*
2. *$x \succcurlyeq_{K_t,s} y$ if and only if $u(x) \geqslant u(y)$ for all $u \in \mathcal{U}_{s,t}$.*

Proof. Fix $s \in S$, and $t \leqslant T$. By Proposition 3.1, $(x,x) \in \succcurlyeq_s$ for all $s \in S$ and $x \in X$. So $(x,x) \in \bigcap_{s' \in R_t(s)} \succcurlyeq_{s'} = \succcurlyeq_{K_t,s}$. $\succcurlyeq_{K_t,s}$ is reflexive. Now let, $(x,y),(y,z) \in \succcurlyeq_{K_t,s}$. So, for each $s' \in R_t(s)$, $(x,y),(y,z) \in$

$\succcurlyeq_{s'}$. By Proposition 3.1, $\succcurlyeq_{s'}$ is transitive, so $(x, z) \in \succcurlyeq_{s'}$. Hence, $(x, z) \in \bigcap_{s' \in R_t(s)} \succcurlyeq_{s'} = \succcurlyeq_{K_t,s}$. $\succcurlyeq_{K_t,s}$ is transitive.

Now, to establish the representation claim: $u(x) \geqslant u(y)$ for all $u \in \mathcal{U}_{s,t}$ if and only if $u_{s'}(x) \geqslant u_{s'}(y)$ for all $s' \in R(s)$, where $u_{s'}$ represents $\succcurlyeq_{s'}$. By the definition of a representation, this is if and only if $x \succcurlyeq_{s'} y$ for all $s' \in R_t(s)$, the definition of implicit preference. ■

Moreover, any incomplete preference (over a finite domain) can be generated by preferential logical model. This can be seen by constructing, for each $u \in \mathcal{U}$ some s_u such that $\succcurlyeq_s$ is the order induced by u. Then any model such that $S \supseteq \{s_u | u \in \mathcal{U}\}$ and $R(s) = \{s_u | u \in \mathcal{U}\}$ will suffice. As such, simple models of incomplete preferences can be seen as models of epistemic uncertainty: a DM who is aware of all of her options but is unsure about the true state of affairs. Her incompleteness is derived form a lack of knowledge, even though her preferences exist and she knows her preferences exist. The set of preference relations she considers in her multi utility model are her true preferences in each world she considers possible.

4 Awareness Structures

To directly incorporate unawareness, add two time indexed modal operators, A_t and E_t for $t = 1 \ldots T$, to the logic. Starting with the same set of atomic propositions as above, refer to the resulting language as $\mathcal{L}^A(\mathcal{X})$. The interpretation of the modal operators is as in Fagin and Halpern (1988). K_t is implicit knowledge at time t; an agent knows φ, denoted $K_t\varphi$, if φ is true in all state she considers possible at time t. A_t is awareness at time t; $A_t\varphi$ is interpreted as the DM is aware of φ at time t. Lastly, E_t is explicit knowledge —the conjunction of K_t and A_t. The DM explicitly knows φ if she implicitly knows it and is aware of it: $K_t\varphi \wedge A_t\varphi$. Set P_t^E as short hand for explicit possibility, or $\neg E_t \neg$.[13]

The choice of semantics for awareness in a first order model is non-trivial. One choice is, for example, to assume awareness is comprised of some subset of the domain over which variables range –implying if the decision maker is aware of all the objects specified in a formula, she is aware of the formula itself (i.e., “Object Based Unawareness” investigated by Board and Chung (2011); Board *et al.* (2011)). But this is very limiting, as it prohibits the DM from from being aware of an object but unaware of some of its attributes –a necessary feature to model many economic environments. For example, a DM is choosing between a Mac and a PC, and currently does not know her preference. However, after learning that her job requires her to use a particular piece of software, which runs only on Windows, and which she had been unaware of, she prefers the PC.

Another option, as in the logic of Halpern and Rêgo (2009), is that awareness is comprised of some subset of the language (a set of formulae or sentences). This is closer to ideal but still not exactly right. The reason being, I wish to interpret awareness semantically (i.e., by assigning meaning to the variables in a formula), rather than purely syntactically (i.e., leaving variables as variables). To see why, consider

$$A_0 \forall a (a \succcurlyeq a), \tag{4.1}$$

$$A_0 \forall a (a \succcurlyeq b), \tag{4.2}$$

[13] Explicit possibility states the DM does not explicitly know the negation of a formula –hence she would not act in response to its negation. Note that while explicit knowledge is a stricter requirement that implicit knowledge, the opposite relation is true for possibly. For this reason, Remark 5.4 and Theorem 5.6 require that the DM explicitly knows the explicitly possibility in question.

where both a and b are variables. The first formula states the DM is aware $\succcurlyeq$ is a reflexive relation. The assignment of variables plays no role in the interpretation of the first formula, since the only variable, a, is under the scope of a quantifier. The second formula, however, has a variable b that is not under the scope of any quantifier. To provide this formula with semantic meaning, first b must be given a meaning via a semantic interpretation. The intention is that the DM is aware of information, statements that have specific interpretations, and when b is left as a variable, (4.2) does not have a specific interpretation.

My approach strikes a balance between these two options. I formally construct the awareness structure such that it ranges over formulae (unlike Board and Chung (2011); Board *et al.* (2011)), but *after* they have been assigned a semantic meaning (unlike Halpern and Rêgo (2009)). This construction uses a second language intimately related to $\mathcal{L}^A(\mathcal{X})$ –the language is exactly $\mathcal{L}^A(\mathcal{X})$ but with free variables replaced with logical constants.

4.1 A Semantic Language: $\mathcal{L}(X)$

The following is done with the intention of representing the DM's awareness over semantic statements, formulae that have interpretations –whose variables are assigned to elements of X. Of course, in order to allow for introspection, the DM must be able to be aware of formulae that contain bound variables. To do this, I define an auxiliary language $\mathcal{L}(X)$. The construction of $\mathcal{L}(X)$ is as follows. First, any n place predicate followed by n *members of* X is a well formed *atomic* formula. Then for well formed formulae, $\bar{\varphi}$ and $\bar{\psi}$, $\neg\bar{\varphi}$, $\bar{\varphi} \wedge \bar{\psi}$, $\Box\bar{\varphi}$ are also well formed. Lastly for any formula $\bar{\varphi}$, $\forall a\bar{\varphi}[x/a]$ is well formed, where $\bar{\varphi}[x/a]$ is the formula $\bar{\varphi}$ with all (and possibly no) occurrences of x replaced with a, for some $x \in X$ and $a \in \mathcal{X}$. Notationally, $\bar{\varphi}$ (with a bar) will always denote a formula in $\mathcal{L}(X)$, whereas φ (without a bar) will denote a formula in $\mathcal{L}^A(\mathcal{X})$.

The formulae of $\mathcal{L}(X)$ correspond precisely the formulae of $\mathcal{L}^A(\mathcal{X})$ but with semantic interpretation attached to the free, and only free, variables. To see this, consider *reduced assignments*, which are purely syntactic transformations of formulae. A reduced assignment $\bar{\mu}$, which corresponds to the assignment μ, is a function that takes as an input a well formed formula $\varphi \in \mathcal{L}(\mathcal{X})$ and returns the string of characters created by taking φ and replacing every free variable, a with $\mu(a)$. Lemma 1 states the output of a reduced assignment is always in $\mathcal{L}(X)$, and every formula in $\mathcal{L}(X)$ is the output of some reduced assignment acting on some formula of $\mathcal{L}^A(\mathcal{X})$.

This connection allows us to describe awareness as a set, $\mathcal{A} \subseteq \mathcal{L}(X)$, and hence naturally endowed with a semantic interpretation. Under a given assignment μ, $\mathcal{A}$ represents the DM's awareness of the formulae $\{\varphi \in \mathcal{L}(X) | \bar{\mu}(\varphi) \in \mathcal{A}\}$. Lemma 1, ensures this conception is not limiting: for any $\Gamma \subseteq \mathcal{L}(X)$ there exist some $\mathcal{A}(\Gamma, \mu) = \{\bar{\varphi} \in \mathcal{L}(X) | \bar{\varphi} = \bar{\mu}(\varphi), \varphi \in \Gamma\} \subseteq \mathcal{L}(X)$ such that DM is aware of exactly Γ, given the semantic interpretation μ.

4.2 The Semantics of Awareness

A DM with an awareness structure is described as $M = \langle S, X, \mathcal{V}, \{R_t\}_{t \leqslant T}, \{\succcurlyeq_s\}_{s \in S}, \{\mathcal{A}_t\}_{t \leqslant T}\rangle$, where previously introduce components are as they were before and $\mathcal{A}_t : S \to 2^{\mathcal{L}(X)}$ indicates the set of propositions of which the DM is aware in state s at time t. The interpretation is captured by the following semantics, which are added to the definition of $\models$, in Section 3.2:

$$(M,s) \models_\mu A_t\varphi \qquad \text{iff } \bar{\mu}(\varphi) \in \mathcal{A}_t(s),$$

$$(M,s) \models_\mu E_t\varphi \qquad \text{iff } (M,s) \models_\mu K_t\varphi \text{ and } (M,s) \models_\mu A_t\varphi.$$

At times, it will be helpful to speak of the truth of a formula in $\mathcal{L}(X)$, provided $M \in \mathcal{M}^X$. This will be represented by

$$(M,s) \models_X \bar{\varphi} \qquad \text{iff } (M,s) \models_\mu \varphi, \text{ for some } (\mu,\varphi) \text{ such that } \bar{\mu}(\varphi) = \varphi.$$

Lemma's 1 and 2 guarantee $\models_X$ is well defined. Indeed, by Lemma 1, there always exists a (μ,φ) that stands in the correct relation to $\bar{\varphi}$ and by Lemma 2 the choice of such (μ,φ) does not impact the truth value.

Notice, if φ is a sentence in $\mathcal{L}^A(\mathcal{X})$, then $(M,s) \models_\mu \varphi$ implies $(M,s) \models \varphi$ and $(M,s) \models_X \varphi$ (where $\varphi \in \mathcal{L}(X)$ since it has no free variables). This is well known in general first order modal logic, but it worth pointing out, the same holds in the presence of awareness and explicit knowledge modalities. To see this note the only case not covered by previous arguments is if $\varphi = A_t\psi$. Since φ is a sentence, so too must be ψ. But then, $\bar{\mu}(\psi) = \psi$ for all reduced assignments $\bar{\mu}$, and so, the truth of $A_t\psi$ does not depend on the interpretation.

Fagin and Halpern (1988) showed this semantic structure of unawareness is captured axiomatically by

[A0] $E_t\varphi \iff (K\varphi \wedge A_t\varphi)$.

Axiom A0 simply states that explicit knowledge is the conjunction of implicit knowledge and unawareness. Even with first order models and semantic awareness, this is in the only additional axiom needed for general awareness structures.

Proposition 4.1. $\forall \boldsymbol{K_T} \cup$ A0 *is a sound and complete axiomatization of* $\mathcal{L}^A(\mathcal{X})$ *with respect to* $\mathcal{M}(\mathcal{X})$.

Proof. In appendix C. ∎

Take note, Lemma's 1 and 2 imply that $\forall \boldsymbol{K_T} \cup$ A0 is also a complete and sound axiomatization of $\mathcal{L}(X)$ with respect to $\mathcal{M}^X$.

4.3 Explicit Preference

Just as implicit knowledge gave rise to the DM's implicitly known preference, explicit knowledge will likewise define her explicitly known preferences. As a prerequisite, define the preferences of which the DM is aware.

Definition. *Let M be a model of decision making. Define*

$$\succcurlyeq_{A_t,s} = \{(x,y) | (x \succcurlyeq y) \in \mathcal{A}_s(t)\} \tag{4.3}$$

Just as explicit knowledge is the intersection of knowledge and awareness, explicit preference is the intersection of implicit preference and awareness.

Definition. *Let M be a model of decision making. Define the explicit preference relation, as*

$$\succcurlyeq_{E_t,s} = \succcurlyeq_{A_t,s} \bigcap \succcurlyeq_{K_t,s} \tag{4.4}$$

Remark 4.2. *$x \succcurlyeq_{E_t,s} y$ if and only if $(M, s) \models_\mu E_t(a \succcurlyeq b)$ holds for any μ such that $\mu(a) = x$ and $\mu(b) = y$.*

Just as the DM could have a true preference and not implicitly know it, if $(M, s) \models_X K_t(x \succcurlyeq y) \wedge \neg A_t(a \succcurlyeq y)$, then at time t, in state s, the DM implicitly knows her preference but is unaware of it, and hence does not explicitly know it. As argued earlier implicit knowledge is necessary to act upon a preference. Similarly, when awareness is taken into account, the DM must *explicitly* know her preference in order to act in accordance. The major contribution of this paper, and what is meant as the behavioral implications of unawareness, is the contrast between $\succcurlyeq_{K_t,s}$ and $\succcurlyeq_{E_t,s}$.

Without further restrictions the sets $\succcurlyeq_{E_t,s}$ can any arbitrary subset of $\succcurlyeq_{K_t,s}$. However, given the intent to describe a DM's preferences under unawareness, there are several natural assumptions on the structure of awareness. At this stage, I wish to examine the behavioral implications of a DM who is unaware of some outcomes and contingencies, but acts as rationally as possible with respect to those of which she is aware. In Section 2.1, without the underlying epistemic framework, LCmp stated the DM was aware of some $Y \subseteq X$, over which she had complete preferences. A similar restriction can be made in this environment. For example, if the DM is aware of a preference ranking x weakly to y, $A_t(x \succcurlyeq y)$, it seems only natural she is also aware of the preference ranking y weakly to x. Likewise, if the DM is aware of rankings over x and y, and over z and w, she ought to be aware of rankings over x and z.

[A1] $\forall a \forall b (A_t(a \succcurlyeq b) \iff (A_t(a \succcurlyeq a) \wedge A_t(b \succcurlyeq b)))$.

This restriction does not say anything about the DM's actual or implicitly known preferences –only about her awareness. Nonetheless, there is an intuitive connection between A1 and LCmp: there exists some $Y \subseteq X$, such that the DM is aware of all preference relations among Y, and none outside of Y. This connection can be seen by the following result:

Proposition 4.3. *Let M be a model of $\forall \boldsymbol{K} \cup \boldsymbol{P} \cup$ A0 $\cup$ A1, then $\succcurlyeq_{E_t,s}$ satisfies LRfx and Trv.*

Proof. Let $(x, y) \in \succcurlyeq_{E_t,s}$. Then $(x, y) \in \succcurlyeq_{A_t,s}$: $(M, s) \models_X A_t(x \succcurlyeq y)$. So by A1, $(M, s) \models_X A_t(x \succcurlyeq x)$, or, $(x, x) \in \succcurlyeq_{A_t,s}$. By Remark 3.3, $(x, x) \in \succcurlyeq_{K_t,s}$. So, $(x, x) \in \succcurlyeq_{A_t,s} \bigcap \succcurlyeq_{K_t,s} = \succcurlyeq_{E_t,s}$. The proof for transitivity similar.

■

4.4 The Structure of Awareness

By a similar motivation, if the DM is aware of the statement φ implies x is preferred to y, $A_t(\varphi \implies (x \succcurlyeq y))$, then reasonably she should be aware of the statement φ and the preference $(x \succcurlyeq y)$. This is captured by the notion that $\mathcal{A}$ is closed under "subformulae".

$$[\mathrm{A}\!\downarrow] \begin{cases} \big((A_t \neg \varphi) \vee (A_t \, \square \, \varphi) \vee (A_t(\varphi \wedge \psi)) \vee (A_t(\psi \wedge \varphi))\big) \implies A_t \varphi \\ A_t \forall a \varphi \implies \neg \forall b \neg A_t \varphi[[a/b]] \end{cases}$$

A↓ specifies the DM is able to extrapolate downwards: if the DM is aware of φ, her understanding is full enough to contemplate on the constituent parts of φ. A↓ is not exactly closure under subformulae, since this would entail $A_t \forall a \varphi \implies \forall b A_t \varphi[[a/b]]$: the DM is aware of every instance embodied by $\forall a \varphi$. This is clearly too strong, it implies the DM who is aware her preferences are reflexive, is aware of every object of consumption. To accommodate this subtly, the quantified case is relaxed so that the DM only need be aware of some instance embodied by $\forall a\varphi$. There are two notable consequences of A↓. First, the DM's explicit knowledge is closed under implication: $(E_t\varphi \wedge E_t(\varphi \implies \psi)) \implies E_t\psi$. Second, the DM is unaware of what she is unaware of.

The converse of A↓ can also be dictated.

[A↑] $(A_t\varphi \wedge A_t\psi) \implies \big((A_t\neg\varphi) \wedge (A_t \Box \varphi) \wedge (A_t(\varphi \wedge \psi)) \wedge (A_t(\psi \wedge \varphi)) \wedge A_t\forall a\varphi[[b/a]]\big)$

A↑ specifies the DM is able to extrapolate upwards: if the DM is aware of all of the constituent parts of φ, her understanding is full enough to be aware of φ itself. While A↓ does not seem to be a particularly restrictive notion, A↑ is slightly more controversial as it implies that the DM who is aware of anything is aware of formula of arbitrary complexity.

As in the propositional case, if both A↓ and A↑ are jointly satisfied then $\mathcal{A}_t$ is generated by a set of primitive propositions (i.e., generated according to the construction rules of $\mathcal{L}(X)$). The unawareness of DM who satisfies A↓ and A↑ does not arise from cognitive limitations (she is aware of all of the logical entailments of her own awareness). Instead, her unawareness regards statements completely disjoint from her current world view, things she has *never heard of*.

Lastly, consider the axiom that dictates the DM knows what she is aware of,

[KA] $\nabla_t(A_t\varphi)$.

It is well known that KA distinguishes models in which $\mathcal{A}_t(s) = \mathcal{A}_t(s')$ for all $s' \in R_t(s)$. The interpretation here is that the DM can always (implicitly) reflect on her on awareness, and therefore distinguish between states in which her awareness differs. Of course, this delineation does not necessarily occur at the explicit knowledge level.

Finally, let $\boldsymbol{A}$ denote the axiom system A0 $\cup$ A↓ $\cup$ A↑ and $\boldsymbol{A}^\star$ as $\boldsymbol{A}$ $\cup$ A1 $\cup$ KA.

5 Contingent Planning, or, Finally, The Behavioral Implication of Awareness of Unawareness

With the requisite foundational matters taken care of, I turn to delineating the behavioral effects of incorporating awareness structures. Because I would like to talk about maximality of contingent plans, assume "utilities" are bounded from above by some outcome. This is captured axiomatically by

[Bnd] $\exists a\forall b(a \succcurlyeq b)$.

and, when the DM's preferences is to be bounded when considering the restriction to her awareness set,

[ABnd] $\exists a\forall b(A_t(a \succcurlyeq a) \wedge (A_t(b \succcurlyeq b) \implies (a \succcurlyeq b)))$.

These axioms ensure the decision maker, when fully informed about her preference, would have a optimal consumption choice –so any lack of maximality is not arising from a lack of closure. Notice, also, in any model of full awareness, Bnd and ABnd are materially equivalent.

5.1 Implicitly Known Contingent Plans

It is in this set up that I show awareness of unawareness produces a natural incompleteness in rankings over contingent plans, in particular, even when states are fully contractable. A mapping $(\bar{\varphi} \mapsto x)_t$, where $\bar{\varphi} \in \mathcal{L}(X)$ and $x \in X$, is the commitment to consume x in period t if $\bar{\varphi}$ is true. This is a *partial contingent* *plan*, since it does not specify what happens in states where $\bar{\varphi}$ is not true. Regardless, the following dictates when a rational DM would refuse accept committing to $(\bar{\varphi} \mapsto x)_t$.

Definition. *A partial contingent plan, $(\bar{\varphi} \mapsto x)_t$, is **implicitly unacceptable** to a decision maker, M, in state s, if*

$$(M, s) \models_\mu P_0\big(\varphi \wedge \exists a K_t(a \succ b)\big), \tag{5.1}$$

whenever $\bar{\mu}(\varphi) = \bar{\varphi}$ and $\bar{\mu}(b) = x$.

Parsing the above definition, a DM is unwilling to commit to $(\bar{\varphi} \mapsto x)_t$ if at time 0 the decision maker considers it possible $\bar{\varphi}$ is true and by time t, she will know the existence of an object y, such that y is preferred to x. There are three relevant outcomes at time t: (1) $\bar{\varphi}$ is not true, so her commitment does not bind, (2) $\bar{\varphi}$ is true but she does not know any object she prefers to x, or, (3) $\bar{\varphi}$ is true and she knows an object she prefers to x. So, the DM is unwilling to commit to $(\bar{\varphi} \mapsto x)_t$ if she believes it is possible (3) might occur – by waiting until time t and then making a decision she could be strictly better off. On the other hand, say a contingent plan is acceptable if it is not unacceptable.

Definition. *A partial contingent plan, $(\bar{\varphi} \mapsto x)_t$, is **implicitly acceptable** to a decision maker, M, in state s, if*

$$(M, s) \models_\mu K_0\big(\varphi \implies \forall a P_0(b \succcurlyeq a)\big) \tag{5.2}$$

whenever $\bar{\mu}(\varphi) = \bar{\varphi}$ and $\bar{\mu}(b) = x$.

Parsing this formula, a DM is willing to commit if she knows at time 0, whenever $\bar{\varphi}$ is true she will consider it possible at time t that x is preferred to any other element. In other words, she knows she will not know φ is true and an object y is preferred to x. If the DM knows $\neg\varphi$ then $(\varphi \implies \forall a P_0(b \succcurlyeq a))$ is always true and the DM finds the partial contingent plan acceptable.

Stringing together these partial contingent plans can produce a complete contingent plan. Of course, to be well defined the formulae ought to partition the state space.

Definition. *A finite set of formulae, $\Lambda \subset \mathcal{L}(X)$, is a **contractable set**, or simply contractable, for some class of models, $\mathcal{N} \subseteq \mathcal{M}^X$, if*

1. *$\bigvee_{\bar{\varphi} \in \Lambda} \bar{\varphi}$ is valid in $\mathcal{N}$, and,*
2. *for any distinct $\bar{\varphi}, \bar{\psi} \in \Lambda$, $\bar{\varphi} \wedge \bar{\psi}$ is unsatisfiable in $\mathcal{N}$.*

Sometimes a contractable set will be referred to as (Λ, Γ, μ), under the acknowledgement that Λ is contractable, $\Gamma \subset \mathcal{L}(\mathcal{X})$ and $\bar{\mu}$ defines a bijection between Λ and Γ. An obvious example of the basis for a contingent plan is $\{\bar{\varphi}, \neg\bar{\varphi}\} \subset \mathcal{L}(X)$. Using Λ as a basis, a *period t contingent plan*, c_t, is a mapping from $c_t : \Lambda \to X$. We can define implicit acceptability and unacceptability of complete contingent plans by similar conditions as above.

Definition. *A contingent plan, $c_t : \Lambda \to X$, based on (Λ, Γ, μ), is **implicitly acceptable** to a decision maker, M, in state s, if*

$$(M, s) \models_\mu K_0 \bigwedge_{\varphi \in \Gamma} \Big(\varphi \implies \forall a P_t\big(c_t(\varphi) \succcurlyeq a\big)\Big) \tag{5.3}$$

whenever $\bar{\mu}(\varphi) = \bar{\varphi}$ and $\bar{\mu}(c(\varphi)) = c(\bar{\varphi})$, and provided there is no free occurrence of a in φ, for all $\varphi \in \Lambda$. It is ***implicitly unacceptable****, if, under the same conditions,*

$$(M, s) \models_\mu P_0 \bigvee_{\varphi \in \Gamma} \Big(\varphi \wedge \exists a K_t\big(a \succ c_t(\varphi)\big)\Big). \tag{5.4}$$

Intuitively, a contingent plan is acceptable if it provides outcomes that are no worse than what could have been selected by the DM had she waited until time t and then made a decision. It is unacceptable if this is not the case. Lemma 3 shows, in any model of $\forall \boldsymbol{K_T} \cup$ Cmp, in a given state, every contingent plan is either acceptable or unacceptable (logically, (5.3) and (5.4) are material negations is such models).

Of course, for a contingent plan to be deemed acceptable, the set of formulae on which it is based must be rich enough so that it could mimic any decision making process the DM could have implemented without a contingent plan (i.e., by waiting until t and making a single decision). If this condition is not met by a set Λ, then there may not exist any acceptable contingent plan. Before formalizing this idea, consider the following example.

Example 2. *Let $T = 1$. There are two states of the world $S = \{s, s'\}$. The domain consists of 2 distinct objects, $X = \{x, y\}$. The preference relations, as induced by $\mathcal{V}(\succcurlyeq)$, in each state are given by $\succcurlyeq_s = \{(x,x),(x,y),(y,y)\}$ and $\succcurlyeq_{s'} = \{(x,x),(y,x),(y,y)\}$. The accessibility relations are $R_0 = S^2$, and $R_1 = \{(s,s),(s',s')\}$; the DM is initially uncertain about the state but will be able to distinguish in period 1. It is easily checked that this is a model of $\boldsymbol{S5} \cup \boldsymbol{P} \cup \forall \cup$ Bnd $\cup$ F $\cup$ S_1.*

For any $\bar{\varphi} \in \mathcal{L}(\Delta)$ let $\Lambda = \bar{\psi} = (\bar{\varphi} \vee \neg\bar{\varphi})$. Let μ be any assignment such that $\bar{\mu}(\psi) = \bar{\psi}$ and $\bar{\mu}(b) = x$. Since, ψ is a tautology, and there are no two distinct elements of Λ, Λ is contractable. There are only two contingent plans that can be based on Λ, $\bar{\psi} \mapsto x$, and $\bar{\psi} \mapsto y$. Now, notice

$$(M, s') \models_{\mu'} (a \succ b),$$

for the a-variant of μ where $\mu'(a) = y$. Since, $R_t(s') = \{s'\}$,

$$(M, s') \models_{\mu'} K_t(a \succ b).$$

So, by definition, (and the fact that φ is tautological),

$$(M, s') \models_\mu \varphi \wedge \exists a K_t(a \succ b).$$

Now and finally, since $s' \in R_0(s'')$ for all $s'' \in S$,

$$(M, s') \models_\mu P_0\big(\varphi \wedge \exists a K_t(a \succ b)\big).$$

But since μ was arbitrary, $\bar{\psi} \mapsto x$ is unacceptable in any state $s'' \in S$. Of course the same argument goes to show $\bar{\psi} \mapsto y$ is likewise unacceptable, and hence, there is no acceptable contingent plan.

Example 2 raises the question as to the existence of restrictions on a contractable set Λ that will guarantee the existence of an acceptable contingent plan. In Example 2, no contingent plan allows the DM to consume different objects in different states, which of course, she would want to do. If, however, she waited until period 1, she would be able to make an informed decision and choose the optimal consumption. In order to ensure the existence of acceptable contingent plans, I will consider contractable sets that are rich enough that any decision making process the decision maker could implement without a contingent plan can reproduced with one.

Definition. *Fix a model M. Let $s^{K_t} = \{\bar{\varphi} \in \mathcal{L}(X) | (M, s) \models_X K_t\bar{\varphi}\}$. A contractable set Λ is* ***t-separable*** *if*

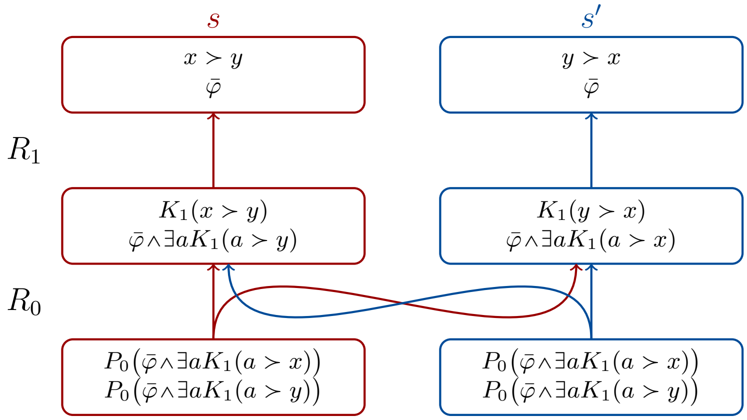

Figure 2: A visual representation of Example 2. The top line provides the truth value of relevant atomic statements; the second line, first order knowledge (in time t); the third, second order knowledge (in time 0 about time t). A statement is possible if it is true in some state linked an arrow.

whenever $s^{K_t} \neq s'^{K_t}$ *then* $(M, s) \models_X \bar{\varphi}$ *implies* $(M, s') \models_X \neg\bar{\varphi}$*, for any* $\varphi \in \Lambda$.

Theorem 5.1. *Let M be a model $\forall \boldsymbol{K_T} \cup \boldsymbol{P} \cup$ T $\cup$ Bnd. For any t-separable and contractable set Λ, there exists a contingent plan, c_t, based on Λ that is acceptable at every s.*

Proof. In appendix D. ∎

Theorem 5.1 states that so long as the contractable set is rich enough to allow the DM to make decisions with at least as much distinction as if she had waited until time t, then she will be willing to commit to something. The intuition is simple: there exists a contingent plan that can implement her optimal choice behavior in time t, so the optimal contingent plan must be weakly better. But since the optimal contingent plan weakly out preforms waiting, she is willing to commit to it.

Without further qualification, Theorem 5.1 does not help to establish the existence of an optimal contingent plan since it requires the existence of a t-separable contingent plan, itself not guaranteed to exist. The next result mollifies this concern, at least in finite models.

Proposition 5.2. *Let M be a model of $\forall \boldsymbol{K_T} \cup \boldsymbol{P} \cup$ T $\cup$ Bnd with a finite state space, S. Then there exists a t-separable contractable set containing $|S|$ formulae.*

Proof. In appendix D. ∎

Putting these two results together, provides a behavioral characterization for full awareness regarding the acceptability of contingent plans.

Corollary 5.3. *Let M be a model of $\forall \boldsymbol{K_T} \cup \boldsymbol{P} \cup$ T $\cup$ Bnd with a finite state space, S. Then for all t, there exists a contractable set $\Lambda(t)$ and a contingent plan there based, c_t, that is acceptable in every $s \in S$.*

While t-separability is sufficient to ensure existence, it is not necessary. For example, consider the case where preferences are constant across states, and so, fully known. Then, even if Λ is a single formula, an acceptable contingent plan exists. This indicates that perhaps the notion of t-separability can be weakened to provide a necessary condition. Indeed, this can be done,[14] but the resulting restriction is more convoluted than insightful. Further, in light of Proposition 5.2, this additional complexity is superfluous. In particular, as it is my focus to show explicit acceptability is harder to satisfy than its implicit counterpart, it suffices to provide any weak condition that provides existence of implicitly acceptable contingent plans in a large class of models.

5.2 Explicitly Known Contingent Plans

The takeaway from the previous section is that if the decision maker is acting according to her implicitly known preference, then she is always willing to commit to *something.* There exists an acceptable contingent plan based on any t-separable set of formulae, itself guaranteed to exist in finite models. The intuition is clear: if a contingent plan can be written in a sufficiently flexible way (i.e., such that it will allow the decision maker to use all available information) there is no reason not to commit. Of course, this line of reasoning relies on the dictate the DM knows what it is she might learn. In other words, the contingent plan allows the DM to specify consumption in the event she learns a particular piece of information, and so it is requisite she know (at time the contingent plan is written) every piece of information she might learn. This is markedly impossible in the event she is unaware of things, and aware of her unawareness!

This is, finally, the behavioral implication of unawareness: the unwillingness to commit to *any* contingent plan, even under circumstances that make knowledge very well behaved. The acceptance of a contingent plan is given by the following, in analogy to (5.3) and (5.4):

Definition. *A contingent plan, $c_t : \Lambda \to X$, based on (Λ, Γ, μ), is **explicitly acceptable** to a decision maker, M, in state s, if*

$$(M, s) \models_\mu E_0 \bigwedge_{\varphi \in \Gamma} \Big(\varphi \implies \forall a P_t^E\big(c_t(\varphi) \succcurlyeq a\big)\Big) \tag{5.5}$$

*whenever $\bar{\mu}(\varphi) = \bar{\varphi}$ and $\bar{\mu}(c(\varphi)) = c(\bar{\varphi})$, and provided there is no free occurrence of a in φ, for all $\varphi \in \Lambda$. It is **explicitly unacceptable**, if, under the same conditions,*

$$(M, s) \models_\mu P_0^E \bigvee_{\varphi \in \Gamma} \Big(\varphi \wedge \exists a E_t\big(a \succ c_t(\varphi)\big)\Big). \tag{5.6}$$

With implicit knowledge, DM a contact is either acceptable or unacceptable, as shown in Lemma 3. With explicit knowledge, the relation is weaker. If a contingent plan is acceptable it is not unacceptable, and if a contingent plan is unacceptable it is not acceptable (each implies the negation of the other), but it is possible neither condition holds.[15] As such, it is not enough to show a contingent plan is not acceptable, instead, it is required to show (5.6) holds.

There exists one additional concern when relaxing full awareness. It may be the DM is unaware of some aspect of the contingent plan itself, and therefore could not make reasonable choices regarding it. For instance, if the formulae on which the plan was based are not in the DM's awareness set. Therefore, it is necessary to impose that contingent plans are articulable to the DM. That is to say, the DM is aware of the constituent parts of the contingent plan (the contingencies, Λ, and the outcomes, $\mathrm{Im}(c)$).

[14] Specifically, if two states, s and s', can be separated via a contingent plan whenever the set of maximal elements of $\succcurlyeq_{K_t,s}$ and $\succcurlyeq_{K_t,s'}$ have an empty intersection.

[15] In particular, this arises when a contingent plan is implicitly acceptable (therefore not explicitly unacceptable), but the DM is unaware of the necessary statements for (5.5) to hold. This misalignment is a consequence of the fact that the awareness modality, A, does not respect material equivalence (i.e., $\varphi \iff \psi$ is a theorem does not imply $A_t\varphi \iff \mathcal{A}_t\psi$) is a theorem). Of course, this property is potentially desirable as it is a very natural relaxation of logical omniscience.

Definition. *Given a DM, M, a contingent plan $c : \Lambda \to X$ is* ***articulable*** *(in state s) if*

1. $\Lambda \subseteq \mathcal{A}_0(s)$.
2. $(x \succcurlyeq y) \in \mathcal{A}_0(s)$ *for any* $x, y \in \mathrm{Im}(c)$.

If a contingent plan is articulable, then DM is able to conceive of it at the time when she would have to commit. If a DM was asked to report her set of potential contingent plans, these are the contingent plans which she would be able to articulate; the contingent plans that are constructible given the language of which she is currently aware. Under full awareness, every contingent plan is articulable. As stated in the introduction, the behavioral characterization relies only on rankings over objects the DM can articulate, and so I will restrict my attention to articulable contingent plans.

The following example shows even under very well behaved knowledge, the existence of unawareness can render *every* articulable contingent plan unacceptable. Of course, this trivially holds when $A_0 = \varnothing$ (since there is no explicit knowledge *and* no articulable contingent plans). Hence, to make a meaningful claim, I am obliged to show something more strict: even when articulable contingent plans exist, they may all be unacceptable and the DM will explicitly know this!

Remark 5.4. *There exist models of $\boldsymbol{S5} \cup \boldsymbol{PA}^\star \cup$ Bnd, that admit articulable contingent plans and such that the DM explicitly knows every articulable contingent plan is explicitly unacceptable.*

Proof. Example 3. ∎

Example 3. *Let S be a single state, s, assessable from itself in every period. Let $X = \{x, y, z\}$, with $\succcurlyeq_s$ defined by $z \succ_s y \succ_s x$. $\mathcal{A}_0$ be the closure (under the construction rules for $\mathcal{L}(X)$) of $\big\{\{x, y\}^2\big\}$, $\mathcal{A}_t$ be the closure of $\big\{\{x, y, z\}^2\big\}$. Let $\bar\varphi$ be any statement of the form $(\bar\psi \vee \neg\bar\psi)$ for some $\bar\psi \in \mathcal{A}_0$. This defines a model of $\boldsymbol{S5} \cup \boldsymbol{P} \cup \forall \cup \boldsymbol{A}^\star \cup$ Bnd $\cup$ F $\cup$ S_t (note: F and S_t are defined in Section 6).*

Note, it is without loss of generality to consider contingent plans of the form $\bar\varphi \mapsto x$ or $\bar\varphi \mapsto y$; since there is only one state there will always be a unique element of the contractable set that is satisfied in the model. Fix, (Λ, Γ, μ) such that $\bar\mu(\varphi) = \bar\varphi$ and a does not occur free in φ. And denote by a, b, c variables such that $\bar\mu(a) = x$, $\bar\mu(b) = y$, and $\bar\mu(c) = z$. We will show $\bar\varphi \mapsto y$ is unacceptable (the argument for $\bar\varphi \mapsto x$ is the same).

By assumption $(z \succ y) \in A_t(s)$:

$$(M, s) \models_\mu A_t(c \succ b)). \tag{5.7}$$

Moreover, since both $\bar\varphi$ and $(x \succ y)$ are in $A_0 \subset A_t$, $P_0^E(\bar\varphi \wedge \exists a E_t(a \succ y))$ is also in $A_0 \subset A_t$:

$$(M, s) \models_\mu A_0 P_0^E(\varphi \wedge \exists a E_t(a \succ b)). \tag{5.8}$$

By assumption $(M, s) \models_\mu (c \succ b)$. Since $R_t(s) = \{s\}$, this implies, $(M, s) \models_\mu K_t(c \succ b)$. Combining this with (5.7), we have $(M, s) \models_\mu E_t(c \succ b)$. This means, for any $\mu' \sim_a \mu$ such that $\mu'(a) = \mu(c)$, $(M, s) \models_{\mu'} E_t(a \succ b)$. By definition,

$$(M, s) \models_\mu \exists a E_t(a \succ b).$$

Applying the fact $R_t(s) = \{s\}$ twice, (and, φ is true at s) implies

$$(M, s) \models_\mu K_0 P_0^E\big(\varphi \wedge \exists a E_t(a \succ b)\big). \tag{5.9}$$

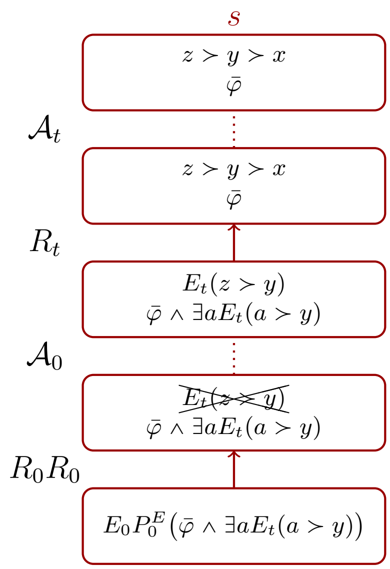

Figure 3: A visual representation of Example 2. Awareness structures are indicated by dotted lines –if a statement is true but not included in $\mathcal{A}$, it is delineated with a strikethrough. Note that the bottom most line applies R twice.

Combining (5.8) *and* (5.9)*,*

$$(M, s) \models_\mu E_0 P_0^E\big(\varphi \wedge \exists a E_t(a \succ b)\big).$$

as desired.

Example 3 shows once awareness is introduced, there is no longer a guarantee of acceptability. The introduction of unawareness (and subsequent focus on explicit rather than implicit acceptability) has fundamentally changed the behavior of the DM –creating a preference for delay that cannot be assuaged by allowing the DM to make conditional decisions. And, by Proposition 5.2, this behavior, unlike incompleteness or a preference for flexibility, cannot be explained in a framework with full awareness, no matter how much uncertainty exists. It is a behavioral trait that indicates the presence of unawareness.

Recall, unacceptability refers here only to articulable contingent plans. In the example, all of the conditions for Proposition 5.2 are met; there exists an implicitly acceptable contingent plan, namely $\bar{\varphi} \mapsto z$. However, this contact is inarticulable because the DM is unaware of z. If the modeler were to ask the DM about $\bar{\varphi} \mapsto z$, she would, after some reflection, be willing to accept it. Of course, in doing so, the modeler would have *made* her aware of z, thereby changing the structure of the very entity whose identification is of interest. It is this subtlety that motivates the departure from the standard revealed preference framework.

As informally argued in the introduction, an unwillingness to commit to any articulable contingent plan is the result of a language that is not rich enough to specify the optimal contingent plan (unawareness), but is rich enough to articulate this fact (awareness of unawareness). The following results formalize this intuition.

Theorem 5.5. *Let M be a model of $\forall \boldsymbol{K_T} \cup \boldsymbol{P} \cup \boldsymbol{A^\star} \cup$ Bnd, such that $(x \succcurlyeq x), (y \succcurlyeq y) \in \mathcal{A}_0(s)$ for some $x, y \in X$. Then if c, based on (Λ, Γ, μ) is an articulable and implicitly acceptable contingent plan in state s,* *it is explicitly acceptable in state s.*

Proof. In appendix D. ∎

Theorem 5.5 states that if the DM can articulate an optimal contingent plan, she will find it explicitly acceptable. Since the contingent plan is implicitly acceptable, the consumption alternatives it specifies are optimal in each state. Moreover, since the DM is aware of these alternatives, she is also aware of the statements professing their optimality (by her ability to extrapolate from explicitly known statements, i.e., A↑). Putting these two facts together delivers the result. This result shows the given formulation of awareness (to wit, under $\boldsymbol{A^\star}$), places clear limits on how unaware a DM can be. Because of the structure of explicit knowledge, the DM cannot be explicitly uncertain if a consumption alternative, x, is optimal: if $\forall a(x \succcurlyeq a) \in \mathcal{A}$ and x is indeed optimal (at all s), the DM explicitly knows this. More generally, unawareness, captured as in this paper, does not allow the DM to question statements she implicitly knows. This limitation is discussed in more detail in Halpern and Rêgo (2013).

Theorem 5.5 places an upper-bound on the DM's awareness such that she finds all articulable contingent plans explicitly unacceptable; Theorem 5.6 places the corresponding lower-bound.

Theorem 5.6. *Let M be a model of $\forall \boldsymbol{K_T} \cup \boldsymbol{P} \cup \mathrm{T} \cup \boldsymbol{A^\star} \cup$ ABnd with a finite state space. Then if M admits articulable contingent plans in state s, and the DM explicitly knows every articulable contingent plan is unacceptable, the DM explicitly knows it is possible she is unaware. Specifically,*

$$(M,s) \models E_0 P_0^E \left(\exists a \left(\neg A_0(a \succcurlyeq a) \wedge A_t(a \succcurlyeq a)\right)\right).$$

Proof. In appendix D. ∎

A DM cannot be so unaware she is not even aware waiting will afford her a more complete world view. That is, the DM must be introspectively unaware. The intuition of this result is straightforward. The DM, acting on explicit knowledge, must explicitly know all contingent plans are unacceptable and this requires she is aware she will have more choices if she does not commit.

6 A Preference for Flexibility

One interpretation of Kreps (1979) is the anticipation of learning induces a *preference for flexibility.* That is, the DM's preference over menus (i.e, subsets of X), respects set inclusion: if $m' \subseteq m \subseteq X$ then m is preferred to m'. A DM who expects to learn her true preference, but is currently uncertain, will prefer the flexibility to make choices contingent on the information she learns. The Krepsian model has a clear connection to the notion of contingent planning (a menu is a restriction on which contingent plans are feasible) as well as more generally to the epistemic framework where the anticipation of learning can be defined precisely. In this section, I will show that the Krepsian framework can be faithfully reproduced as a special case of the general model outlined above. In particular, this special case is one of full awareness; as such, the unforeseen contingencies interpretation is not strictly needed, and a preference for flexibility in not alone the behavioral indication of unawareness.

To reproduce the anticipation of learning, two axioms are needed. First,

[F] $K_t\varphi \implies K_{t'}\varphi$ *for all* $t' \geqslant t$.

The restriction F captures learning by ensuring the decision maker knows (weakly) more at later time periods.

Proposition 6.1. *$\forall \boldsymbol{K_T} \cup$ F is a sound and complete axiomatization of $\mathcal{L}^A(\mathcal{X})$ with respect to the class of models such that $R_{t'} \subseteq R_t$ for all $t' \geqslant t$ (denoted $\mathcal{M}^f$). Moreover, in the $\boldsymbol{S5}$ framework, $\{R_t\}_{t\in T}$ is a filtration of S.*

Proof. In appendix C. ■

Second, that all uncertainty will be realized by time t,

[S_t] $\nabla_t(x \succcurlyeq y)$.

Axiom S_t dictates the DM implicitly knows her preference at time t. From a modeling perspective S_t corresponds to the class of models such that $\succcurlyeq_s = \succcurlyeq_{s'}$ if $s' \in R_t(s)$.

Remark 6.2. *Let M be a model of $\forall \boldsymbol{K_T} \cup \boldsymbol{P} \cup S_t$. The implicit preferences in time t, $\{\succcurlyeq_{K_t,s}\}_{s\in S}$, are complete. Moreover, if M satisfies A0 $\cup$ A1, then $\succcurlyeq_{E_t,s}$ satisfies LCmp.*

Proof. By Proposition 3.1, $\succcurlyeq_s$ is a preference relation. By S_t, $\succcurlyeq_{K_t,s} = \succcurlyeq_s$, and so trivially inherits completeness. Now, let $x \succcurlyeq_{E_t,s} y$ or $y \succcurlyeq_{E_t,s} x$. By A0, this implies $(M,s) \models_X A_t(x \succcurlyeq y) \vee A_t(x \succcurlyeq y)$, and so, by A1, $(M,s) \models_X A_t(x \succcurlyeq x) \wedge A_t(y \succcurlyeq y)$. Moreover, since $\succcurlyeq_{K_t,s}$ is reflexive, $(M,s) \models_X K_t(x \succcurlyeq z) \wedge K_t(y \succcurlyeq u)$. Finally, this implies, $(M,s) \models_X E_t(x \succcurlyeq x) \wedge E_t(y \succcurlyeq y)$, as desired. Conversely, let $x \succcurlyeq_{E_t,s} x$ and $y \succcurlyeq_{E_t,s} y$. Then by A1, $(M,s) \models_X A_t(x \succcurlyeq y)$, and, switching the roles of x and y, $(M,s) \models_X A_t(y \succcurlyeq x)$. The completeness of $\succcurlyeq_{K_t,s}$ delivers the result. ■

To see how axiom F$\cup S_t$ can generate a preference for flexibility consider the simple two period case; let $T = 1$. It is immediate that if $\succcurlyeq_{K_0,s}$ is complete for any s, it follows $\succcurlyeq_{K_1,s}$ is likewise complete. The interesting case is when $\succcurlyeq_{K_0,s}$ is incomplete. So let x, y be elements for which the DM does not have implicitly known preference at time 0 in state s. So:

$$\begin{aligned}
(M,s') &\models_X (x \succcurlyeq y),\\
(M,s'') &\models_X (y \succcurlyeq x),\\
M &\models_X K_1(x \succcurlyeq y) \vee K_1(y \succcurlyeq x),\\
M &\models_X K_0(K_1(x \succcurlyeq y) \vee K_1(y \succcurlyeq x)),
\end{aligned}$$

where $s', s'' \in R_0(s)$. The first two lines rely on Cmp, and the third from S_1 (and Cmp). The fourth line follows directly. This last line shows if the DM is going to learn her true preference by time 1, then at time 0, she knows she will learn her true preference (this is *not* the case once awareness is introduced). As such, she would like the option of choosing either x or y, contingent on what she learns. A preference for flexibility is not the product of learning alone, but requires also the DM acknowledge the possibility she will learn.[16] Now, consider the problem of a DM choosing a menu in period 0 to be the choice set in period t. If the DM knows in period 0 she will know in period t one element of m is preferred to every element in m' then she prefers m to m'. This behavior is captured by the definition of dominance.

[16]This sentiment is echoes the fact that unawareness alone in not sufficient to perturb behavior, and introspective unawareness is required. A preference for flexibility could be seen as the behavioral marker of introspective uncertainty.

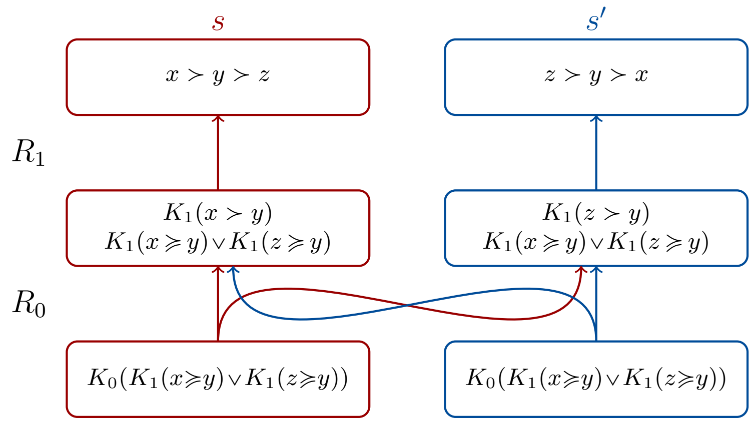

Figure 4: A visual representation of Example 4. The top line provides the truth value of relevant atomic statements; the second line, first order knowledge (in time t); the third, second order knowledge (in time 0 about time t).

Definition. *A menu, m, **s-dominates** a menu, m', (at period t), if and only if*

$$(M, s) \models_X K_0 \bigvee_{z \in m} K_t \bigwedge_{z' \in m'} (z, z'). \tag{6.1}$$

*Further, m **strictly** s-dominates m', if it dominates m' and m' does not dominate m.*

That is, m dominates m' if the DM knows no matter what state of affairs is the true state, she will choose out of the menu m rather than m'.[17] Of course, she does not need to know which element is the maximal one. For example:

Example 4. *Let $S = \{s, s'\}$ and $X = x, y, z$. Then, $x \succ_s y \succ_s z$ and $z \succ_{s'} y \succ_{s'} x$. $R_0 = S^2$ and $R_1 = \{(s, s), (s', s')\}$. M is a model of $\forall \boldsymbol{K_1} \cup \boldsymbol{P} \cup \mathrm{F} \cup S_1$. Notice, $\{x, z\}$ strictly s-dominates $\{y\}$. Indeed, $(M, s) \models K_1(x \succcurlyeq y)$ and $(M \succcurlyeq s') \models K_1(z \succcurlyeq y)$. So, $M \models K_0(K_1(x \succcurlyeq y) \vee K_1(z \succcurlyeq y))$. So, the DM knows, in the true state of affairs, either x or z is preferred to y, but does not know which preference is her true preference.*

The dominance relation is generally incomplete. So, to connect this statement to previous work on preference over menus, where the ranking is usually a weak ordering, s-dominance must be extended to a complete and transitive relation.

Definition. *A preference $\succcurlyeq$ over menus (i.e., a subset of $2^X \times 2^X$) is **$\boldsymbol{FS_1}$-generated** if $\succcurlyeq$ is complete and transitive and there exists some some model, M, of $\forall \boldsymbol{K_1} \cup \boldsymbol{P} \cup \mathrm{F} \cup S_1$, and some state s thereof, such that if m s-dominates m' then $m \succcurlyeq m'$, and if m strictly s-dominates m' then $m \succ m'$.*

The following result shows that beginning with the s-dominance relation, generated by some epistemic model, and extending it to a weak order captures exactly the "preference for flexibility" described in Kreps (1979). In fact, the converse it also true: every Krepsian decision maker can be formulated as the extension of an s-dominance relation with respect to some epistemic model.

[17]Under the restriction S_t, the DM's knowledge (regarding her preference) in period t is exactly her preference, and so inclusion of K_t in (6.1) is superfluous. It is included for conceptual reasons; the DM must know her preference in order to act on it, so if she remains ignorant in period t, there is no reason to desire non-singleton menus. This becomes important to the analysis when S_t is relaxed.

Theorem 6.3. *The preference $\geqslant$ satisfies Kreps' axioms if and only if it is FS_1-generated. Moreover, for any FS_1-generated $\geqslant$, there exists a strictly increasing aggregator $\Gamma : \mathbb{R}^{|R_0(s)|} \to \mathbb{R}$ such that*

$$m \geqslant m' \iff \Gamma(\{\max_{x\in m} u_{s'}(x)\}_{u_{s'}\in\mathcal{U}_{s,0}}) \geqslant \Gamma(\{\max_{x\in m'} u_{s'}(x)\}_{u_{s'}\in\mathcal{U}_{s,0}}), \tag{6.2}$$

where $\mathcal{U}_{s,0}$ is as in (3.2).

Proof. In appendix D. ∎

Notice that the dominance relation, projected onto singleton menus, produces a reflexive and transitive relation on X. In fact, in such models, $x \succcurlyeq_{K_{s,0}} y$ if and only if $\{x\}$ s-dominates $\{y\}$. Hence a preference for flexibility can be seen as a natural extension of multi-utility models.

Finally, I investigate the connection between implicit acceptability and the dominance relation over menus defined in Section 6. Intuitively, when the menu is thought of as the image of a contingent plan, then the contingent plan specifies how the DM will choose out of the menu. Under this interpretation, Theorem 6.4 the provides the connection between having a well defined preference over menus and being willing to accept a contingent plan.

Theorem 6.4. *Let M be a model of $\forall \boldsymbol{K_T} \cup \boldsymbol{P} \cup S_t \cup \mathrm{T} \cup \mathrm{F} \cup$ Bnd. A finite menu $m \subseteq X$ is not strictly s-dominated (with regards to time t consumption) if and only if it is the image of an acceptable (in state s) contingent plan, $c_t : \Lambda \to X$.*

Proof. In appendix D. ∎

If a menu is undominated (according to the definition given by (6.1)), one must be able to construct an acceptable contingent plan from it, and if a contingent plan is acceptable it must induce an undominated menu. This formally establishes the behavior being captured by a preference for flexibility is exactly what facilitates a contingent plan as being acceptable (or not) in models of full awareness. In fact, this result can prove Corollary 5.3 for the case with finite X: since $m = X$ must be undominated it must contain the image of an acceptable contingent plan, and hence, such an object must exist. Conversely, this result suggests a well defined preference (i.e., weak order) over menus is only sensible if acceptable contingent plans exist. It is this last observation (as well as Theorem 6.3) that serves as motivation for moving away from a theory predicated on a preference over menus.

7 Unawareness and Contracts

This section contains a simple example to show how the framework presented above could be used in applications. Assume there are two players: a *principal* (player p), who is offering a take-it-or-leave-it contract to an *agent* (player a). The model takes place in an interactive awareness structure

$$M = \langle S, X, \mathcal{V}, \{R_{t,i}\}_{t\leqslant T, i=1,2}, \{\succcurlyeq_{s,i}\}_{s\in S, i=1,2}, \{\mathcal{A}_{t,i}\}_{t\leqslant T, i=1,2}\rangle,$$

in which players' knowledge and awareness are defined over atomic statements and both their own and their opponents knowledge and awareness.[18] It is in this framework that I will show the principal has an incentive to conceal mutually beneficial information. The intuition being that, although certain novel actions are Pareto improving in every ex-post scenario, the agent will react to the discovery of novel information by becoming more sensitive to her own unawareness, hence increasing her aversion to commitment. In other

[18] It is clear that the same axiomatization will suffice, simply by adding additional indexes to the modalities.

words, the display of surprising outcomes indicates to the agent that the novel outcomes are more valuable than she previously thought; the added value to waiting (and taking an outside option) is greater than the value added by the novel outcome itself. Further, I will show that this incentive can naturally lead to the optimality of incomplete contracts.

If M is any interactive model of decision making under unawareness, define $Y_{t,s,i} \equiv \{x | (x \succcurlyeq_i x) \in \mathcal{A}_{s,i}(t)\} \subseteq X$, the set of outcomes player i is aware of at time t is state s. Further, to make matters simple and tractable, assume each player has a state dependent (expected) utility index from $u_{s,i} : X \to \mathbb{R}$ that represents $\succcurlyeq_{s,i}$, and $\mu_t \in \Delta(S)$, such that

$$\sum_{s \in S} \mu_{t,s,i}(s) u_{s,i}(x)$$

is a completion of $\succcurlyeq_{K_{t,s,i}}$. From Remark 3.3, we know $\mu_{t,s,i}$ must put positive probability on s' if and only if $s' \in R_{t,i}(s)$. Lastly, let $\bar{u}_{s,i} \equiv u_{s,i}|_{Y_{t,s,i}}$. Proposition 4.3 guarantees that $\sum_{s \in S} \mu_{t,s,i}(s) \bar{u}_{s,i}(x)$ is a completion of $\succcurlyeq_{E_{t,s,i}}$ on $Y_{t,s,i}$.

The timing is as follows. In period 0, the principal offers the agent a continent plan to be executed in period 1, which I assume is simply a function from S to X. If the offer is rejected the the agent can take an outside offer, some action in $Q \subseteq X$. If the contract is accepted the principal's and agent's evaluations of contracts are based on their explicitly known preferences:

$$\sum_{s' \in S} \mu_{0,s,i}(s') \bar{u}_{s,i}(c(s'))$$

If the contract is rejected, p gets a utility of 0, and a gets a utility according to her outside option. Of course, to make our problem well defined, we also have quantify the agents perceived value of the outside option. In the case of full awareness (or, naive unawareness, as in Auster (2013)), the agent would have a well defined expected utility over the outside option. This is not the case with introspective unawareness, as the DM is aware of the possibility that waiting will afford novel actions. So consider a mapping $\delta_{t,s,i} : 2^{\mathcal{L}(X)} \to \mathbb{R}$ with the restriction that

$$\delta_{t,s,i}(\mathcal{A}_{t,i}(s)) \geqslant \sum_{s \in S} \max_{x \in Q \cap Y_{0,s,a}} \mu_{t,s,i}(s) \bar{u}_{s,i}(x), \tag{7.1}$$

which, by the characterization in section 5, holds with equality if the DM is fully aware or naively unaware. δ captures the DM's attitude towards unawareness, her perceived value to the objects that she is currently unaware of (also, implicitly, the likelihood of discovering these novel actions).[19]

I focus on the case where actions are verifiable and the principal is fully aware. As such, the principal's problem is simply to offer the acceptable contract that maximizes his payoff. That is, maximize his payoff subject to a participation constraint on behalf of the agent. Moreover, we will assume the agent explicitly knows the principal is fully aware. However, and unlike prior application of awareness, the agent is introspectively unaware.

7.1 The Principal's Problem

Let $S = \{s_1, s_2\}$ and $X = \{x, y, z, w\}$, $T = \{0, 1\}$. Assume that neither agent knows the state in period 0, but both will know the state in period 1 (this is a model of $\boldsymbol{S5} \cup \mathsf{S}_t \cup \mathrm{F} \cup \textsc{Bnd}$). Let $Q = X$. Assume

[19] δ is in principle elicitable via the method described in previous sections –the additional utility the DM would require to accept some contingent plan–although certainty not for counterfactual worlds. Nonetheless, behavioral and experimental data could easily capture the "shape" of δ. Also, under the interpretation of Q as a menu of actions to be chosen from in period 1, the lower bound for δ is the value such a menu (assuming awareness stays the same) as given by (6.2).

that the belief in period 0, for both players and both states, is $(\frac{1}{2}, \frac{1}{2})$, and that utilities are given by,

$u_{s,p}$

	x	y	z	w
s_1	3	1	3	0
s_2	1	3	4	6

$u_{s,a}$

	x	y	z	w
s_1	3	1	3	6
s_2	1	3	4	0

Since we will only consider a limited number of awareness sets, let, for some $D \subseteq X$, $\mathcal{A}(D)$ denote the awareness structure in which the agent is aware only (and all) statements that contain objects in D. In period 0, let $\delta_0(\mathcal{A}(D)) = 3$ if $z, w \notin D$ and $\delta_0(\mathcal{A}(D)) = 5$ otherwise. Since the game ends after period 1, δ_1 is given by (7.1), holding with equality. Consider the case where the agent's initial awareness is $\mathcal{A}(\{x, y\})$ and, if left unperturbed by the principal's offer, it remains her awareness in period 1. That is, without any outside influence, the agent does not become aware of any new objects. However, if the principal offers the contract c, then the agent's awareness set in period 1 is $\mathcal{A}(\{x, y\} \cup Im(C))$.

What is the principal's optimal strategy, given that he is constrained to offer complete contracts? Notice that if he offers the contract $c = (x, y)$ (read: x in state 1, y in state 2), it is accepted and the expected utility (for both players) is 3. It is accepted because the agent is indifferent between accepting and rejecting, given that her perception of the value of the outside option is also 3. It is easy to verify that this is the best the principal can do. To see this, note that if the principal offers a contract containing either z or w, he must provide the agent with a utility of at least 5 (the agent's new value participation constraint). Clearly, this is only possible with the contract $c = (w, z)$. But this gives the principal an expected utility of 2, worse than $c = (x, y)$.

Nonetheless, the contract $c = (x, z)$ makes both players strictly better off, providing a utility of 3.5. Hence, when the principal is constrained to offer a compete contract, he willingly conceals a Pareto improving contract. The intuition is simple: expanding the agent's awareness makes her more aware of her own awareness, and hence she displays a larger aversion to commitment. This second effect outweighs the first, so the principal chooses not to disclose the information.

Now, consider the case where the principal can offer an incomplete contract. Such a contract does not provide any alternative for a particular state, upon the realization of which the players renegotiate. Now the principle can offer the contract $c = (x, \cdot)$ (read: x in state 1, re-negotiate in state 2). This is acceptable to the agent, since $\delta_{0,s_1} = 3 = u_{s_1,a}(x)$. In period 1, if state s_2 is realized, the principal offers the new contract $c = z$. This is again acceptable since $\delta_{1,s_2} = 4 = u_{s_2,a}(z)$. Therefore by appealing to incomplete contracts, the principal can implement his unconstrained optimal contract.

7.2 Incompleteness Requires Introspection

The above example, while highly stylized, is indicative of a general phenomena. Although the effect of unawareness can be quantified via δ, and delay can be calculated, unawareness introduces behavior that intrinsically different than uncertainty. Unlike in the more standard framework, the value of delay (i.e., the outside option) changes with the agents epistemic state, and therefore is itself a function of the contract being offered. As such, there may exist feasible contracts which are initially individually rational, but cease to be so when offered. It is this effect, driven by introspective unawareness, that can make incompleteness strictly beneficial.

Remark 7.1. *Assume the principal is fully aware. Further assume that contracts take the form* $S \to X \backslash Q$

(that is, there are a distinct set of outside options). If there is an incomplete contract c that is strictly better for the principal than any complete contract, then the agent is introspectively unaware.

Proof. We will show there exists some $D \subset X$ such that

$$\delta_{t,s,i}(\mathcal{A}(D)) > \sum_{s \in S} \max_{x \in Q \cap D} \mu_{t,s,i}(s)\bar{u}_{s,i}(x).$$

Since by definition (7.1) holds with equality when the agent is fully aware or naively unaware, this suffices to prove the claim. Assume there was no such D. Then for all $s \in S$ and $D' \subseteq X$ it follows from the definition of δ:

$$\delta_{0,s,a}(\mathcal{A}(D')) = \sum_{s \in S} \max_{x \in Q \cap D'} \mu_{0,s,a}(s)\bar{u}_{s,a}(x). \tag{7.2}$$

Let $\hat{c} : S \to X \backslash Q$, denote the contract that assigns the ex-post choice of the agent (i.e., what was chosen after any renegotiation). Let $\hat{D} = Im(\hat{c}) \cup Y_{0,s,a}$. With regards to $\hat{D}$ assume that (7.2) holds. Since $\hat{c}(s)$ was either acceptable as part of the partial contract c, and so,

$$\bar{u}_{s,a}(\hat{c}(s)) \geqslant \max_{x \in Q \cap Y_{0,s,a}} \bar{u}_{s,a}(x),$$

or chosen ex-post,

$$\bar{u}_{s,a}(\hat{c}(s)) \geqslant \max_{x \in Q \cap Y_{1,s,a}} \bar{u}_{s,a}(x) = \max_{x \in Q \cap Y_{0,s,a}} \bar{u}_{s,a}(x).$$

It follows immediately that $\hat{c}$ would have been acceptable. ■

It is worth briefly addressing the relation between this environment and previous work connecting awareness with incomplete contracts. There is a large body of literature on incomplete contracts arising from the indescribably of states, leading to the well known discussion of Maskin and Tirole (1999). They show show, so long as players understand the utility consequences of states, indescribability should not matter. Here, of course, the unawareness is in the domain of actions rather than states, and so directly and intrinsically regards the understanding of utility. In other words, the utility of the outside option is defined in an articulable manner. Incompleteness, even in the single agent case (i.e., over contingent plans) is welfare improving.

More specifically, when the stipulation that $Q \cap Im(c) = \varnothing$ is dropped, naive awareness can also induce the optimality of incomplete contracts. There, the principal may withhold information strategically, as the novel outcomes may be of direct value as an outside option. This is similar in spirit to the arguments put forth in Filiz-Ozbay (2012) and Auster (2013), where the agents are naively unaware. However, leaving in the stipulation, we see that the discovery of novel outcomes can effect the value of the outside option *indirectly*.

One obviously missing component from the above example, is how the agent's perception of unawareness reacts to the offer of an incomplete contract. It is reasonable to assume the agent believes when such a contract is offered, it must be due to to strategic concerns relating to options outside if her current awareness. Hence the offer of an incomplete contract is itself reason to change her perception of the value of delay. This effect cannot be captured at all by naive unawareness, and highlights the importance of creating a richer epistemic framework. However, because this behavior is complicated, and the agent's shifting perception is likely subject to equilibrium effects, I leave any formal analysis to future work.

8 Literature Review

This paper is within the context of two distinct, albeit related, literatures: that on epistemic logic and unawareness, and that on unawareness and unforeseen contingencies in decision theory. Unawareness was first formalized within modal logic by Fagin and Halpern (1988), who introduced the modal operator for awareness, A, and explicit knowledge, E. This was extended later by Halpern and Rêgo (2009) to include first order statements that allow for introspective unawareness, and extended further by Halpern and Rêgo (2013), to allow the agent to be uncertain about whether she has full awareness or not.

The structure of quantification and awareness in Halpern and Rêgo (2009) is significantly different than the one presented here. They use a logic where quantification is over formulae. Because the foremost concern is over the alternatives that can be consumed, (and over which preferences can be defined), I use a first order modal logic where the variables range only over some fixed domain and are present in formulae only under the scope of predicates. Moreover, I am interested in the awareness of particular objects and so explicitly consider awareness of formula that contain free variables; awareness is a subset of the auxiliary semantic language $\mathcal{L}(X)$.

Relatedly, Board and Chung (2011) and Board *et al.* (2011) propose an alternate structure of awareness that is, like the one presented here, based on objects and predicates. They, however, assume awareness is fully characterized by a set of objects. Then the DM is aware of a statement like "phone x is preferred to phone y" if she is aware of x and y. This does not allow, as I do, for the DM to be aware of an object but not some of its attributes; for example, a DM cannot be aware of x and y but unaware of statements like "phone x has a higher pixel density than phone y." For many practical applications, it seems necessary to disconnect the DM's awareness of attributes of objects from her awareness of objects themselves. Note, this issue cannot be resolved by a simple relabeling of alternatives: for example, phone x' is similar to phone x in all respects save its pixel density. This resolution does not work because phone x has all of its properties regardless of the DM's awareness –a commitment to consume x is a commitment to the relevant properties.

The logic presented in this paper, is a happy medium between that of Halpern and Rêgo (2009) and of Board and Chung (2011). It allows for general awareness structures, ranging over formulae, but *after* the formulae have been given semantic interpretations. Therefore, we can speak of the DM's awareness of particular objects, but not necessarily all of the attributes, or relations between, these objects.

In economics, *state space models* –the semantic structure that include states, and define knowledge and unawareness as operators thereon, as in this paper– have been of particular interest. Modica and Rustichini (1994) and Dekel *et al.* (1998) both provide beautiful, albeit negative, results in this domain. They show, under mild conditions, unawareness must be in some sense trivial; the DM is either fully aware or fully unaware. While Modica and Rustichini (1994) consider a specific awareness modality, Dekel *et al.* (1998) show, under reasonable axioms, state-space models do not allow any non-trivial unawareness operator. As stated, this would be a very damning result for this paper, as it would imply either $\succcurlyeq_E = \succcurlyeq_K$ or $\succcurlyeq_E = \varnothing$, either way, not making for an interesting decision theory. The resolution comes from disentangling explicit and implicit knowledge. Considering these forms of knowledge separately reasonably avoids ever simultaneously satisfying the necessary axioms for their negative result. A far more succinct and intuitive discussion than I could hope to achieve is found in Section 4 of Halpern and Rêgo (2013), and so, I refer the reader there.

Beyond the separation of implicit and explicit knowledge, there have been other approaches to the formalization of unawareness that circumvent the problems outlined in the pervious paragraph. Modica and Rustichini (1999) propose propose models in which the DM is aware only of a subset of formulae (similar in spirit to the awareness sets proposed here, albeit necessarily generated by primitive propositions), and entertains a subjective state space (a coarsening of the objective state space) in which the DM cannot distinguish between any two states that differ only by the truth of a proposition of which she is unaware. Heifetz *et al.* (2006) and Heifetz *et al.* (2008) consider a lattice of state spaces that are ordered according to their expressiveness. In this way, unawareness is captured by events that are not expressible from different spaces –events that are not contained in the event nor the negation of the DM's knowledge. Li (2009) also provides a model with multiple state spaces, where the DM entertains a subjective state space (similar to the above papers, the set of possible state spaces forms a lattice). This allows the DM to be unaware of events in finer state spaces, while having non-trivial knowledge in coarser state spaces.

The decision theoretic take on unawareness is primarily based on a revealed preference framework, and so, unlike its logical counterpart does not dictate the structure of awareness but rather tries to identify it from observable behavior. The first account of this approach (and which predates the literature by a sizable margin) is Kreps (1979). Kreps considers a DM who ranks menus of alternatives, and whose preferences respect set inclusion. The motivation being larger menus provide the DM with the flexibility to make choices after *unforeseen contingencies*. This interpretation, while not strictly ruled out by the model, is certainly not its most obvious interpretation, especially in light of the titular representation theorem. That Krepsian behavior can always be rationalized in a model without appealing to unawareness is shown formally in Theorem 6.3; a longer discussion in relation to this paper is found in Section 6.

More recently there has been a growing interest in modeling the unaware DM. Kochov (2015) posits a behavioral definition of unforeseen contingencies. He considers the DM's ranking over streams of acts (function from the state space to consumption). An event, E, is considered foreseen if all bets on E do not distort otherwise prefect hedges. That is to say, an event is unforeseen if the DM cannot "properly forecast the outcomes of an action" contingent on the event. Kochov shows the events a DM is aware of form a coarsening of the modeler's state space. In a similar vein, Minardi and Savochkin (2015) also contemplate a DM who has a coarser view of the world than the modeler. This coarse perception manifests itself via imperfect updating; the DM cannot "correctly" map the true event onto an event in her subjective state space. The events that are inarticulable in the subjective language of the DM can be interpreted as unforeseen. However, in these works, the objects of which the DM is supposedly unaware are encoded objectively into the alternatives she ranks. Because of this, I argue they are behavioral models of *misinterpretation* rather than unawareness.

Karni and Vierø (2016); Grant and Quiggin (2014) are more explicit about modeling unawareness, and, along with their companion papers, are (to my knowledge) the only decision theoretic paper that deals with unawareness of consumption alternatives, rather than contingencies. They examine a DM who evaluates acts which may specify an alternative explicitly demarcated as "something the DM is unaware of," and who can be interpreted as possessing probabilistic belief regarding the likelihood of discovering such an outcome. They observe the DM's preferences over acts, both before and after the discovery of a new alternative. Of particular interest to the authors, is the process by which the DM updates her beliefs; in particular they provide the axiomatic characterization of *reverse Bayesianism* first developed by the same authors in Karni and Vierø (2013). The assertion of the existence of a novel consumption alternative without dictating what that alternative is has clear parallels with this paper: the DM in their paper is necessarily introspectively unaware. However, their framework takes as given the epistemic state of the decision maker –naive in Karni and Vierø (2013) and introspective in Karni and Vierø (2016).

Grant and Quiggin (2012) develop a general model to deal with unawareness in games, founded on a modal logic which incorporates unawareness in a similar way to Modica and Rustichini (1994). They show that while this model is rich enough to provide non-trivial unawareness, it fails to allow for introspective unawareness, even when first order quantification is permitted. (This limitation arises because of the desired interplay between the structure of knowledge and the structure of awareness as facilitated by the game theoretic environment.) By relaxing the connection with the modal underpinnings, they then consider possible heuristics that a player might exhibit when she inductively reasons that she is introspectively unaware. In a companion paper, Grant and Quiggin (2014) provide a (decision theoretic) axiomatization of such heuristics.

Several recent papers examine the value of information (and expanded awareness) when agents have bounded perception. Quiggin (2015) defines the value of awareness, in analogy to the value of information, and shows the two measures are perfectly negatively correlated. Galanis (2015) examines the value of information under unawareness and shows, in contrast to standard results, the value of information can be negative. Extending this line of thought to the multi agent case (with risk sharing) Galanis (2016), shows, under unawareness, public information might be treated asymmetrically, allowing some agents to prosper at the expense of others.

There are also (very few) economic papers that directly investigate the connection between observable choice and the underlying logical structure. Morris (1996) works somewhat in the reverse direction of the current paper, providing a characterization of different logical axioms (for example, K, T, 4, etc) in terms of preferences over bets on the state of the world. Schipper (2014) extends this methodology to include unawareness structures as described in Heifetz *et al.* (2006). In a similar set up, Schipper (2013) constructs an expected utility framework to elicit (or reveal) a DM's belief regarding the probability of events (when she might be aware of some events). Schipper concludes, the behavioral indication of unawareness of event E is that the DM treats both E *and* its complement as null. This is a very nice result, and the intuition maps nicely to the idea of unawareness. To achieve such a representation, the objects of choice are maps from the "true" state-space to outcomes, and as such, raise the previously discussed issues of observability and the limits of an unaware modeler.

9 Conclusion

This paper contemplates a framework that separates a DM's knowledge from her awareness, in such a way that allows the DM to reason about her own ignorance. Within this environment, I assume the DM has a ranking over consumption alternatives that is informed by her epistemic state (i.e., what she knows and what she is aware of). I show these *epistemic preferences* can serve as a foundation for well known models. The main result is a characterization of the effect of unawareness on observable choice, and the provision of the requisite domain for identification. In static environments, or when the DM is blind to her own unawareness, the presence of unawareness does not produce any local changes to behavior. However, in dynamic contexts and when the DM is introspectively unaware, she will be unwilling to commit to making future choices, even when given the flexibility write a contingent plan that executes a choice conditional on the truth of any piece of information.

A List of Axioms and Inference rules

Axioms		
Axiom	Systems	Notes
Prop	$\boldsymbol{K_T}$, $\boldsymbol{S5}$	Necessary for propositional logic.
K	$\boldsymbol{K_T}$, $\boldsymbol{S5}$	Necessary for propositional modal logic.
D	$\boldsymbol{S5}$	R is serial.
T	$\boldsymbol{S5}$	R is reflexive.
4	$\boldsymbol{S5}$	R is transitive.
5	$\boldsymbol{S5}$	R is euclidian.
Cmp	$\boldsymbol{P}$	$\succcurlyeq$ is complete.
Trv	$\boldsymbol{P}$	$\succcurlyeq$ is transitive.
S_t	FS	$\succcurlyeq$ is known at time t
F	FS	$R_{t'} \subset R_t$; with $\boldsymbol{S5}$, $\{R_t\}_{t\in T}$ is a filtration of S.
Bnd		$\succcurlyeq$ is bounded.
A0	$\boldsymbol{A}$, $\boldsymbol{A}^\star$	Explicit knowledge is the conjunction of implicit knowledge and awareness.
A↓	$\boldsymbol{A}$, $\boldsymbol{A}^\star$	$\mathcal{A}$ is closed under subformulae; with A↑, $\mathcal{A}$ is generated by atomic propositions.
A↑	$\boldsymbol{A}$, $\boldsymbol{A}^\star$	$\mathcal{A}$ is closed under formula construction rules; with A↓, $\mathcal{A}$ is generated by atomic propositions.
A1	$\boldsymbol{A}^\star$	$\succcurlyeq_{E_t}$ is a transitive and locally reflexive; with S_t, $\succcurlyeq_{E_t}$ is a transitive and locally complete.
KA	$\boldsymbol{A}^\star$	$\mathcal{A}$ is implicitly known.
ABnd		$\succcurlyeq$ is bounded when restricted to $\mathcal{A}$.
1∀	∀	Necessary for first order logic.
Barcan	∀	Necessary for constant domain modal logic.
Inference Rules		
Rule	Systems	Notes
MP	$\boldsymbol{K_T}$	Necessary for propositional logic.
GenK	$\boldsymbol{K_T}$	Necessary for propositional modal logic.
Gen∀	∀	Necessary for first order logic.

B Supporting Results

Lemma 1. *$\bar{\varphi} \in \mathcal{L}(X)$ if and only if there exists some reduced assignment $\bar{\mu}$ and formula in $\varphi \in \mathcal{L}(\mathcal{X})$ such that $\bar{\varphi} = \bar{\mu}(\varphi)$.*

Proof. The proof is by induction on the construction of $\bar{\varphi}$. First enumerate the elements of $\mathcal{X}$. Assume $\bar{\varphi}$ is atomic in $\mathcal{L}(X)$. Then $\bar{\varphi} = \alpha x_1 \dots x_n$, for some (not necessarily distinct) $x_1 \dots x_n \in X$. So let φ be the formula $\alpha = a_1 \dots a_n$, with $a_1 \dots a_n$ (note, the a_i's are distinct elements); and μ, any assignment that extends the mapping $\mu : a_i \mapsto x_i$ for $i = 1 \dots n$. It is immediate that $\bar{\varphi} = \bar{\mu}(\varphi)$. In the other direction, if for some assignment μ, $\bar{\varphi} = \bar{\mu}(\varphi) = \alpha\mu(a_1)\dots\mu(a_n)$, then $\bar{\varphi}$ is an atomic $\mathcal{L}(X)$.

So assume the theorem holds for arbitrary $\bar{\varphi}$ and $\bar{\psi}$, with corresponding formulae φ, ψ, and assignments μ and ν. Then $\neg\bar{\varphi} = \neg\bar{\mu}(\varphi) = \bar{\mu}(\neg\varphi)$ and $\Box\bar{\varphi} = \Box\bar{\mu}(\varphi) = \bar{\mu}(\Box\varphi)$, by nature of the fact that μ acts only on variables. Proving both directions.

Next, let $k(\bar{\varphi})$ be the number of occurrences of elements of D in $\bar{\varphi}$ and $k(\bar{\psi})$ the number of occurrences of elements of D in $\bar{\psi}$. Then, let τ be any assignment that extends the mapping

$$\tau : \begin{cases} a_i \mapsto \mu(a_i) \text{ for } i = 1\dots k(\bar{\varphi}) \\ a_i \mapsto \nu(a_i) \text{ for } i = k(\bar{\varphi}) + 1\dots k(\bar{\varphi}) + k(\bar{\psi}). \end{cases}$$

Then, $\bar{\varphi} \wedge \bar{\psi} = \bar{\tau}(\varphi \wedge \tau[a_i/a_{i+k(\bar{\varphi})}])$. In the other direction, assume there exists some assignment μ and $\eta \in \mathcal{L}(\mathcal{X})$ such that $\bar{\mu}(\eta) = \bar{\varphi} \wedge \bar{\psi}$. Since, μ acts only on variables and one character at a time, it is immediate that η must be for the form $\varphi \wedge \psi$ where ψ and ψ are such that $\bar{\mu}(\varphi) = \bar{\varphi}$ and $\bar{\mu}(\psi) = \bar{\psi}$ and so, by the inductive hypothesis, $\bar{\varphi}, \bar{\psi} \in \mathcal{L}(X)$ and therefore so is $\bar{\varphi} \wedge \bar{\psi}$.

Finally, let $C = \{c \in \mathcal{X} | \mu(c) = x, c \text{ occurs in } \varphi\}$ (if x appears in $\bar{\varphi}$, then by the inductive hypothesis, C is non empty). Let $\zeta \in \mathcal{L}(\mathcal{X})$ be the formula that coincides with φ except all (and possibly no) free occurrence of a are replaced with free occurrences of $b \in \mathcal{X}$, where b does not occur (free or bound) in φ, and all (and possibly no) free occurrence of any $c \in C$ are replaced with free occurrences of a (notice there are no free occurrences of any $c \in C$ in ζ (except possible if $a \in C$)). Then let τ be an assignment that coincides with μ everywhere but for a and b, where $\tau(a) = x$ and $\tau(b) = \mu(a)$. Then,

$$\begin{aligned} \bar{\tau}(\forall a\zeta) &= \forall a\bar{\tau}(\zeta)[\tau(a)/a] \\ &= \forall a\bar{\mu}(\varphi)[\tau(a)/a] \\ &= \forall a\bar{\varphi}[\tau(a)/a] \\ &= \forall a\bar{\varphi}[x/a], \end{aligned}$$

where the first equality is definitional, since τ acts only on free variables and there are no free $c \in C$ in ζ, the second follows from the construction of ζ and τ, which ensures $\bar{\mu}(\varphi) = \bar{\tau}(\zeta)$, the third from the inductive hypothesis, and the fourth since $\tau(a) = x$.

In the other direction, assume there exists some assignment μ and $\eta \in \mathcal{L}(\mathcal{X})$ such that $\bar{\mu}(\eta) = \forall a\bar{\varphi}[x/a]$. Since, μ acts only on variables and one character at a time, it is immediate that η must be for the form $\forall a\varphi$, were $\bar{\mu}(\varphi) = (\bar{\varphi}[x/a])[a/\mu(a)]$. So let $\tau \sim_a \mu$ and $\tau(a) = x$. Then, $\bar{\tau}(\varphi) = (\bar{\varphi}[x/a])[a/\tau(a)] = (\bar{\varphi}[x/a])[a/x] = \bar{\varphi}$. So $\bar{\varphi} \in \mathcal{L}(X)$, by the inductive hypothesis, and so, $\forall a\bar{\varphi}[x/a] \in \mathcal{L}(X)$. ■

Lemma 2. *Fix some $\bar{\varphi} \in \mathcal{L}(X)$. Let (φ, μ) and (φ', μ') be such that $\bar{\mu}(\varphi) = \bar{\mu}'(\varphi') = \bar{\varphi}$. Then $(M, s) \models_\mu \varphi$ if and only if $(M, s) \models_{\mu'} \varphi'$*

Proof. The proof is by induction. First, assume $\bar{\varphi}$ is an atomic proposition, i.e., $\bar{\varphi} = \alpha x_1 \dots x_n$. Then $\varphi = \alpha a_1 \dots a_n$ and $\varphi' = \alpha b_1 \dots b_n$, where $\mu(a_i) = \mu'(b_i) = x_i$. Then $(M, s) \models_\mu \varphi$ if any only if $(\mu(a_1)\dots\mu(a_n), s) = (x_1 \dots x_n, s) = (\mu'(b_1)\dots\mu'(ba_n), s) \in \mathcal{V}(\alpha)$, if and only $(M, s) \models_{\mu'} \varphi'$.

So assume the result holds for all formulae of order n. Since reduced assignments preserve the structure of $\neg, \wedge, \Box$, we need only consider the case of $\forall a\bar{\varphi}[x/a]$. So let $\bar{\mu}(\forall a\varphi) = \bar{\mu}'(\forall a\varphi') = \forall a\bar{\varphi}[x/a]$, and $(M, s) \models_\mu \forall a\varphi$. Then for any $\mu^a \sim_a \mu$, $(M, s) \models_{\mu^a} \varphi$. Let $\mu'^a \sim \mu'$ be such that $\mu^a(a) = \mu'^a(a)$. Then notice $\bar{\mu}^a(\varphi) = \bar{\mu}'^a(\varphi')$ (this follows from the fact that when $\mu^a(a) = \mu'^a(a) = x$, it must be that $\bar{\mu}^a(\varphi) = \bar{\mu}'^a(\varphi') = \bar{\varphi}$). So by the inductive hypothesis, $(M, s) \models_{\mu'^a} \varphi'$, for all a-variants of μ'. ■

Lemma 3. *Let M be a model of $\forall \boldsymbol{K_T} \cup \boldsymbol{P}$. Then for any contingent plan, c_t, based on the contractable set (Λ, Γ, μ), in each state, c_t is either implicitly acceptable or implicitly unacceptable.*

Proof. Assume c_t is not implicitly acceptable in state s. Then,

$$(M, s) \models_\mu \neg K_0 \bigwedge_{\varphi \in \Gamma} \big(\varphi \implies \forall a P_t(b \succcurlyeq a)\big).$$

Utilizing $\psi \equiv \neg\neg\psi$, we have

$$(M, s) \models_\mu P_0 \neg \bigwedge_{\varphi \in \Gamma} \big(\varphi \implies \forall a P_t(b \succcurlyeq a)\big).$$

Applying De Morgan's Law, and the definition of $\implies$,

$$(M, s) \models_\mu P_0 \bigvee_{\varphi \in \Gamma} \neg\big(\neg\varphi \vee \forall a P_t(b \succcurlyeq a)\big),$$

and De Morgan's once more, and $\neg\exists a \equiv \forall a \neg$,

$$(M, s) \models_\mu P_0 \bigvee_{\varphi \in \Gamma} \big(\varphi \wedge \exists a \neg P_t(b \succcurlyeq a)\big).$$

Then, since, $\neg P_t \equiv K_t \neg$,

$$(M, s) \models_\mu P_0 \bigvee_{\varphi \in \Gamma} \big(\varphi \wedge \exists a K_t \neg(b \succcurlyeq a)\big).$$

and lastly, from the material equivalence of $\neg(b \succcurlyeq a)$ and $(a \succ b)$ (under $\boldsymbol{P}$, by Proposition 3.1), and the fact that K_t respects material equivalence,

$$(M, s) \models_\mu P_0 \bigvee_{\varphi \in \Gamma} \big(\varphi \wedge \exists a K_t(a \succ b)\big),$$

as desired. ∎

C Soundness and Completeness Results

Because propositions 3.1, 4.1, and 6.1 all relate to the soundness and completeness of particular axiomatizations they are grouped together. Moreover, since Proposition 4.1 is the most general result, it is proven first. The notation there introduced is used without reintroduction in the subsequent proofs.

Proof of Proposition 4.1. The proof of soundness is standard (with perhaps the A0 as the exception, which is immediate). Towards completeness, we will construct a canonical structure. To this end, define $\mathcal{L}^+(\mathcal{X}, \mathcal{Y})$ and the extension of $\mathcal{L}^A(\mathcal{X})$ that contains as atomic formulae exactly the same predicates, and a set of variables that contains $\mathcal{X}$ but, in addition, countably many variables, $\mathcal{Y}$, not in $\mathcal{X}$. A set, $\Lambda \subset \mathcal{L}^+(\mathcal{X}, \mathcal{Y})$ is *admissible* if for every $\varphi \in \mathcal{L}^+(\mathcal{X}, \mathcal{Y})$ for every $a \in \mathcal{X} \cup \mathcal{Y}$, there exists some $b \in \mathcal{X} \cup \mathcal{Y}$ such that $(\varphi[[a/b]] \implies \forall a \varphi) \in \Lambda$.

Now, let S^c be the set of all admissible and maximally $\forall \boldsymbol{K_T} \cup$ A0 consistent sets of formulae in $\mathcal{L}^+(\mathcal{X}, \mathcal{Y})$. Note, by Theorem 14.1 of Hughes and Cresswell (1996), if $\Delta \subset \mathcal{L}(\mathcal{X})$ is consistent then there exists an $s \in S$ such that $\Delta \subseteq s$. Also, define $s^{K_t} = \{\varphi \in \mathcal{L}^+(\mathcal{X}, \mathcal{Y}) | K_t\varphi \in s\}$ and $s^{A_t} = \{\varphi \in \mathcal{L}^+(\mathcal{X}, \mathcal{Y}) | A_t\varphi \in s\}$.

Define the canonical model $M^c = \langle S^c, X^c, \mathcal{V}^c, \{R^c_t\}_{t \leqslant T}, \{\succcurlyeq^c_s\}_{s \in S^c}, \{\mathcal{A}^c_t\}_{t \leqslant T}\rangle$, where S^c is defined as above, $X^c = \mathcal{X} \cup \mathcal{Y}$, $\mathcal{V}^c$ is defined by $(a_1 \ldots a_n, s) \in \mathcal{V}^c(\alpha)$ if and only if $\alpha a_1 \ldots a_n \in s$, R^c_t is defined by $sR^c_t s'$ if and only if $s^{K_t} \subseteq s'$, $\succcurlyeq^c_s = \{(a, b) | (a \succcurlyeq b) \in s\}$, and $\mathcal{A}^c_t(s) = s^{A_t}$ for all t. Finally, define the conical assignment as the identity, $\mu^c : a \mapsto a$.

We will now show, for any $s \in S^c$ and $\varphi \in \mathcal{L}^+(\mathcal{X}, \mathcal{Y})$, $(M^c, s) \models_{\mu^c} \varphi$ if and only if $\varphi \in S$. This will complete the proof, because any $\forall \boldsymbol{K_T} \cup$ A0 consistent $\varphi \in \mathcal{L}(\mathcal{X})$ is contained in some $s \in S^c$, and hence satisfiable. The proof is by induction on the construction of φ. For the base case, note, for any atomic $\alpha a_1 \ldots a_n$ we have $(M^c, s) \models_{\mu^c} \alpha a_1 \ldots a_n$ if any only if $(\mu^c(a_1) \ldots \mu^c(a_n), s) \in \mathcal{V}^c(\alpha)$ if and only if $(a_1 \ldots a_n, s) \in \mathcal{V}^c(\alpha)$ if and only if $\alpha a_1 \ldots a_n \in s$, as desired. Likewise, for any $(a \succcurlyeq b)$, $(M^c, s) \models_{\mu^c} (a \succcurlyeq b)$ if any only if $\mu^c(a) \succcurlyeq_s^c \mu^c(b)$ if and only if $a \succcurlyeq_s^c b$ if and only if $(a \succcurlyeq b) \in s$.

Assume this holds for arbitrary φ and ψ. The inductive step for the cases of $\neg\varphi$, $\varphi \wedge \psi$, $K_t\varphi$ and $\forall a\varphi$ are exactly as in the proof of Theorem 14.3 of Hughes and Cresswell (1996). So assume $(M^c, s) \models_{\mu^c} A_t\varphi$. This is if and only if $\bar{\mu}^c(\varphi) \in \mathcal{A}_t^c(s)$. Since μ is the identity, so is $\bar{\mu}$, and therefore, this is if and only if $\varphi \in \mathcal{A}_t^c(s) = s^{A_t}$ which is if and only if $A_t\varphi \in s$. Lastly, let $(M^c, s) \models_{\mu^c} E_t\varphi$. This is if and only if $(M^c, s) \models_{\mu^c} K_t\varphi$ and $(M^c, s) \models_{\mu^c} A_t\varphi$, so, by the induction hypothesis, if and only if $K_t\varphi \in s$ and $A_t\varphi \in s$. By the properties of maximally consistent sets, $(K_t\varphi \wedge A_t\varphi) \in s$, which by the validity of A0 is if and only if $E_t\varphi \in s$. ■

Proof of Proposition 3.1. The evaluation of propositions in $X \times X$ via of the semantic structure $\{\succcurlyeq_s\}_{s \in S}$ imposes no additional restrictions (namely, no restrictions on R); we could just as well begin with $\mathcal{V}$ and derive $\{\succcurlyeq_s\}_{s \in S}$ there from. In addition, Cmp and Trv, do not involve statements regarding K_t and so impose no restrictions on K_T, and therefore $\{R_t\}_{t \leqslant T}$. Hence, the result when $\mathcal{C}$ is restricted to be in $\{$T, 4, 5$\}$ is as in the well known case (see for example Hughes and Cresswell (1996)).

So let $\{R_t\}_{t \leqslant T}$ be arbitrary. Let $M \in \mathcal{M}^{cmp}$. Let μ be arbitrary. Assume $(M, s) \models_\mu \neg(a \succcurlyeq b)$. By definition $(\mu(a), \mu(b)) \notin \succcurlyeq_s$, and by the completeness of $\succcurlyeq_s$ this implies $(\mu(b), \mu(a)) \in \succcurlyeq_s$, so, $(M, s) \models_\mu (b \succcurlyeq a)$: since μ was arbitrary, Cmp is valid in $\mathcal{M}^{cmp}$. So $\boldsymbol{K} \cup$ Cmp is sound with respect to M^{cmp}.

To show completeness, we construct the canonical structure, $M^{c:cmp}$. To this end, let $S^{c:cmp}$ denote the set of all maximally $\forall \boldsymbol{K_T} \cup$ Cmp consistent sets of formulae in $\mathcal{L}^+(\mathcal{X}, \mathcal{Y})$. Let the rest of the canonical model be defined as in the proof of Proposition 4.1. The result follows if $\succcurlyeq_s^{c:cmp}$ is complete for all s, since then $M^{c:cmp} \in \mathcal{M}^{cmp}$, implying any $\forall \boldsymbol{K_T} \cup$ Cmp consistent formula is satisfiable in $\mathcal{M}^{cmp}$. Fix, $(a, b) \in X^c \times X^c$. Since s is maximally consistent it contains Cmp and either $(a \succcurlyeq b)$ or $\neg(a \succcurlyeq b)$. If $(a \succcurlyeq b) \in s$ then $a \succcurlyeq_s^{c:cmp} b$ and we are done. If $\neg(a \succcurlyeq b) \in s$, then, since s contains Cmp and every instance of 1$\forall$ it contains $\neg(a \succcurlyeq b) \wedge (\neg(a \succcurlyeq b) \implies (b \succcurlyeq a)) \in s$ and consequently, $(b \succcurlyeq a)$. Therefore, $b \succcurlyeq_s^{c:cmp} a$. $\succcurlyeq_s^{c:cmp}$ is complete, as desired.

Now, let $M \in \mathcal{M}^{trv}$, and μ be arbitrary. Assume $(M, s) \models_\mu (a \succcurlyeq b) \wedge (b \succcurlyeq c)$. By definition, this implies $(\mu(a), \mu(b)), ((\mu(b), \mu(c)) \in \succcurlyeq_s$; by the transitivity of $\succcurlyeq_s$ this implies $(\mu(a), \mu(c)) \in \succcurlyeq_s$, so, $(M, s) \models_\mu (a \succcurlyeq c)$. Since μ was arbitrary, $\forall \boldsymbol{K} \cup$ Trv is sound with respect to M^{trv}.

Again, we construct the canonical structure, $M^{c:trv}$, as usual. Assume $a \succcurlyeq_s^{c:trv} b$ and $b \succcurlyeq_s^{c:trv} c$. So, $(a \succcurlyeq b), (b \succcurlyeq c) \in s$. Since s is maximally $\forall \boldsymbol{K} \cup$ Trv consistent it contains Trv and every instance of 1$\forall$, therefore $((a \succcurlyeq b) \wedge (a \succcurlyeq b)) \wedge (((a \succcurlyeq b) \wedge (b \succcurlyeq c)) \implies (a \succcurlyeq c)) \in s$. This implies, $(a \succcurlyeq c) \in s$. Therefore, $\succcurlyeq_s$ is transitive, as desired. ■

Proof of Proposition 6.1. Let $M \in \mathcal{M}^f$, and μ be arbitrary. Let $t \geqslant t'$. Assume $(M, s) \models_\mu K_t\varphi$. So, $(M, s') \models \varphi$ for all $s' \in R_t(s)$. In particular, $(M, s'') \models_\mu \varphi$ for all $s'' \in R_{t'} \subseteq R_t$. By definition, $(M, s) \models_\mu K_{t'}\varphi$. F is valid in $\mathcal{M}^f$; $\forall \boldsymbol{K_T} \cup$ F is sound with respect to M.

Towards completeness, we construct the canonical structure, $M^{c:f}$, as usual. The result follows if $R_{t'}(s) \subseteq R_t(s)$ is true for all $S^{c:f}$, and $t \leqslant t'$. So fix some $s \in S^{c:f}$, and let s' be such that $sR^{c:f}_{t'}s'$. By definition this implies $s^{K_{t'}} \subseteq s'$. Now, let $\varphi \in s^{K_t}$: by definition $K_t\varphi \in s$. Since s contains every instance of F, $(K_t\varphi \implies K_{t'}\varphi) \in s$, and consequently, $K_{t'}\varphi \in s$. By definition $\varphi \in s^{K_{t'}}$ Since φ was arbitrary, $s^{K_t} \subseteq s^{K_{t'}} \subseteq s'$, implying $sR^{c:f}_t s'$, as desired. ■

D Omitted Proofs

Proof of Theorem 5.1. Let $\Lambda \subset \mathcal{L}(X)$ be t-separable and contractable, and (Γ, μ) be such that $\Gamma \subset \mathcal{L}(\mathcal{X})$ and $\bar{\mu}$ defines a bijection between Λ and Γ such that there is no free occurrence of a in any formula of Γ. For each state s, let $x(s)$ denote any $\succcurlyeq_s$ maximal element, guaranteed to exists by Bnd. Denote by $\bar{\Lambda}$ the subset of Λ that are satisfied in some state: $\bar{\Lambda} = \{\bar{\varphi} \in \Lambda | (M, s) \models_X \bar{\varphi}, s \in S\}$. For each $\bar{\varphi} \in \bar{\Lambda}$, define, $\bar{c}_t(\bar{\varphi})$ to be any element of $\{x(s) | s \in S$, such that $(M, s) \models_X \bar{\varphi}\}$. Finally, let c_t be any extension of $\bar{c}_t$ to Λ. It remains to show c_t is acceptable.

It suffices to show there no state such that $(M, s) \models_\mu \neg \bigwedge_{\varphi \in \Gamma} \big(\varphi \implies \forall a P_0(c_t(\varphi) \succcurlyeq a)\big)$. So assume to the contrary, this was true for some state s'. Applying De Morgan's, $(M, s') \models_\mu \bigvee_{\varphi \in \Gamma} \big(\varphi \wedge \exists a K_t(a \succ c_t(\varphi))\big)$. By the definition of contractable, there is a unique $\bar{\psi} \in \Lambda$ (with corresponding $\psi \in \Gamma$) such that $(M, s') \models_X \bar{\psi}$, and so, it must be that $(M, s') \models_\mu \psi \wedge \exists a K_t(a \succ c_t(\psi))$. Therefore, for some $\mu' \sim_a \mu$,

$$(M, s') \models_{\mu'} K_t(a \succ c_t(\psi)), \tag{D.1}$$

Further, by T, $(M, s') \models_{\mu'} (a \succ c_t(\psi))$. i.e., $c_t(\bar{\varphi})$ is not $\succcurlyeq_s$ maximal.

By the construction of c_t, (and the fact that $\bar{\psi} \in \bar{\Lambda}$) there must be some other state, s'', such that $c_t(\varphi)$ is $\succcurlyeq_{s''}$ maximal and $(M, s'') \models_X \bar{\psi}$. But then, by (the contrapositive of) t-separability, $s'^{K_t} = s''^{K_t}$. So, by (D.1), $(M, s'') \models_{\mu'} K_t(a \succ c_t(\psi))$, a contradiction, via T, to the $\succcurlyeq_{s''}$ maximality of $c_t(\varphi)$. ■

Proof of Proposition 5.2. Let M be such a model. Let $\sim_{K_t}$ be the equivalence relation on S defined by $s \sim_{K_t} s'$ if $s^{K_t} = s'^{K_t}$. Let $S/{\sim_{K_t}}$ denote the resulting quotient space of S, with elements $[s]$. Enumerate the elements of $S/{\sim_{K_t}}$. The proof is by induction on the number of elements in $S/{\sim_{K_t}}$. If $S/{\sim_{K_t}}$ contains a single element, any single tautological statement provides t-separable contractable set.

Assume the result hold for n, with the corresponding set $\Lambda_n = \{\bar{\lambda}_{1,n} \ldots \bar{\lambda}_{n,n}\}$. Finally, let $S/{\sim_{K_t}}$ contain $n+1$ elements. By definition of $S/{\sim_{K_t}}$, it must be that for each $[s_i]$, $i \leqslant n$, there exists some statement φ_i, such that (abusing notation: $[s]$ denoting any of its elements),

1. $\bar{\varphi}_i \in [s_i]^{K_t} \backslash [s_{n+1}]^{K_t}$, or,
2. $\bar{\varphi}_i \in [s_{n+1}]^{K_t} \backslash [s_i]^{K_t}$.

So, for each $i \leqslant n$, define $\bar{\psi}_i = \neg K_t \bar{\varphi}_i$ if (1) holds, and $\bar{\psi}_i = K_t \bar{\varphi}_i$ if (2) holds. Define,

$$\bar{\lambda}_{n+1,n+1} = \bigwedge_{i \leqslant n} \bar{\psi}_i,$$

$$\bar{\lambda}_{i,n+1} = \bar{\lambda}_{i,n} \wedge \neg \bar{\lambda}_{n+1,n+1},$$

for $i \leqslant n$. We claim $\Lambda_{n+1} = \{\bar{\lambda}_{1,n+1} \ldots \bar{\lambda}_{n+1,n+1}\}$ is a t-separable contractable state. So let (Γ, μ) be any set such that $\Gamma \subset \mathcal{L}(\mathcal{X})$ and $\bar{\mu}$ defines a bijection between Λ and Γ. Let $\lambda_{i,n+1}$ be the corresponding element to $\bar{\lambda}_{i,n+1}$. It must be that $\lambda_{n+1,n+1}$ is of the form $\bigwedge_{i \leqslant n} \psi_i$, and from the properties of μ, we know $\bar{\mu}(\psi_i) = \bar{\psi}_i$.

Towards contractibility, let M' be any model satisfying the conditions of the theorem, and s' any state thereof. We claim $(M', s') \models_\mu \neg(\lambda_{j,n+1} \wedge \lambda_{k,n+1})$ for $j \neq k$ and $j \neq n+1$. Indeed, if $k = n+1$ this is immediate. If $k \neq n+1$ then the fact that the conjunction of any two distinct formulae in $\Lambda(n)$ is unsatisfiable provides the claim. Further, we claim $(M', s') \models_\mu \bigwedge_{j \leqslant n+1} \lambda_{j,n+1}$. If $(M', s') \models_\mu \lambda_{n+1,n+1}$ we are done, if not, then the validity of the disjunction of all formulae of $\Lambda(n)$ provides the claim.

Towards t-separability, assume $\bar{\psi}_i$ is of the form $\neg K_t \bar{\varphi}_i$ (i.e., (1) holds), and let φ be the corresponding formula of $\mathcal{L}^A(\mathcal{X})$ such that $\psi_i = \neg K_t \varphi$. Then $\bar{\varphi} \notin [s_{n+1}]^{K_t}$. So there does not exist any assignment μ' and φ' such that $\bar{\mu}(\varphi') = \bar{\varphi}_i$ and $(M, [s_{n+1}]) \models_{\mu'} K_t \varphi'$. In particular, this is true for μ and φ; therefore $(M, [s_{n+1}]) \models_\mu \psi_i$. Now, assume $\bar{\psi}_i$ is of the form $K_t \bar{\varphi}_i$ (i.e., (2) holds), again with corresponding φ. Then $\bar{\varphi} \in [s_{n+1}]^{K_t}$. So there exists some μ' and φ' such that $(M, [s_{n+1}]) \models_{\mu'} K_t \varphi'$. But, $\mu'(\varphi') = \varphi = \mu(\varphi)$, so by lemma 2, $(M, [s_{n+1}]) \models_\mu \psi_i$. Hence,

$$(M, [s_{n+1}]) \models_\mu \bar{\lambda}_{n+1,n+1}$$

A similar logic applies to show, for each $i \leqslant n$, $(M, [s_i]) \models_\mu \neg\psi_i$, and so, by the inductive hypothesis,

$$(M, [s_i]) \models_\mu \bar{\lambda}_{i,n+1}$$

so $\Lambda(n+1)$ is t-separable.

■

Proof of Theorem 5.5. Let c_t be articulable and implicitly acceptable –so (5.3) holds. So for every $s' \in R_0(s)$,

$$(M, s') \models_\mu \bigwedge_{\varphi \in \Gamma} \big(\varphi \implies \forall a P_t(c_t(\varphi) \geqslant a)\big)$$

Since possibility implies explicit possibility, (i.e., $P_t\varphi \implies P_t^E \varphi$ is a theorem), we have

$$(M, s') \models_\mu \bigwedge_{\varphi \in \Gamma} \big(\varphi \implies \forall a P_t^E(c_t(\varphi) \geqslant a)\big)$$

which implies,

$$(M, s) \models_\mu K_0 \bigwedge_{\varphi \in \Gamma} \big(\varphi \implies \forall a P_t^E(c_t(\varphi) \geqslant a)\big). \tag{D.2}$$

Now, notice, $(M, s) \models_\mu A_0(a \geqslant a)$ and $(M, s) \models_\mu A_0(b \geqslant b)$ (by the assumption there are two distinct elements of D of which the DM is aware), and $(M, s) \models_\mu A_0\varphi$ for every $\varphi \in \Gamma$ since c_t is articulable. This implies by A↑, and A1,

$$(M, s) \models_\mu A_0 \bigwedge_{\varphi \in \Gamma} \big(\varphi \implies \forall a P_t^E(c_t(\varphi) \geqslant a)\big). \tag{D.3}$$

Combining (D.2) and (D.3), provides the result.

■

Proof of Theorem 5.6. Assume this was not true. First note, since M admits articulable contingent plans, there must be at least some $x \in X$ such that $(x \geqslant x) \in \mathcal{A}_0(s)$. Let $\mu(a) = x$. By A↑ and KA this implies that, for all $s' \in R_0(s)$

$$(M, s') \models_\mu A_0 P^E \left(\exists a \left(\neg A_0(a \geqslant a) \wedge A_t(a \geqslant a)\right)\right),$$

Hence, by our hypothesis, and A0, we have, for some $s' \in R_0(s)$

$$(M, s') \models_\mu E_t \neg (\exists a (\neg A_0(a \succcurlyeq a) \wedge A_t(a \succcurlyeq a))),$$

so by A0 again, we have, for all $s'' \in R_0(s')$,

$$(M, s') \models_\mu \neg (\exists a (\neg A_0(a \succcurlyeq a) \wedge A_t(a \succcurlyeq a))),$$

Clearly, this implies

$$\{(x \succcurlyeq x) : (x \succcurlyeq x) \in \mathcal{A}_t(s'')\} \subseteq \mathcal{A}_0(s'') = \mathcal{A}_{0'} = \mathcal{A}_0(s), \tag{D.4}$$

where the last two equalities come from KA and the fact that $s' \in R_0(s')$ for all $s' \in R_0(s)$, by T.

Order the states in $R_0(s)$, $s_1 \ldots s_n$. So, for each $s_i \in R_0(s)$, let $x(s_i)$ denote any $\succcurlyeq_{s_i}$ maximal element of $\{x : (x \succcurlyeq x) \in \mathcal{A}_t(s')\}$ guaranteed to exist by ABnd. Let μ be such that $\mu(a_i) = x(s_i)$ for each $i \leqslant n$. By (D.4), we have $(M, s) \models_\mu A_0(a_i \succcurlyeq a_i)$ and so by A1,

$$(M, s) \models_\mu A_0(a_i \succcurlyeq a_j), \tag{D.5}$$

for any $i, j \leqslant n$.

Now define for each $i \leqslant n$, define recursively $\bar{\varphi}_i = \bigwedge_{j \leqslant n}(x(s_i) \succcurlyeq x(s_j)) \wedge \neg \bigvee_{j<i} \bar{\varphi}_j$. By Proposition 3.1, $\succcurlyeq_s$ is a weak order and so clearly, $\Lambda = \{\bar{\varphi}_i | i \leqslant n\}$ forms the a contractable set. Moreover, by (D.5) and A$\uparrow$, $c_t : \bar{\varphi}_i \mapsto x(s_i)$ is articulable. But, by construction, is clear c_t is explicitly acceptable, a contradiction. ∎

Proof of Theorem 6.3. First, fix some any model, M of $\forall \boldsymbol{K_1} \cup \boldsymbol{P} \cup \mathrm{F} \cup S_1$. Note, by the completeness of $\succcurlyeq$ we have $m \succcurlyeq m$ for all m, so in order for the definition of an FS_1 generated preference to be well defined, we need dominance to be reflexive. Indeed, for each $s \in S$, $\succcurlyeq_s$ is a preference relation by Proposition 3.1. So for each $s \in S$ let $\bar{m}(s)$ be the $\succcurlyeq_s$-maximal element of m. So by definition of $\succcurlyeq_s$-maximal, and S_1, we have, for all $s \in S$, $(M, s) \models_X K_1 \bigwedge_{z \in m}(\bar{m}(s) \succcurlyeq z)$. Since this holds for all $s \in S$ (in particular, for all $s' \in R_0(s)$), and $\bigcup_{s \in S} \bar{m}(s) \subseteq m$, it follows that $(M, s) \models_X K_0 \bigvee_{y \in m} K_1 \bigwedge_{z \in m}(y \succcurlyeq z)$.

Now, to show sufficiency, we will prove that the representation (6.2) holds. Since $\succcurlyeq$ is complete and transitive, there exists some $V : 2^X \to \mathbb{R}$ that represents it. Define $\xi(m) \equiv \{\max_{x \in m} u_{s'}(x)\}_{u_{s'} \in \mathcal{U}_{s,0}}$. So let Γ be any strictly increasing extension of the map: $\xi(m) \mapsto V(m)$. It remains to show that Γ is well defined. Indeed, if $\xi(m) = \xi(m')$, then for all $s' \in R_0(s)$, we have $\bar{m}(s') \sim_{s'} \bar{m}'(s')$, implying (via S_1),

$$(M, s') \models_X K_1 \bigwedge_{z' \in m'} (\bar{m}(s') \succcurlyeq z') \wedge K_1 \bigwedge_{z \in m} (\bar{m}'(s') \succcurlyeq z).$$

It follows that m s-dominates m' and that m' s-dominates m, so by the requirements of an FS_1 generated preference, $V(m) = V(m')$. Now if $\xi(m) > \xi(m')$ (i.e,. component wise inequality with some strict), we have likewise have for all $s' \in R_0(s)$, $\bar{m}(s') \succcurlyeq_{s'} \bar{m}'(s')$, (with some strict preference) implying (via S_1),

$$(M, s') \models K_1 \bigwedge_{z' \in m'} (\bar{m}(s') \succcurlyeq z'),$$

and for at least one state $s'' \in R_0(s)$,

$$(M, s'') \models \neg K_1 \bigwedge_{z \in m} (\bar{m}'(s'') \succcurlyeq z').$$

It follows that m strictly s-dominates m', so by the requirements of an FS_1 generated preference, $V(m) > V(m')$, as desired.

Towards necessity, we will construct the FS_1 model that generates $\succcurlyeq$. So let $\succcurlyeq$ satisfy the axioms of Kreps (1979), and so, the representation therein holds, (i.e., of the form of (6.2), with an arbitrary state space, Ω). It is easy to check the following model suffices, $S \cong \Omega$, $\mathcal{V}$ can be arbitrary, $R_0 = \Omega^2$, $R_1 = \bigcup_{\omega\in\Omega}(\omega,\omega)$, and for each ω, let $\succcurlyeq_\omega$ be the order generated by u_ω in the initial representation. ∎

Proof of Theorem 6.4. First, assume m is the image of such a contingent plan: $m = \mathrm{Im}(c_t)$, with c_t based on (Λ, Γ, μ). By way of contradiction, assume m is strictly dominated by m'. So,

$$(M,s) \models_\mu \neg K_0 \bigvee_{b\in\mu^{-1}(m)} K_t \bigwedge_{a\in\mu^{-1}(m')} (b \succcurlyeq a),$$

which implies for some $s' \in R_0(s)$, and all $b \in \mu^{-1}(m)$ we have

$$(M,s') \models_\mu \neg K_t \bigwedge_{a\in\mu^{-1}(m')} (b \succcurlyeq a),$$

which in turn implies, for some $s'' \in R_t(s')$ and some $a \in \mu^{-1}(m')$ and all $b \in \mu^{-1}(m)$, we have $(M,s'') \models_\mu (a \succ b)$. Now S_t implies

$$(M,s') \models_\mu K_t(a \succ b).$$

or, $(M,s') \models_\mu \exists a K_t(a \succ b)$. Let $\bar\psi$ (with $\psi = \bar\mu^{-1}(\bar\psi)$) denote the unique element of Λ such that $(M,s') \models_X \bar\psi$. By assumption, $c_t(\bar\psi) \in \mu^{-1}(m)$, so, $(M,s') \models_\mu \psi \wedge \exists a K_t(a \succ (c_t(\psi))$. This clearly implies

$$(M,s') \models_\mu \bigvee_{\varphi\in\Gamma} \varphi \wedge \exists a K_t(a \succ c_t(\varphi)).$$

Lastly, since s' is assessable from s at 0, we have a contradiction to the acceptability of c.

Now, assume m is an undominated menu. For each $s \in S$, let $x(s)$ denote any $\succcurlyeq_s$ maximal element in m. Let Λ be a t-separable set. The proof proceeds as in Theorem 5.1. Denote by $\bar\Lambda$ the subset of Λ that are satisfied in some state: $\bar\Lambda = \{\bar\varphi \in \Lambda | (M,s) \models_X \bar\varphi\}$. For each $\bar\varphi \in \bar\Lambda$, define, $\bar c_t(\bar\varphi)$ to be any element of $\{x(s) | s \in S$, such that $(M,s) \models_X \bar\varphi\}$. Finally, let c_t be any extension of $\bar c_t$ to Λ. It remains to show c_t is acceptable.

Assume it was not. Then there exists some $s' \in R_0(s)$ such that,

$$(M,s') \models_\mu \bigvee_{\varphi\in\Gamma} \big(\varphi \wedge \exists a K_t(a \succ c_t(\varphi))\big). \tag{D.6}$$

By the definition of contractable, there must be some $\varphi \in \Gamma$ such that $(M,s') \models_\mu \varphi$. As shown in Theorem 5.1, it must be $c_t(\bar\varphi)$ is $\succcurlyeq_{s'}$ maximal element in m. By (D.6), there must be some $\mu' \sim_a \mu$ such that $(M,s') \models_{\mu'} (a \succ c_t(\varphi))$, and so, by S_t, $(M,s') \models_{\mu'} K_t(a \succ c_t(\varphi))$. It is immediate that $m \cup \bar\mu'(a)$ strictly s-dominates m, a contradiction.

N.B. this argument only shows that m *contains* the image of an acceptable contingent plan. However, we can always add a list of logical contradictions to Λ in order to exhaust the remainder of m (which is finite by assumption).

∎

REFERENCES

Aumann, Robert (1962). Utility theory without the completeness axiom. *Econometrica* **30**(3), 445–462.

Auster, Sarah (2013). Asymmetric awareness and moral hazard. *Games and Economic Behavior* **82**, 503–521.

Board, Oliver J. and Kim Sau Chung (2011). Object-Based Unawareness: Axioms. *Working paper.*

Board, Oliver J., Kim Sau Chung and Burkhard C. Schipper (2011). Two models of unawareness: Comparing the object-based and the subjective-state-space approaches. *Synthese* **179**(SUPPL. 1), 13–34.

Dekel, Eddie, Barton L. Lipman and Aldo Rustichini (1998). Standard State-Space Models Preclude Unawareness. *Econometrica* **66**(1), 159–173.

Dillenberger, David, Juan Sebastián Lleras, Philipp Sadowski and Norio Takeoka (2014). A theory of subjective learning. *Journal of Economic Theory* pp. 1–31.

Dillenberger, David, R. Vijay Krishna and Philipp Sadowski (2015). Dynamic Rational Inattention. *Working paper.*

Eliaz, Kfir and Efe A. Ok (2006). Indifference or indecisiveness? Choice-theoretic foundations of incomplete preferences. *Games and Economic Behavior* **56**(1), 61–86.

Ergin, Haluk and Todd Sarver (2010). A Unique Costly Contemplation Representation. *Econometrica* **78**(4), 1285–1339.

Evren, Özgür and Efe A. Ok (2011). On the multi-utility representation of preference relations. *Journal of Mathematical Economics* **47**(4-5), 554–563.

Fagin, Ronald and Joseph Y. Halpern (1988). Belief, Awarness and Limited Reasoning. *Artificial Intelligence* **34**(1988), 491–501.

Fagin, Ronald, Joseph Y. Halpern, Yoram Moses and Moshe Y. Vardi (2003). *Reasoning About Knowledge.* Vol. 1.

Feinberg, Yossi (2012). Games with unawareness. *Working paper.*

Filiz-Ozbay, Emel (2012). Incorporating unawareness into contract theory. *Games and Economic Behavior* **76**(1), 181–194.

Galanis, Spyros (2015). The value of information under unawareness. *Journal of Economic Theory* **157**, 384–396.

Galanis, Spyros (2016). The value of information in risk-sharing environments with unawareness. *Games and Economic Behavior* **97**, 1–18.

Grant, Simon and John Quiggin (2012). Induction reasoning about unawareness. *Economic Theory* pp. 1–39.

Grant, Simon and John Quiggin (2014). A preference model for choice subject to surprise. *Theory and Decision.*

Halpern, Joseph Y. and Leandro C. Rêgo (2009). Reasoning about knowledge of unawareness. *Games and Economics Behavior.*

Halpern, Joseph Y. and Leandro C. Rêgo (2013). Reasoning about knowledge of unawareness revisited. *Mathematical Social Sciences.*

Heifetz, Aviad, Martin Meier and Burkhard C. Schipper (2006). Interactive unawareness. *Journal of Economic Theory* **130**(1), 78–94.

Heifetz, Aviad, Martin Meier and Burkhard C. Schipper (2008). A canonical model for interactive unawareness. *Games and Economic Behavior* **62**, 304–324.

Heifetz, Aviad, Martin Meier and Burkhard C. Schipper (2013). Unawareness, beliefs, and speculative trade. *Games and Economic Behavior* **77**(1), 100–121.

Hughes, G. E. and M. J. Cresswell (1996). *A New Introduction to Modal Logic.* Vol. 107. Routledge.

Karni, Edi and Marie-Louise Vierø (2013). "Reverse Bayesianism": A choice-based theory of growing awareness. *American Economic Review* **103**(7), 2790–2810.

Karni, Edi and Marie-Louise Vierø (2016). Awareness of Unawareness : A Theory of Decision Making in the Face of Ignorance. *Working paper.*

Kochov, Asen (2015). A Behavioral Definition of Unforeseen Contingencies. *Working paper.*

Kreps, David M (1979). A Representation Theorem for "Preference for Flexibility". *Econometrica* **47**(3), 565–577.

Li, Jing (2009). Information structures with unawareness. *Journal of Economic Theory* **144**(3), 977–993.

Maskin, Eric and Jean Tirole (1999). Unforeseen Contingencies and Incomplete Contracts. *Review of Economic Studies* **66**, 83–114.

Minardi, Stefania and Andrei Savochkin (2015). Subjective Contingencies and Limited Bayesian Updating.

Modica, Salvatore and Aldo Rustichini (1994). Awareness and partitional information structures. *Theory and Decision* **37**(1), 107–124.

Modica, Salvatore and Aldo Rustichini (1999). Unawareness and Partitional Information Structures. *Games and Economic Behavior* **27**(2), 265–298.

Morris, Stephen (1996). The Logic of Belief and Belief Change: A Decision Theoretic Approach. *Journal of Economic Theory* **69**(35), 1–23.

Ok, Efe A. (2002). Utility Representation of an Incomplete Preference Relation. *Journal of Economic Theory* **449**(November), 429–449.

Piermont, Evan, Norio Takeoka and Roee Teper (2015). Learning the Krepsian State: Exploration Through Consumption. *Working paper.*

Quiggin, John (2015). The value of information and the value of awareness. *Theory and Decision* **80**(MAY), 167–185.

Riella, Gil (2013). Preference for Flexibility and Dynamic Consistency. *Journal of Economic Theory* **148**(6), 2467–2482.

Schipper, Burkhard C. (2013). Awareness-dependent subjective expected utility. *International Journal of Game Theory* **42**(3), 725–753.

Schipper, Burkhard C. (2014). Preference-based unawareness. *Mathematical Social Sciences* **70**, 34–41.