WP:115 ISSN:1749-3641 (online)

Modal Constructions in Sociological Arguments*

László Pólos
Durham University

Michael T. Hannan
Stanford University

Greta Hsu
University of California, Davis

January 12, 2009

Abstract

This paper introduces modal logics to a sociological audience. We first provide an overview of the formal properties of this family of models and outline key differences with classical first-order logic. We then build a model to represent processes of perception and belief core to social theories. To do this, we define our multi-modal language and then add substantive constraints that specify the inferential behavior of modalities for perception, default, and belief. We illustrate the deployment of this language to the theory of legitimation proposed by Hannan, Pólos, and Carroll (2007). This paper aims to call attention to the potential benefits of modal logics for theory building in sociology.

*We thank the Stanford Graduate School of Business and Durham Business School for financial support.

Introduction

Sociological theorizing can benefit from logical formalization in three substantially different ways. The first reconstructs existing sociological arguments in standard logics, usually propositional logic or first-order logic[1] to check on the validity of arguments and to sharpen them (see Péli, Bruggeman, Mausch, and O'Nualláin 1994; Hannan 1997; Kamps and Pólos 1999; Péli, Pólos, and Hannan 2000). The second develops new approaches for logical rendering of argumentation in sociological theory development (Pólos and Hannan 2002, 2004; Hannan et al. 2007). The third way focuses on the (formal) languages used to express core concepts in sociological theories. This paper focuses on the third way of using logic to sharpen sociological theory building.

A key choice in this kind of work concerns what family of semantic models can establish a firm imagery of the intended interpretation of the theory (Bach 1986). Such imagery can turn out to be a useful heuristic tool for identifying new empirically testable claims and substantively meaningful insights.

We propose an intensional, multi-modal language as an attractive alternative for expressing core sociological concepts. The semantics for these modal languages (models for modal languages in technical jargon) are usually presented as Kripke frames (also called possible-world semantics); we describe and apply this modeling technique in this paper.

We base our proposal on the belief that retaining the core sociological insight is challenging in translations from natural language to an artificial (constructed) language. Many efforts at formalization introduce assumptions and simplify arguments in ways that fail to resonate closely with intuitions behind the original theories (Hannan 1997). In such cases, the insights that made these arguments appealing in the first place are washed away. A structural mismatch between theories and formalization tools makes this challenge particularly severe.

Logicians design different logics to represent different patterns of reasoning. A core insight might be extremely cumbersome or even impossible to capture with the language of one logic, while a different logic facilitates a parsimonious and sensible interpretation. Finding a language with adequate expressive power is a fundamental issue for any formalization effort (van Benthem 1995a).

[1] An interesting exception is Montgomery's (2005) reconstruction of role theory in a (standard) nonmonotonic logic.

While there are a great variety of logics, the default in modern science and mathematics is classical predicate (first-order) logic. This logic provides a formal characterization of the reasoning patterns typically found in mathematics. It should not be surprising, however, that typical patterns of reasoning within mathematics do not fit easily with those inherent in much sociological theorizing.

A key mismatch arises from the fact that predicate logic builds on the property called *extensionality.* The extension of a predicate refers to the set of objects for which the predicate is true. For instance, the predicate "Is red" has as its extension the set of objects (in the universe of discourse) that are red, that satisfy the property "is red". In an extensional logic the replacement of names by other names that point to the same entities (that have identical denotations) yields logically indistinguishable expressions. Consider a simple version of a well-known example: "Superman" and "Clark Kent" both designate the same (fictional) entity. In an extensional framework, the statements "Superman is faster than a speeding bullet" and "Clark Kent is faster than a speeding bullet" are equivalent (have the same sense, the same truth value).[2] There is no reason to suppose that any agent treats these claims as interchangeable. It is easy to see that sociological intuitions can get lost in the translation when extensional logics are used to interpret natural-language arguments.

Instead of forcing formalizations of sociological theories into an ill-fitting extensional mold or dismissing formalization outright, researchers should consider logics that provide a better fit. Modern logic offers some appealing alternatives. We argue that modal logics deserve attention. The syntax and semantics of appropriately chosen modal logics can remain close to natural-language arguments, and still one can derive implications, prove soundness, check for consistency, and so forth.

First we offer a gentle introduction to modal logic and possible-worlds semantics. Second we discuss informally some of the desired properties of the temporal multi-modal logic we develop. After discussing these desiderata, we

[2] In the original discussion of this idea, Frege (1892) introduced the distinction between *sinn* (sense or meaning) and *bedeutung* (reference) using the example of the "morning star" and the "evening star," which astronomers discovered are the same object—Venus. He argued that the sentences "The 'morning star' is the 'morning star'." and "The 'morning star' is the 'evening star.'" have the same meaning, because the two names point to the same object. However, the first sentence seems to be true merely by virtue of the law that every object is identical to itself; but the second sentence states an astronomical discovery.

lay out a definition of our multi-modal language and of its semantics. These definitions allow us to prove that this logic actually delivers what we wanted from it. The next step in the model building adds some substantive (or material) constraints to the general model so that the belief frames we define offer a reasonably realistic model of the considered perception–default–belief processes. In the final step, we attempt to illustrate the potential value of one modal logic by considering (and sharpening) its role in the theory of legitimation proposed by Hannan et al. (2007).

1 Models of the Language and Models of Reality

Before considering modal languages, we briefly sketch the two alternative approaches to defining a logic and its language; and we explain which we utilize. The syntactic (or axiomatic) approach treats logics as pure formalisms. The semantic (or model-based approach) builds the language of a logic to reflect the truth and falsity of statements about the world.

The syntactic/axiomatic approach to the logical consequence relation is based on derivations of proofs. Specifically this approach holds that the formula ψ follows as a syntactic consequence of another formula φ if and only if there is a proof of ψ from φ. Proofs begin with logical axioms and the premise(s) and end with the proven theorem. A logical axiom can be, for example, any instantiation of the following schema $\varphi \to (\psi \to \varphi)$. (This axiom is sound because it is false only if φ is both true and false.) New formulas can be added to a proof if they can be derived from the logical axioms and premises. The syntactic method has to specify the legitimate derivation steps to tell what can be derived. Two common choices are (1) substitution (in the formula above φ and ψ can be substituted with any proposition and the formula remains an axiom) and (2) modus ponens (if $\varphi \to \psi$ and φ is derivable, then so is ψ).

The other main option, the semantic/model-theoretic approach, defines the logical consequence relation in terms of the meanings of the premise(s) and of the conclusion.[3] The details of such an approach depend on the kind of meanings used in the semantic model. Two typical choices for meanings

[3]A combination of the syntactic and semantic approach tries to prove the completeness theorem, to show that a given syntactic rendering of the relation and a semantic definition actually depict exactly the same relation.

are (1) truth conditions and (2) update conditions. According to the truth-conditional approach, a set of premises semantically imply a conclusion if the truth of the premises guarantees the truth of the conclusion. (We discuss the update-conditional approach below.)

For our purposes the semantic option offers relevant insights. We want to model belief reports, perceptual reports, and taken-for-granted assumptions mean in terms of statements about an agent's information about the world. Therefore we focus on logical consequence within a semantic framework. According to one common-sense and widespread rendering, the logical consequence relation must be *truth preserving:* no valid inference should yield false conclusions from true premises (Etchemendy 1990). Unfortunately it is normally hard to determine the truth of a proposition (in the world). Nonetheless, we can provide a systematic answer to the question of truth (and falsity) by defining formal models that are structured in the same way as the propositions. This means providing what logicians call a *model of the language.*

This model-theoretic strategy yields a recursive definition of the truth of any proposition stated in the language. Consider first the so-called atomic sentences, those that do not have subparts that are themselves sentences. Suppose that all relevant atomic propositions have a "subject-predicate" format, such as "Anchor Brewing Co. is a microbrewer." Then the truth/falsity of any atomic sentence can be determined by knowing (1) the extension of each predicate and (2) the denotations of subjects' names. A sentence stated in this language is true just in case the denotation of the subject of the sentence is an element of the extension of the predicate stated by the sentence. For instance, the sentence referred to above is true if the entity denoted by "Anchor Brewing Co." belongs to the set of "microbrewers" (lies in the extension of that concept) as it does in the real world. For more complex, molecular sentences (those that are not atomic) the truth-conditions can be defined recursively. For example, $\neg\varphi$ is true if and only if φ is false, $\varphi \to \psi$ is true if and only if it is not the case that φ is false and ψ is true, and so forth.

Such a recursive definition of a language yields a *family* of models. Models within a family can differ with respect to the denotations of the names and the extensions of the predicates (which depend on the substantive application, whether we study brewers or banks, for instance), but they all follow the same general pattern. Once we have defined a family of models of the language, a specification of the logical consequence relation almost automatically offers itself. Proposition φ logically implies proposition ψ (in notation: $\varphi \Rightarrow \psi$) if no model (in the family) makes φ true while making ψ false. In other words, the set of propositions composed of the premise(s) and the negation of the conclusion cannot be satisfied. (Because any finite number of premises can be replaced by their conjunction, defining the consequence relation for one premise suffices for the cases we consider in this paper.)

2 Modal Logics

Now we consider the problem of modeling propositions that contain concepts involving an agent's perceptions, taken for granteds, or beliefs. Such concepts are central to sociological theories, including the one we analyze. If we want to argue about (and possibly predict) the consequences of what agents perceive, take for granted, or believe, then a potentially fruitful strategy is turning to modern logic for languages that can express these ideas formally. Logic can be used to produce coherent models of perceptions, defaults, and beliefs. These models are obviously based on idealizations. Whether these idealizations are ultimately justifiable should be tested by examining their implications empirically. If these models predict empirically, then they are justified as (at least) heuristic tools.

Belief statements typically express a relationship between an agent and a proposition. Consider for instance the statement "John believes that Anchor Brewing Co. is a microbrewer." This sentence expresses a relationship between the agent "John" and the proposition "Anchor Brewing Co. is a microbrewer". Linguists refer to such sentences as propositional attitude reports. We focus on a subtype of such reports, which we call *belief reports*.

We try to utilize some of the insights developed in linguistics and semantics in studies of propositional attitudes. Attitude verbs such as "believes," "doubts," "perceives," and "assumes" are functional expressions[4]

The dominant Fregean tradition in formal semantics recognizes two types of functional expressions: extensional and intensional. An *extensional* functor operates on the denotations of its arguments; an *intensional* functor operates on the connotation (intension) of its argument. (We explain the formal notion of intension below.)

---

[4]A functional expression takes certain linguistic expressions and combines them into a new expression. For example, the conjunction takes two formulas and produces a new formula with two argument slots: one for a noun phrase (a name) and one for a sentence (a proposition).

Clearly, attitude verbs are extensional in their first arguments: replacing "John" with a different name (e.g., John's social security number) does not change the truth value of the belief statement so long as the new name refers to the same agent as "John" does. The second argument slot plays a different role: it contains a declarative sentence. If we replace "Anchor Brewing Co. is a microbrewer" with another proposition with the same truth value, then we cannot always predict the truth value of the resulting proposition. Both "Anchor Brewing Co. is a microbrewer" and "Sierra Nevada Brewing Co. is a microbrewer" happen to be true in the actual world. It might still be the case that the proposition "John believes that Anchor Brewing Co. is a microbrewer" is true and "John believes that Sierra Nevada Brewing Co. is a microbrewer" is false. Such conditions violate the property of extensionality. Logicians refer to attitudes with this kind of property as intensional attitudes.

Because belief attitudes are intensional, it is natural to argue that intensional logics have potential value as tools for building systematic models of argumentation about belief reports. And we pursue this path. But we face a complication. Devising adequate criteria for identifying whether a pair of attitude reports point to the same scenario has proven to be difficult. It seems safe to argue for a sufficient criterion: (1) if two propositions are logically equivalent, then they refer to the same scenario, and (2) if proposition φ logically implies proposition ψ (in notation, $\varphi \Rightarrow \psi$), then the scenario referred to by proposition φ is a part of the scenario referred to by proposition ψ.[5] Taking this consideration into account, we conclude that belief reports should satisfy a principle of logical closure. Consider the belief attitude report "John believes φ" and suppose that φ logically implies (entails) ψ (in notation $\varphi \Rightarrow \psi$). Then it is also the case that "John believes ψ." In technical terms, belief reports should be closed under logical deduction.

In short, we think that appropriate languages for analyzing belief reports contain sentential operators (operators that take sentences and yield other sentences) that are (1) intensional and (2) closed under logical deduction. These two properties define modalities.

In logic, the term modality originally was used (by C. S. Lewis and others) to refer to qualities of the truth of an expression, especially the possibility and

[5]Those who assume that two scenarios that are mutually part of each other are identical would find the first consideration redundant. We prefer to follow a noncommittal approach on this issue.

necessity of the truth of an expression. A statement might be possibly true, necessarily true, and so forth. The technical apparatus developed to analyze logics that contain operators for possibility and necessity were successfully generalized for analysis of statements about an agent's attitude toward an object or relation, and the term modality now generally includes expressions of perceptions, beliefs, and valuations. We use this extended sense of modality.

The various models for modal logics all build on the technical notion of *possible worlds.* In constructions using this notion, the model concerns a set of possible worlds, where each world consists of a distinct state of affairs defined over the propositions to be analyzed.[6] For instance, suppose we consider the possible worlds defined over two statements about empirical facts, say φ and ψ. Each statement can be true or false. So we have four possible worlds: (1) $\varphi \wedge \psi$ (both sentences are true), (2) $\neg(\varphi \vee \psi)$ (both are false), (3) $\varphi \wedge \neg\psi$ and (4) $\neg\varphi \wedge \psi$.

Agents often have incomplete and inaccurate perceptions, taken for granteds, and beliefs about the world around them. For instance, they do not perceive some aspects of reality, and some of what they perceive as true is false in the actual world. For example, someone might believe $\varphi \wedge \psi$ when in reality $\varphi \wedge \neg\psi$ is true. Because the standard predicate logic deals only with facts (extensional statements), it does not provide a way to express such misperception. Since social and behavioral theories often assume that agents base their behaviors on their perceptions and recognize that perceptions are fallible, a language that can express possible worlds might improve formal translation efforts in sociology.

Formal Properties of Modal Logics

To this point we have treated modalities informally. Now we begin to spell out their formal properties.

As we noted above, modalities are a class of sentential (or propositional) operators. These operators take sentences as input and produce other sentences

[6] Opinions differ on the ontological status of the possible worlds. At one extreme, Lewis argues that the possible worlds are a set of actual worlds, much as some contemporary physicists argue for the existence of parallel universes. Others regard the set of worlds as nothing more than an index over a set of possible alternative assignments of truth values. We adopt the latter view, as will make clear below.

as output. Modal operators are sentential operators that are closed under logical deduction. (A sentential operator ω lacks closure under logical deduction if it is possible that $\omega(\varphi)$ is true for some proposition φ but $\omega(\psi)$ is false even though φ logically implies ψ.)

The modalities first studied by philosophers (at least as far back as Aristotle and Diodoros Cronos) are the metaphysically important modalities for *necessity* ("it is necessary that") and *possibility* ("it is possible that"). The customary notation uses $\Box$ for the necessity operator and $\Diamond$ for the possibility operator.

These modalities are duals of each other. "It is necessary that" means that "it is not possible that it is not" (in notation, $\Box\varphi \Leftrightarrow \neg\Diamond\neg\varphi$, and conversely $\Diamond\varphi \Leftrightarrow \neg\Box\neg\varphi$).

What other properties ought to be attributed to these modalities is controversial. Consider, for example, the principle that anything that is necessary is also true: $\Box\varphi \Rightarrow \varphi$. If "necessary" is interpreted as "it ought to be the case," then the principle is not valid. But, other accepted interpretations do make the principle valid. Modal logics with this property are called alethic logics; those that do not incorporate this principle but use the weaker principle $\Box\varphi \Rightarrow \Diamond\varphi$ are called deontic logics.

We think that a sensible model of belief reports as modalities requires use of deontic logic. This is because we need to allow for the possibility of false beliefs, and the alethic principle cannot hold in general if beliefs can be mistaken. On the other hand, we want beliefs to be consistent. That is, if an agent believes that the proposition φ, then we want to constrain the situation so that the agent also believes that $\neg\varphi$ is false. This is exactly what is needed to grant the deontic principle.

Another debated principle, imposed in some modal logics, holds that necessarily true statements are necessarily necessarily true: $\Box\varphi \Rightarrow \Box\Box\varphi$. For belief attitudes this principle has potential value for distinguishing perceptions, defaults, and beliefs on the basis of positive introspection. For example, if $\Box\varphi \Rightarrow \Box\Box\varphi$ holds and $\Box\varphi$ is interpreted as "the focal agent believes that φ is the case", then the same agent "believes that she believes that φ" is implied by "she believes that φ is the case." But we do not see any reason to stipulate that such a strong positive introspection characterizes our agents, and we do not impose this principle.

The study of modal logics flourished in the second half of the twentieth century after Saul Kripke discovered systematic ways of making sense of these proliferating logics and proposed a (formal) semantics for modalities. The easiest way to explain Kripke models uses Leibniz's idea of possible worlds. The now-

Table 1: Example of all of the possible worlds for the case of three propositions expressed in terms of truth values.

w1	w2	w3	w4
$\lvert\varphi\rvert = T$	$\lvert\varphi\rvert = T$	$\lvert\varphi\rvert = T$	$\lvert\varphi\rvert = F$
$\lvert\psi\rvert = T$	$\lvert\psi\rvert = T$	$\lvert\psi\rvert = F$	$\lvert\psi\rvert = T$
$\lvert\chi\rvert = T$	$\lvert\chi\rvert = F$	$\lvert\chi\rvert = T$	$\lvert\chi\rvert = T$
w5	**w6**	**w7**	**w8**
$\lvert\varphi\rvert = F$	$\lvert\varphi\rvert = F$	$\lvert\varphi\rvert = T$	$\lvert\varphi\rvert = F$
$\lvert\psi\rvert = F$	$\lvert\psi\rvert = T$	$\lvert\psi\rvert = F$	$\lvert\psi\rvert = F$
$\lvert\chi\rvert = T$	$\lvert\chi\rvert = F$	$\lvert\chi\rvert = F$	$\lvert\chi\rvert = F$

standard view holds that $\Box\varphi$ is true if and only if φ is true in all possible worlds and that $\Diamond\varphi$ is true just in case φ is true in at least one possible world. This sounds straightforward, provided that we are willing to assume a multiplicity of possible worlds. Any effort to collect information about different possible worlds, however, is problematic, because we can study empirically only one of the possible worlds—the actual one.

For our purposes it is sufficient to consider possible worlds as different "interpretations" of the language of a (non-modal) propositional logic; this allows us to avoid further complicating the setup. In propositional logic, as we explained above, all propositions are built from atomic propositions by application of functions such as negation, conjunction, disjunction, material implication, and so forth. For a classical propositional logic an *interpretation of the language* is nothing but the assignment of truth values to the atomic propositions. The truth value of any complex proposition is determined by the truth values of its atomic propositions. Suppose, for example, that we have three atomic propositions (φ, ψ, and χ) whose truth values are indicated by $|\varphi| = T$ (for true), $|\varphi| = F$ (for false), etc. Table 1 shows the set of possible worlds (alternative interpretations) for this language. In general, n-many propositional atoms yield 2^n possible worlds.

In classical logic, the truth of a proposition depends only on the chosen model of the language. A model of the language of classical first-order logic recursively defines truth values for all the propositions of the language as we noted above. First the atomic propositions get their truth values defined (interpreted, as we used the term above), and then the complex propositions get their truth values following their logical form (their syntactic composition). For example, a negated proposition is true if the proposition is false, and so forth. This construction of the language relies on a property called *truth functionality:* the truth (or falsity) of the components together with the (fixed) definitions of the logical operators defines the truth of compositions.

Modal logics are more complicated because some compositions do not necessarily respect truth functionality. Because truth values differ among possible worlds by construction, truth and falsity in an intensional case depends both on the model of the language *and* on the choice of the world. Questions of truth and falsity for modal logics focus on a function, called an *intension*, that tells for every possible world whether a given proposition is true or false in that world. An intension associates a proposition, say φ, and a possible world, say w, with a truth-value of 1 or 0, for truth and falsity, respectively. In the common notation for intensions, $[[\varphi]]_w = 1$ indicates that ψ is true in the possible world w, and $[[\varphi]]_w = 0$ indicates that φ is false in that world. Consider Table 1 and notice that reading across the rows of the table that concern φ shows concretely an example of an intension: the set of the eight ordered pairs whose first element is a world and second element is the indicated truth or falsity of φ in that world.

Clearly any propositional atom is possible—each is true in exactly half of the possible worlds. On the other hand, no propositional atom is necessary, because all are false in half of the possible worlds. So is any proposition necessary? Yes, for example, $(\varphi \to (\psi \to \chi)) \to ((\varphi \to \psi) \to (\varphi \to \chi))$ is true in all of the possible worlds, not surprisingly, because it is a (propositional) tautology. Now the following question rises rather naturally: are there any non-tautological necessary truths? It turns out that there are none in this system.

In scientific theory building, we are interested in contingent (non-tautological) necessities, those that can be shown to be false. To obtain these we need to introduce an accessibility relation, $\mathscr{R}$, on the set of possible worlds. The choice of accessibility relation(s) lies at the core of defining a modal logic for substantive theory, as we explain below. Given a choice of an accessibility relation, we relativize necessity and possibility to given possible worlds as depending on truth values in the worlds accessible from them. This means admitting that certain propositions are necessarily true under some circumstances but not under others.

The definition of the truth of a modal proposition works as follows. *A propo-*

sition of the form $\Box\varphi$ is true in the world w just in case the proposition φ is true in all the worlds w' that are accessible from w, those that satisfy $\mathscr{R}(w, w')$.

Different choices of an accessibility relation (different assumptions about its algebraic properties) define interestingly different modalities. Take, for example, the property of reflexivity: the assumption that $\forall w[\mathscr{R}(w, w)]$. If every possible world is accessible from itself, then any proposition that is necessary in a world w (and therefore true in all the possible worlds that are accessible from w) must be true in w as well: $\Box\varphi \rightarrow \varphi$. Reflexivity of the accessibility relation defines the semantic models for alethic logics; the lack of reflexivity characterizes deontic logics.[7] We argue below that imposing reflexivity for belief attitudes leads to substantively implausible implications.

Note that the lack of reflexivity differs from irreflexivity. The absence of reflexivity means only that there might be worlds that are not accessible from themselves, while irreflexivity requires that no world is accessible from itself. Irreflexivity obviously imposes a stronger requirement. In the case of beliefs, imposing this property means postulating that beliefs are always mistaken about any world (because no world is accessible to itself). Such faulty beliefs surely occur, but this is not the only possibility our model should accommodate.

Similarly, transitivity of accessibility relation has interesting consequences. Suppose that $\forall w, w', w'' [\mathscr{R}(w, w') \wedge \mathscr{R}(w', w'') \rightarrow \mathscr{R}(w, w'')]$. The transitivity assumption implies that $\Box\varphi \rightarrow \Box\Box\varphi$. This is easy to show by showing that assuming the opposite leads to a contradiction.[8]

To get deontic logics we need to exclude isolated worlds, those from which no other world is accessible. (In such a world every proposition is necessary, but no proposition is possible.) The property of the accessibility relation called seriality rules out isolated worlds: $\forall w \exists w' [\mathscr{R}(w, w')]$. Now if $\Box\varphi$ is true in w, then φ is true in all w' for which $\mathscr{R}(w, w')$ holds and there exists such a w'. In other words, there is a w' for which $\mathscr{R}(w, w')$ holds and φ is true in w', which is the truth condition for $\Diamond\varphi$ being true in w.

We now move from these very general considerations to discussion of the

---

[7] To define deontic logics one also needs to impose seriality—see below.

[8] For the sake of the argument, assume that in the world w $\Box\varphi$ is true and $\Box\Box\varphi$ is false. Since $\Box\Box\varphi$ is false in w, there must be a w' such that $\mathscr{R}(w, w')$ and $\Box\varphi$ is false in w'. However if $\Box\varphi$ is false in w', then there exists a w'' such that $\mathscr{R}(w', w'')$ and φ is false in w''. Now, due to transitivity we have $\mathscr{R}(w', w'')$, that is $\Diamond\neg\varphi$ is true in w, which contradicts the fact that $\Box\varphi$ is true in w.

modalities needed for our substantive applications.

3 Belief Attitudes for Perception, Default, and Belief

We refer to an agent's information state about a factual situation as a set of beliefs. Beliefs depend upon both perceptions and defaults. Specifically we define two basic modalities, for perception and default, and a third modality, for belief, that builds on them.

Perception is generally partial. Although any well-formed proposition must be either true or false in reality, the best one can say about perception is that it makes some propositions true, others false, and leaves open the truth/falsity of others. This kind of partiality means that even factually correct perception cannot identify exactly what the actual world looks like. Even if everything that is perceived is true, certain facts might not be observed. At best, perception can achieve an accurate *circumscription* of the actual world. Perception can also be inaccurate. In other words, perception can paint a picture of a world, but it is not the real world—just a different possible world. For these reasons, it makes sense to regard perception as providing a set of possible worlds (and the real world might even be absent from the set due to faulty perception).

These arguments lead to the conclusion that, if perception-based behavior differs from fact-based behavior, then it is useful to construct different possible worlds and specify the relations among them. Note that the accessibility relation between possible worlds and the fact that perceptions (as well as default and beliefs) are relative to the actual world provides a useful way to express the context dependence of these belief attitudes.

As perceptions are partial, sometimes an agent perceives neither φ nor its negation, $\neg\varphi$. Because the partiality of perception generates uncertainty, it is natural that mechanisms sometimes emerge that eliminate gaps. Default assumptions turn out to be useful for filling gaps in perception. If an agent lacks perceptual evidence about the value of a relevant fact but does have an applicable default, then she uses the default to "fill in" the missing facts.

Defaults shape beliefs only in the absence of a current perception of the facts in question. Thus defaults are sticky: the first applied assumption defines the belief. On the other hand, audience members recognize whether they assumed a fact or perceived it directly, at least in the past. Although beliefs shape behavior, beliefs based on taken-for-granted assumptions get exposed to revision due to direct perceptions that conflict with the assumed facts. The temporal order of perceptions matters. In case of conflict, more recent perceptions replace older ones. In this respect perceptions do not have the same sort of stickiness as defaults.

The applications considered by Hannan et al. (2007) and the one analyzed in this paper concern producers/products and audiences in domains. The agents care about the values of certain features of the producers or products in their focal domain. So the language has to contain atomic formulas of the following form: $f(x,t) = v$, where f stands for a (relevant) feature, x refers to a product or producer, t is a time point, and v is a value of f. We use the classical logical constants, such as $\wedge, \vee, \rightarrow, \leftrightarrow, \exists, \forall$, and $=$ in the usual manner. We introduce three new logical constants that are defined for an (arbitrary) audience member y and a sentence (formula) φ. We use the following notation for these new logical constants:

$\boxed{\mathrm{P}}_y\, \varphi$ stands for "The agent y perceives that φ is the case."

$\boxed{\mathrm{D}}_y\, \varphi$ stands for "The agent y takes for granted that φ is the case."

$\boxed{\mathrm{B}}_y\, \varphi$ stands for "The agent y believes that φ is the case."

In what follows we often omit the lower indexes if this simplification does not lead to any confusion.

So far, we have offered an informal description of how the three belief attitudes relate to one another. To make sure that all relevant parts of the desired imagery are in the picture, we need to provide a model that characterizes the inferential behavior of these belief operators in terms of a systematic semantics. That is, we define a logical model for the language containing these operators.

In building a model, we seek to satisfy the following constraints:

1. Perception is partial at all time points.

2. Beliefs must be grounded in either perception or taken-for-granted assumptions.

3. As seeing is believing, perception (at least temporarily) overrides earlier beliefs.

4. Defaults shape beliefs (unless there is perceptual evidence to the contrary).

5. Beliefs come from perceptions or defaults.

6. Lasting beliefs develop if lasting taken-for granted assumptions are not contradicted by perceptual evidence.

Accessibility Relations

To model the three belief attitudes we introduce three accessibility relations: $\mathscr{R}^{\mathrm{P}}$, $\mathscr{R}^{\mathrm{D}}$, and $\mathscr{R}^{\mathrm{B}}$. We denote the full set of possible worlds by W.

These accessibility relations must share some properties (introduced above in sketching modal logic) to meet all the stated desiderata. First, isolated worlds must be ruled out. In the case of perception, isolation would mean that agents perceive every proposition to be true, in contradiction with the partiality of perception. In isolated worlds, taken-for-granted assumptions are useless too, because absurd (contradictory) feature values are taken for granted. In the case of beliefs some might find this situation ideal until it is recognized that the negations of propositions are propositions too, so beliefs would be inconsistent. Therefore, we postulate seriality for each accessibility relation (for any agent).

Postulate 3.1 (Seriality)**.** *there are no isolated worlds.*

$$\forall \mathscr{R}, w\,[(\mathscr{R} \in \{\mathscr{R}^{\mathrm{P}}, \mathscr{R}^{\mathrm{B}}, \mathscr{R}^{\mathrm{D}}\}) \wedge (w \in W) \rightarrow \exists w'\,[(w' \in W) \wedge \mathscr{R}(w, w')]].$$

A second important general consideration argues that we rule out reflexivity for the accessibility relations to avoid imposing the restriction that whatever is perceived is true: $\boxed{\mathrm{P}}\varphi \rightarrow \varphi$. This would be a ludicrous constraint on perception. The damage of allowing reflexivity for the two other modalities would also be considerable. In case of defaults, it would imply that taken for grantedness yields satisfaction. To avoid these highly undesirable consequences we add the following postulate.

Postulate 3.2 (Lack of reflexivity)**.** *The accessibility relations are not reflexive.*

$$\forall \mathscr{R}\,\exists w\,[(\mathscr{R} \in \{\mathscr{R}^{\mathrm{P}}, \mathscr{R}^{\mathrm{P}}, \mathscr{R}^{\mathrm{D}}\}) \wedge (w \in W) \wedge \neg\mathscr{R}(w, w)].$$

For our model to make sense, the three modalities cannot be independent. For example, we required that any perception leads to belief. To guarantee this we need the following postulate:

Postulate 3.3 (Seeing (perceiving) is believing)**.** *The worlds that are belief-accessible (from some unspecified world) are among the perception-accessible worlds (from that world).*

$$\mathscr{R}^{\mathrm{B}} \subseteq \mathscr{R}^{\mathrm{P}}.$$

It might be helpful to explain the reasoning behind the formal statement of this postulate. Suppose that this postulate is false (for some focal world) for the proposition φ. The negation of the formula expression Postulate 3.3 would mean that the φ is true in all of the perception-accessible worlds but false in at least one belief-accessible world. In such a case, seeing is not believing; and we want to rule out this possibility.

Our next postulate guarantees that available defaults fill the gaps in a set of beliefs.

Postulate 3.4. *Agents who have available defaults use them to fill what would otherwise be gaps in their beliefs.*

Let $[[\varphi]]$ *denote the set of all the possible worlds where* φ *is true (according to the agent's defaults).*

$$\forall\varphi, w[\forall w'\,[\mathscr{R}^{\mathrm{D}}(w,w') \to (w' \in [[\varphi]])] \wedge \exists w'\,[\mathscr{R}^{\mathrm{B}}(w,w') \wedge (w' \not\in [[\varphi]])] \\ \to \forall w'\,[\mathscr{R}^{\mathrm{P}}(w,w') \to (w' \not\in [[\varphi]])]].$$

The first sub-formula $\forall w'\,[\mathscr{R}^{\mathrm{D}}(w,w') \to (w' \in [[\varphi]])]$ expresses that the formula φ is true in all worlds default-accessible from w. That is, φ is a default in w. The second sub-formula $\exists w'\,[\mathscr{R}^{\mathrm{B}}(w,w') \wedge (w' \not\in [[\varphi]])]$ expresses that φ is false in at least one possible world that is belief-accessible from w. In other words, the agent does not believe φ in w. The postulate states these two conditions jointly imply $\forall w'\,[\mathscr{R}^{\mathrm{P}}(w,w') \to (w' \not\in [[\varphi]])]$, that is, that φ is false in all worlds perceptually accessible from w, that the agent does not perceive φ in w. So the formal statement of the postulate says that the only way that a proposition can be a default yet not be believed to be true is that the agent perceives it to be false. Alternatively, if the agent does not perceive that the proposition is false and treats the truth of the proposition as a default, then the agent believes that the proposition is true. Defaults fill in gaps in perception (unless there is perceptual evidence to the contrary).

Postulate 3.5. *Anything that an agent believes as long as there is no contrary perceptual evidence is a default.*

$$\forall\varphi, w[\forall w'\,[\mathscr{R}^{\mathrm{B}}(w,w') \to (w' \in [[\varphi]])] \wedge \exists w'\,[\mathscr{R}^{\mathrm{D}}(w,w') \wedge (w' \not\in [[\varphi]])] \\ \to \forall w'\,[\mathscr{R}^{\mathrm{P}}(w,w') \to (w' \in [[\varphi]])]]$$

where we use the $[[\varphi]]$ *notation as above.*

The first sub-formula, $\forall w' [\mathscr{R}^{\mathrm{B}}(w, w') \rightarrow (w' \in [\![\varphi]\!])]$, expresses that the formula φ is true in all of the worlds that are belief-accessible from w. That is, the agent believes φ in w. The second sub-formula, $\exists w' [\mathscr{R}^{\mathrm{D}}(w, w') \wedge (w' \notin [\![\varphi]\!])]$, expresses that φ is false at least one possible world that is default-accessible from w. In other words, the agent does not take φ for granted in w. The postulate states these two conditions jointly imply $\forall w' [\mathscr{R}^{\mathrm{P}}(w, w') \rightarrow (w' \in [\![\varphi]\!])]$. That is, φ is true in all worlds perceptually accessible from w, which means that the agent does perceive φ in w. So the formal statement of the postulate says that the only way that a proposition can be believed yet not be a default is that the agent perceives it to be true.

Recall that we want to eliminate from the model those beliefs that have nothing to do with perceptions or taken-for-granted assumptions. We do so by introducing the following postulate.

Postulate 3.6. *Worlds that are both perceptually accessible and default accessible are also belief accessible.*

$$\mathscr{R}^{\mathrm{D}} \cap \mathscr{R}^{\mathrm{P}} \subseteq \mathscr{R}^{\mathrm{B}}.$$

Figure 1 illustrates the joint effects of Postulates 3.3 and 3.6.

The foregoing postulates define the elements of what is generally called a Kripke frame. In this case, it is a frame for beliefs.

Definition 3.1 (Belief frame)**.**

Let W be the set of possible worlds (for a propositional language) and $\mathscr{R}^{\mathrm{P}}$, $\mathscr{R}^{\mathrm{D}}$, and $\mathscr{R}^{\mathrm{B}}$ be the accessibility relations (for an unspecified agent). We call

$$\langle W, \mathscr{R}^{\mathrm{P}}, \mathscr{R}^{\mathrm{D}}, \mathscr{R}^{\mathrm{B}} \rangle,$$

the belief frame of the agent if the accessibility relations $\mathscr{R}^{\mathrm{P}}$, $\mathscr{R}^{\mathrm{D}}$, and $\mathscr{R}^{\mathrm{B}}$ satisfy Postulates 3.1–3.6.

With this definition of the belief frame, we can derive a set of theorems that characterize the interactions among the modalities and that relate to the desiderata that we spelled out above. The five items on our list of desiderata are logical consequences of the Kripke belief frame.

Theorem 3.1. *Perception is partial.*

$$\forall y \exists \varphi, t, x [\neg \boxed{\mathrm{P}}_y \varphi(x, t)].$$

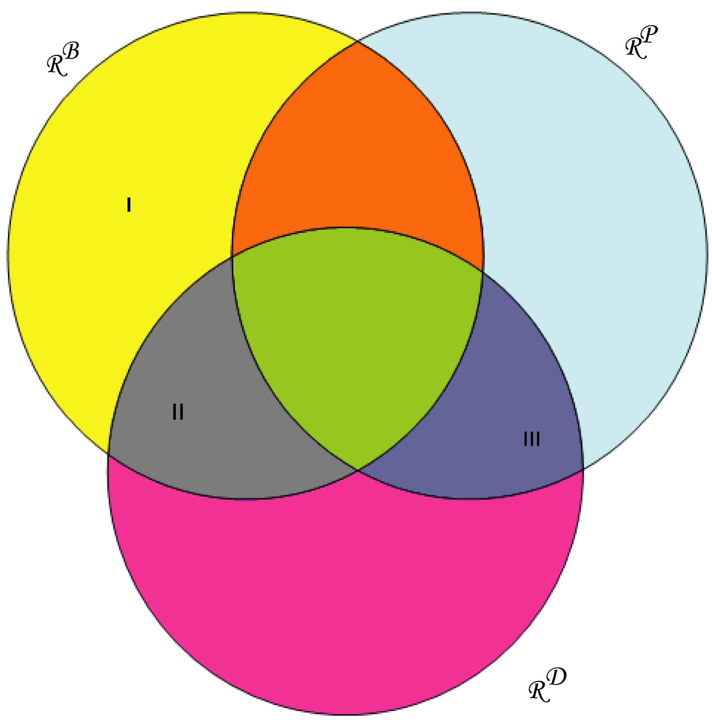

Figure 1: Relationships among the accessibility relations for perception, default, and belief according to Postulates 3.3 and 3.6. Regions I and II are empty according to Postulate 3 and Postulate 5 makes III empty.

Proof. According to Postulate 3.1 (Seriality),

$$\forall w\,[(w \in W) \to \exists w'\,[(w' \in W) \wedge \mathscr{R}^{\mathrm{P}}(w, w')]].$$

We can use the perception-accessible world w' to build φ in a way that makes the theorem true. Two cases need to be considered. First suppose that the state w' is such that some atomic propositions are true (this would allow the first seven of the worlds in the example given earlier in Table 1, but not the eighth). Take the disjunction of the negations of all of these atomic formulas. Because the resulting formula is false (by construction) in w', it cannot be the case that the agent perceives that this formula is true in w. This follows from the definition of the truth of a modal operator (a proposition is true in a world just in case it is true in all of the worlds accessible from that world). In the case of the perception modality, the proposition $\boxed{\mathrm{P}}_y\,\varphi(x,t)$ is true in a world w just in case $\boxed{\mathrm{P}}_y\,\varphi(x,t)$ is true in all of the worlds that are perception-accessible from w.

In the second case, no atomic proposition is true in w'. Then the analysis just presented holds for the conjunction of the (false) atomic propositions. □

Theorem 3.2. *Beliefs are grounded in either perception or taken-for-granted assumptions.*

$$\boxed{\mathrm{B}}_y\,\varphi \to \boxed{\mathrm{D}}_y\,\varphi \vee \boxed{\mathrm{P}}_y\,\varphi.$$

Proof. Suppose this theorem is false: in a world w, $\boxed{\mathrm{B}}_y\,\varphi \wedge \neg\boxed{\mathrm{D}}_y\,\varphi \wedge \neg\boxed{\mathrm{P}}_y\,\varphi$. Then φ is true in all the $\mathscr{R}^{\mathrm{B}}$ alternatives of w; but it is false in some $\mathscr{R}^{\mathrm{P}}$ and in some $\mathscr{R}^{\mathrm{D}}$ alternatives. Because the $\mathscr{R}^{\mathrm{P}}$ and the $\mathscr{R}^{\mathrm{D}}$ alternatives are among the $\mathscr{R}^{\mathrm{B}}$ alternatives (due to Postulate 3.5), this cannot happen. □

Theorem 3.3. *Perception overrides beliefs.*

$$\boxed{\mathrm{P}}_y\,\varphi \to \boxed{\mathrm{B}}_y\,\varphi.$$

Proof. Suppose this theorem is false: $\boxed{\mathrm{P}}_y\,\varphi \wedge \neg\boxed{\mathrm{B}}_y\,\varphi$ is true in a world w. Then φ is true in all the $\mathscr{R}^{\mathrm{P}}$ alternatives of w but false in some $\mathscr{R}^{\mathrm{B}}$ alternatives. But the $\mathscr{R}^{\mathrm{B}}$ alternatives are among the $\mathscr{R}^{\mathrm{P}}$ alternatives, which violates Postulate 3.3. □

Theorem 3.4. *Defaults shape beliefs (absent perceptual evidence to the contrary)*

$$\boxed{\mathrm{D}}_y\,\varphi \wedge \neg\boxed{\mathrm{P}}_y\,\neg\varphi \to \boxed{\mathrm{B}}_y\,\varphi.$$

Proof. This theorem is false in w if and only if $\boxed{\mathrm{D}}_y\,\varphi \wedge \neg \boxed{\mathrm{P}}_y\,\neg\varphi \wedge \neg \boxed{\mathrm{B}}_y\,\varphi$ is true in w, which, in turn, requires that φ is true in all the $\mathscr{R}^{\mathrm{D}}$ alternatives of w but false in all $\mathscr{R}^{\mathrm{B}}$ alternatives of w and true in some $\mathscr{R}^{\mathrm{D}}$ alternatives of w. If φ is true in all of the $\mathscr{R}^{\mathrm{D}}$ alternatives to w and is false in all or some $\mathscr{R}^{\mathrm{B}}$ alternatives of w, then the antecedent of Postulate 3.4 is satisfied. So we get

$$\forall w' \, [\mathscr{R}^{\mathrm{P}}(w, w') \to w' \not\in [[\varphi]]].$$

Although this does not contradict the fact that φ true in some $\mathscr{R}^{\mathrm{P}}$ alternatives of w (because the antecedent might not be satisfied), a contradiction can reached if we take account of the seriality of $\mathscr{R}^{\mathrm{P}}$, which excludes the possibility that

$$\forall w' \, [\mathscr{R}^{\mathrm{P}}(w, w') \to w' \not\in [[\varphi]]]$$

is vacuously true (because the antecedent is false). □

Theorem 3.5. *Beliefs originate from either defaults or perceptions.*

$$\boxed{\mathrm{B}}_y\,\varphi \wedge \neg \boxed{\mathrm{D}}_y\,\varphi \to \boxed{\mathrm{P}}_y\,\varphi.$$

Proof. The proof of this theorem analogous to the proof of Theorem 3.4, but uses Postulate 3.6 instead of Postulate 3.4. This theorem is false in w if and only if $\boxed{\mathrm{B}}_y\,\varphi \wedge \neg \boxed{\mathrm{D}}_y\,\varphi \wedge \neg \boxed{\mathrm{P}}_y\,\varphi$ is true in w, which, in turn, requires that φ is true in all the $\mathscr{R}^{\mathrm{B}}$ alternatives of w but false in some $\mathscr{R}^{\mathrm{D}}$ alternatives of w and also false in some $\mathscr{R}^{\mathrm{P}}$ alternatives of w. If φ is true in all the $\mathscr{R}^{\mathrm{B}}$ alternatives of w and false in some $\mathscr{R}^{\mathrm{D}}$ alternatives of w, the antecedent of Postulate 3.6 is satisfied. So we get

$$\forall w' \, [\mathscr{R}^{\mathrm{P}}(w, w') \to w' \in [[\varphi]]].$$

This does not contradict the fact that φ false in some $\mathscr{R}^{\mathrm{P}}$ alternatives of w. □

Notation. Let $\varphi(x, t)$ be a formula, x an object, and t a temporal parameter. (In our substantive application below, x denotes a producer and φ denotes one of its feature values at the time point t.) In the case of the belief modality,

$\exists\, t' \,\forall\, t'' \,[(t' \leq t'' < t) \to \boxed{\mathrm{B}}_y\,\varphi(x, t'')]$ will be abbreviated as $\overleftarrow{\boxed{\mathrm{B}}_y}\,\varphi(x, t)$;

$\exists\, t' \,\forall\, t'' \,[(t < t'' \leq t') \to \boxed{\mathrm{B}}_y\,\varphi(x, t'')]$ will be abbreviated as $\overrightarrow{\boxed{\mathrm{B}}_y}\,\varphi(x, t)$.

The notation is exactly parallel for the other two modalities. When doing so does not cause confusion, we refer to φ or $\overrightarrow{\varphi}(t)$, etc. rather than use the full notation in describing the language.

Theorem 3.6. *Lasting beliefs develop if lasting taken-for-granted assumptions are not contradicted by perceptual evidence.*

$$\neg \overleftarrow{\boxed{\mathrm{B}}_y}\, \varphi(t) \rightarrow (\overrightarrow{\boxed{\mathrm{B}}_y}\, \varphi(t) \longleftrightarrow (\overrightarrow{\boxed{\mathrm{D}}_y}\, \varphi(t) \wedge \neg \overrightarrow{\boxed{\mathrm{P}}_y} \neg \varphi(t))).$$

Proof. Let $\overrightarrow{\boxed{\mathrm{D}}_y}\, \varphi(t)$ and $\neg \overrightarrow{\boxed{\mathrm{P}}_y} \neg \varphi(t)$ be the case. By the definition of the over-arrow notation, $\overrightarrow{\boxed{\mathrm{D}}_y}\, \varphi(t)$ means

$$\exists\, t_1\, \forall\, t_1'\, [(t < t_1' < t_1) \rightarrow \boxed{\mathrm{D}}_y\, \varphi(x, {t_1}')],$$

and $\neg \overrightarrow{\boxed{\mathrm{P}}_y} \neg \varphi(t)$ means

$$\exists\, t_2\, \forall\, t_2'\, [(t < t_2' \leq t_2) \rightarrow \neg \boxed{\mathrm{P}}_y \neg \varphi(x, t_2')].$$

Let t_* denote the smaller of t_1 and t_2: $t_* = \min\{t_1, t_2\}$. With this definition of t_*, it is true that

$$\forall\, t_*'\, [(t < t_*' < t_*) \rightarrow \boxed{\mathrm{D}}_y\, \varphi(x, t_*') \wedge \neg \boxed{\mathrm{P}}_y \neg \varphi(x, t_*')].$$

This formula, taken together with Theorem 3.4, yields

$$\forall\, t_*'\, [(t < t_*' < t_*) \rightarrow \boxed{\mathrm{B}}_y\, \varphi(x, t_*')],$$

i.e., $\overrightarrow{\boxed{\mathrm{B}}_y}\, \varphi(x, t)$. □

To elaborate on the nature of defaults and perception we consider under what circumstances would agents abandon default, stop to take for granted compliance with some assumptions. It seems natural to stipulate that just before a default is abandoned the agents perceives violations instead of compliance with the expectations that are no longer taken for granted. If agent y has the default $\boxed{\mathrm{D}}_y\, \varphi(t_0)$, she abandons it at t_1 if and only if $t_1 = inf\{t | (t_0 < t) \wedge \neg \boxed{\mathrm{D}}_y\, \varphi(t)\}$. Now we propose this condition as an axiom.[9].

[9]Postulate and axiom are often used as synonyms. But we use these terms in parallel to indicate the following difference: Postulates describe, in terms of the accessibility relations, what a belief frames look like. However some of their properties related to the interaction between modal and temporal considerations are too complex for us to offer easily understandable postulates that support to these properties. So we propose two axioms that describe the interaction between the modal and temporal constructions.

Axiom 3.1. *When an agent abandons a default, she does not perceive compliance with the (formerly taken for granted) expectation.*

$$\boxed{\mathrm{D}}_y\,\varphi(x,t_0)\wedge(t_0<t)\wedge\neg\boxed{\mathrm{D}}_y\,\varphi(x,t)\rightarrow$$
$$\exists\epsilon,t_*[(0<t_*)\wedge(|t_*-\inf\{t|(t_*<t)\wedge\neg\boxed{\mathrm{D}}_y\,\varphi(x,t)\}|<\epsilon)\rightarrow\neg\boxed{\mathrm{P}}_y\,\varphi(x,t_*)]\}.$$

We add the notion that perception is episodic and show that adding this consideration gives a further useful characterization of the interaction of the modalities.

Axiom 3.2. *Continuous scrutiny is not possible over any interval that contains time points where a certain fact is not perceived.*

$$\neg\overrightarrow{\boxed{\mathrm{P}}}_y\,\varphi(t)\quad\text{and}\quad\neg\overleftarrow{\boxed{\mathrm{P}}}_y\,\varphi(t).$$

With these axioms we can also prove the converse of Theorem 3.5, so we offer the following summary:

Theorem 3.7. *Lasting beliefs develop if and only if lasting taken-for-granted assumptions are not contradicted by perceptual evidence.*

$$\overrightarrow{\boxed{\mathrm{D}}}_y\,\varphi(t)\wedge\neg\overrightarrow{\boxed{\mathrm{P}}}_y\,\neg\varphi(t)\longleftrightarrow\overrightarrow{\boxed{\mathrm{B}}}_y\,\varphi(t).$$

Proof. Theorem 3.5 grants the $\rightarrow$ direction. So only the $\leftarrow$ needs to be proven. Now suppose $\overrightarrow{\boxed{\mathrm{B}}}_y\,\varphi(x,t)$. Then

$$\exists t_1\forall t_2[(t<t_2\le t_1)\rightarrow\boxed{\mathrm{B}}_y\,\varphi(x,t_2)].$$

Due to Postulate 3.5 we have $\forall t[\boxed{\mathrm{B}}_y\,\varphi(x,t)\rightarrow\boxed{\mathrm{D}}_y\,\varphi(x,t)\vee\boxed{\mathrm{P}}_y\,\varphi(x,t)]$.

If $\forall t_1[\boxed{\mathrm{D}}_y\,\varphi(x,t_1)]$, then the theorem is proven. A proof by contraction in this case shows that $\exists t[\neg\boxed{\mathrm{D}}_y\,\varphi(x,t_1)]$ is inconsistent with $\overrightarrow{\boxed{\mathrm{B}}}_y\,\varphi(x,t)$. From Axiom 3.2 we can conclude that $\exists t_2[(t_2<t_1)\wedge\boxed{\mathrm{D}}_y\,\varphi(x,t_2)]$, so we can apply Axiom 3.1.

$$\exists\epsilon\forall t_*[(t_2<t_*)\wedge(0<t_*-\inf\{t|t\le t_1\wedge\neg\boxed{\mathrm{D}}_y\,\varphi(x,t)\}|<\epsilon)\rightarrow\neg\boxed{\mathrm{P}}_y\,\varphi(x,t_*)].$$

Due to the definition of the infimum,

$$\exists\epsilon'\exists t^*[(t_2<t^*)\wedge(0<t^*-\inf\{t|t\le t_1)\wedge\neg\boxed{\mathrm{D}}_y\,\varphi(x,t)\}|<\epsilon'\wedge\neg\boxed{\mathrm{D}}_y\,\varphi(x,t^*)].$$

Both $\neg\boxed{\mathrm{D}}_y\,\varphi(x,t^*)$ and $\neg\boxed{\mathrm{P}}_y\,\varphi(x,t^*)$ are satisfied for the time point t^*, which in turn makes $\neg\boxed{\mathrm{B}}_y\,\varphi(x,t^*)$ the case. □

Theorem 3.8. *Perception can generate persistent beliefs only via changes of the taken for granteds.*

$$\overrightarrow{\boxed{\mathrm{B}}_y}\,\varphi(t) \to \overrightarrow{\boxed{\mathrm{D}}_y}\,\varphi(t);$$

and similarly

$$\overleftarrow{\boxed{\mathrm{B}}_y}\,\varphi(t) \to \overleftarrow{\boxed{\mathrm{D}}_y}\,\varphi(t).$$

Proof. The $\overrightarrow{\boxed{\mathrm{B}}_y}\,\varphi(t) \to \overrightarrow{\boxed{\mathrm{D}}_y}\,\varphi(t)$ is only a weakening of the $\leftarrow$ direction of Theorem 3.7, so the part that has to be proven is $\overleftarrow{\boxed{\mathrm{B}}_y}\,\varphi(t) \to \overleftarrow{\boxed{\mathrm{D}}_y}\,\varphi(t)$. Because $\overleftarrow{\boxed{\mathrm{B}}_y}\,\varphi(t)$ there exists a t' such that for all t'' $t' < t'' \leq t \to \boxed{\mathrm{B}}_y\,\varphi(t'')$ In other words $\forall t''[t' < t'' < t \to \overrightarrow{\boxed{\mathrm{B}}_y}\,\varphi(t'')]$. In the proof above we showed that $\exists t_1[t'' < t_1 \leq t \wedge \neg\boxed{\mathrm{D}}_y\,\varphi(x, t_1)]$ is inconsistent with $\overrightarrow{\boxed{\mathrm{B}}_y}\,\varphi(x, t'')$, so $\forall t_1[t'' < t_1 \leq t \to \boxed{\mathrm{D}}_y\,\varphi(x, t_1)]$ what in turn implies that $\overleftarrow{\boxed{\mathrm{D}}_y}\,\varphi(t)$

□

A Dynamic Alternative: Information States

Where do the accessibility relations come from? This is not a relevant question for a purely logical approach. From this perspective, the postulates circumscribe these relations sufficiently for the validity of an inference to be ascertained. On the other hand, if one wants to look at the logical models as models of reality, as we do, this question has to be taken seriously. We think that the best way to approach this question is by following the lead of logical dynamics.

Dynamic logics analyze inference from the perspective of information states. Formulas of the language come with update conditions that describe how learning a formula changes an information state. An inference connecting a set of premises and a conclusion is valid if and only if, after updating the empty information state with all the premises, a further update with the conclusion has no effect. In other words, the conclusion is logically implied by the premises if the conclusion does not carry additional information beyond that contained in the premises.

Using this approach for our purposes requires a clear definition of the information states for belief reports. The empty information state will be a possible-world model for which any world is related to *every* other world by all three accessibility relations. Recall that such hyper-accessibility means that nothing is perceived, taken for granted, or believed. Therefore, in the model for our language that corresponds to the empty information state, the focal agent does not perceive, take for granted, or believe anything except the logical tautologies.

One simple—but misleading—view of the update process (for the belief modality, for instance) is that it works as follows. Suppose θ denotes an information state and θ gets updated with $\boxed{\mathrm{P}}_y\,\varphi$. One might imagine that the update results in an information state θ' for which all those (w, w') pairs of possible worlds for which φ is false (i.e., where $w' \not\in [[\varphi]]$) are eliminated from the extension of the $\mathscr{R}^{\mathrm{B}}$ relation.

The problem with this idea is that it does not respect the principle that modalities must be closed under logical deduction. Continuing to focus on beliefs, suppose that θ gets updated with $\boxed{\mathrm{B}}_y\,\varphi$. According to the (mistaken) view we are considering, all those (w, w') pairs of possible worlds where φ is false in w' (i.e., where $w' \not\in [[\varphi]]$) would be pruned from the extension of the $\mathscr{R}^{\mathrm{B}}$ relation. If this is all that we do, the resulting information state will not respect Theorem 3.2, which states $\boxed{\mathrm{B}}_y\,\varphi \to \boxed{\mathrm{P}}_y\,\varphi \vee \boxed{\mathrm{D}}_y\,\varphi$. Unfortunately it is not clear how we should ensure that the updated information state should respect closure under logical deduction. The conditional sentence in Theorem 3.2 says that either the antecedent is false or the consequent is true, but it does not say which. If we update the empty information state with either the negated antecedent ($\neg\boxed{\mathrm{B}}_y\,\varphi$) or the consequent ($\boxed{\mathrm{P}}_y\,\varphi \vee \boxed{\mathrm{D}}_y\,\varphi$), we do too much. The resulting information state would contain more information than is justified in either case because we do not know which alternative is the case.

The answer to this conundrum is recognizing that information states are not particular multi-modal models, but instead *sets* of multi-modal models. If we update an information state with a disjunctive piece of information such as $\boxed{\mathrm{P}}_y\,\varphi \vee \boxed{\mathrm{D}}_y\,\varphi$, then the models in the information state get updated with either $\boxed{\mathrm{P}}_y\,\varphi$ or $\boxed{\mathrm{D}}_y\,\varphi$. So, typically, there will be twice as many models in the information state after the update than before. Any subsequent update should apply to all of the models in the information state. In this manner the update process generates a tree of information states. The leaves on the tree (the terminal nodes) represent all of the different ways that the world could be (given the information added in the previous update)s. This procedure ensures that all of the terminal nodes respect all of the postulates if all update steps are followed by the adjustments required by the postulates.

It is worth noting that complete information about our focal agent's perceptions, taken for granteds, and beliefs tells also what the agent does not perceive, does not take for granted, and does not believe. Updating an information state with such negative bits of information might prune out whole branches of the tree. In the unrealistic case of an update with the complete information, the result is a single multi-modal model whose accessibility relations are fully defined (and the world described by this model is the actual world). This construction provides an indirect characterization of the accessibility relations; information collected about the perceptions, defaults, and beliefs of the agents allow us to construct the set of belief frames that depicts this (partial) characterization.

Examples of Updates for Modal Models

In what follows we illustrate how the axioms help to build information states to represent different attitude reports concerning one (focal) agent's beliefs, perceptions and default assumptions. An information state can comply with reports on the three attitudes of the focal agent in several ways. The relevant information can be seen in diagrams of how the accessibility relations related to these attitudes look with an empty information (belief) state and how the pattern changes with different kinds of updates. For this reason we offer pictures of the maximal accessibility relations that comply with updates. In each case, numerous subsets of these maxima sets are also faithful to the attitude reports.

Figure 2 contains a graphical representation of any one of three maximal accessibility relations, $\mathscr{R}^{\mathrm{B}}$, $\mathscr{R}^{\mathrm{P}}$ and $\mathscr{R}^{\mathrm{D}}$, provided that no attitude reports are available. These are the models for empty information states. In principle[10] there are $(2^8 - 1)^8$ different realizations of all three accessibility relations.[11] In these figures two-sided arrows represents two directional arrows. An arrow pointing from world w1 to world w2 represents the fact that w2 is accessible from w1 or, in other words, w2 is a belief-alternative of w1.

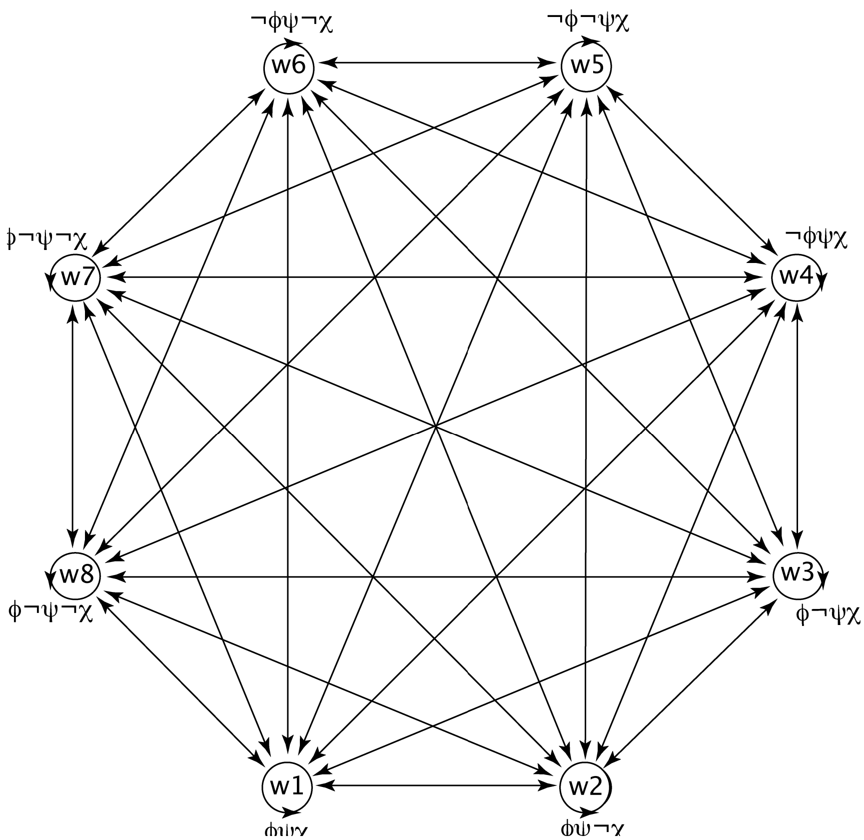

Figure 2: A graphical representation of (any) one of the three accessibility relations without any information of the attitude. In this model the relation is reflexive, so at least one reflexive arrow has to be omitted, but in the empty information state it is not known which.

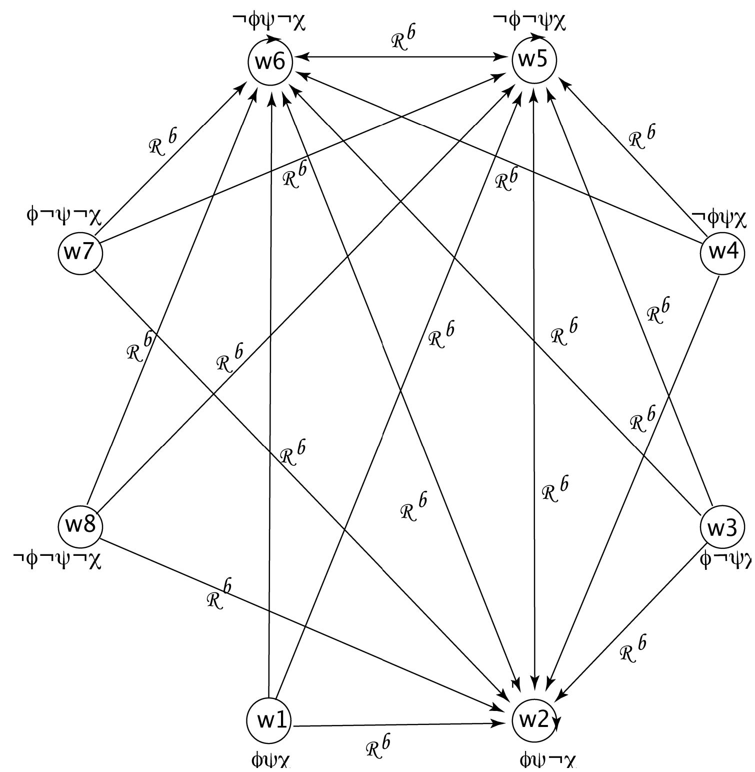

Figure 3: A graphical representation of the maximal belief accessibility relation shown in Figure 2 following the update with $\boxed{\mathrm{B}}_y\,(\psi \wedge \neg\chi) \vee (\neg\varphi \wedge \neg\psi \wedge \chi)$

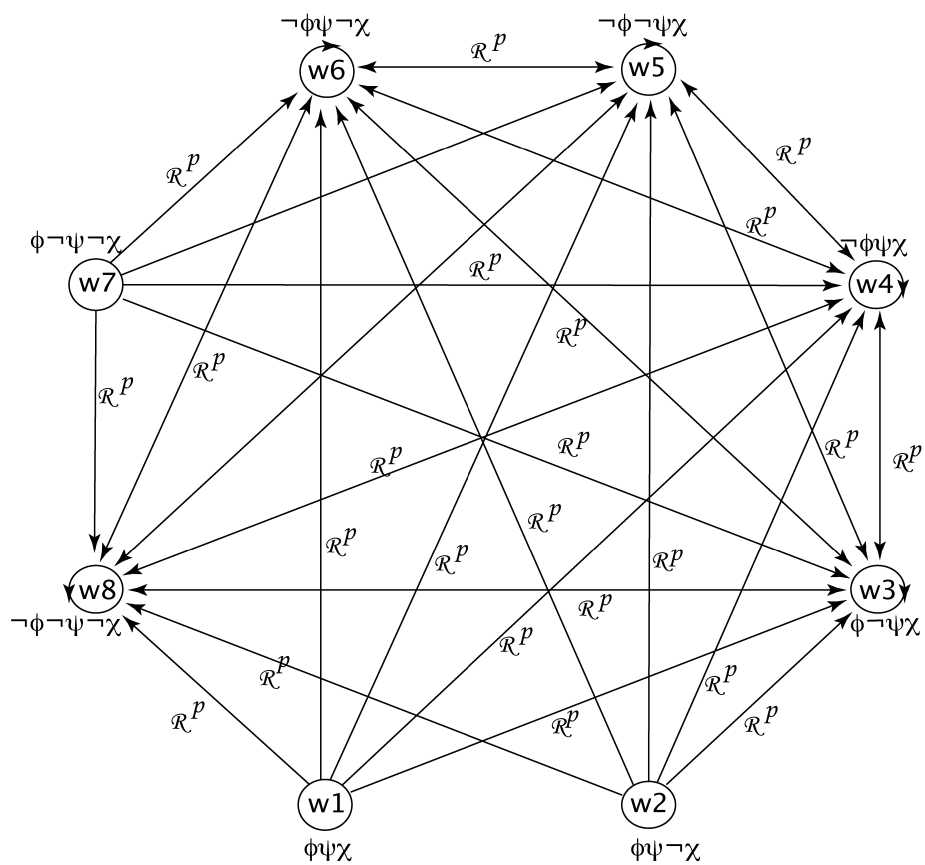

Figure 4: A graphical representation of the maximal perception accessibility relation following the update of the empty information state with $\boxed{\mathrm{P}}_y(\neg\varphi) \vee (\neg\psi \wedge \chi)$

Effect of an update with a belief report

Now suppose we learn that the agent that she believes that either $(\psi \wedge \neg\chi)$ or $(\neg\varphi \wedge \neg\psi \wedge \chi)$ is the case. We can model the maximal accessibility relation that complies with the update of her belief as in Figure 3. The still-belief-possible worlds following the update are w2, defined in Table 1 as the state of affairs $\varphi \wedge \psi \wedge \neg\chi$, and w5 $(\neg\varphi \wedge \neg\psi \wedge \chi)$, and w6 $(\neg\varphi \wedge \psi \wedge \neg\chi)$. w2 and w6 are the (only) two world that are consistent with here belief that $\psi \wedge \neg\chi$ is the case while here alternative belief that (perhaps) $\neg\varphi \wedge \neg\psi \wedge \chi$ is the case, is only satisfied by w5.

The actual accessibility relation is constrained from below too. Due the seriality condition, every possible world must have at least one (not necessarily different) world that is accessible from it. Since in the maximal relation every world has three belief-alternatives one can choose one alternative for a world (in three different ways) or two alternatives (three different ways) or perhaps all three alternatives, so for every world we have seven choices. In total $(2^3-1)^8 = 5,764,801$ realizations of the belief-accessibility relation comply with the above described attitude report.

Further update with a perception report

Next we learn that the agent perceives that $\neg\varphi \vee (\neg\psi \vee \neg\chi)$ is the case. The maximal perceptual-accessibility relation that complies with this attitude report is shown in Figure 4.

This maximal perceptual accessibility relation has also $(2^3-1)^8$ subsets that are faithful realizations of the perception report, just as for beliefs. This is a large enough number that it is time to take into account the interactions due to our axioms. From Axiom 3.3 we know that $\mathscr{R}^{\mathrm{B}} \subseteq \mathscr{R}^{\mathrm{P}}$. Clearly the maximal $\mathscr{R}^{\mathrm{B}}$ relation drawn in Figure 3 does not satisfy this condition. The relation should be an appropriate subset of the relation shown in Figure 4 instead. The change is

[10] Strictly speaking this is not true: in such a scenario all worlds are accessible from themselves so the accessibility relations are reflexive, while Postulate 3.2 requires the opposite. At least one of the arrows that point back to a world has to be removed, and there is a multiplicity of alternatives to choose from. For the sake of simplicity we do not discuss these details.

[11] Exactly eight arrows begin in each world. Any one of them might be in the accessibility relation or not, which gives a total of 2^8 possibilities. To satisfy the seriality condition at least one of these arrows have to be chosen, which gives us 2^8-1 possibilities per possible worlds, and there are 8 possible worlds.

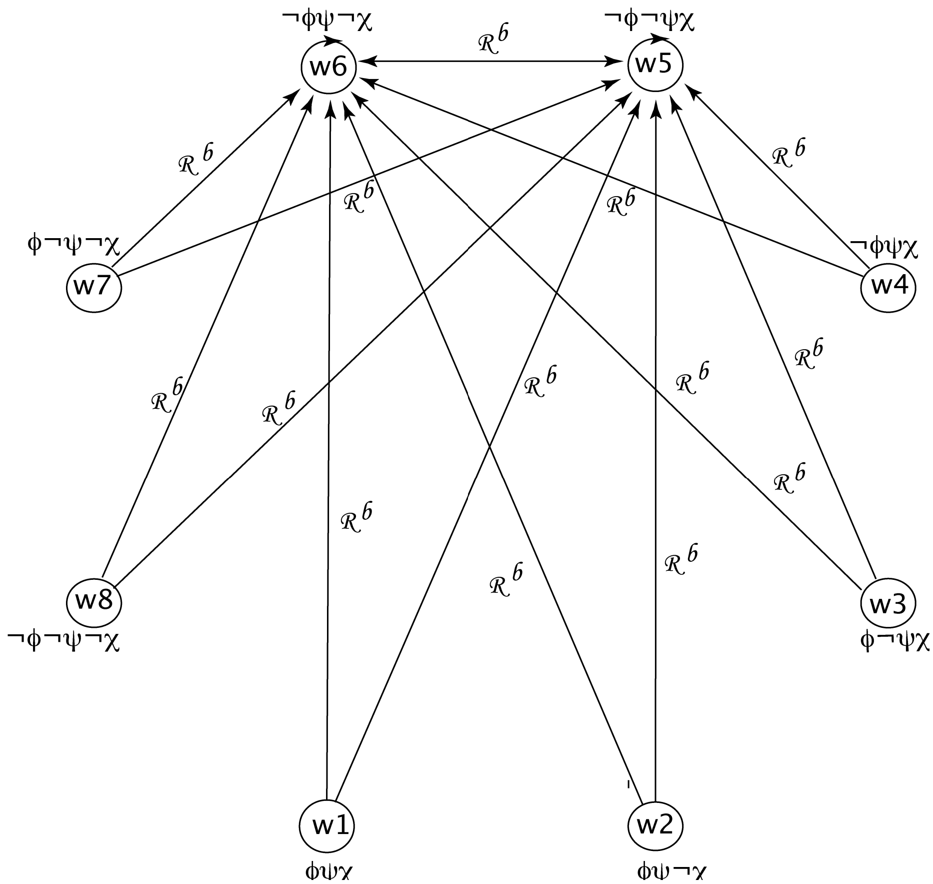

Figure 5: A graphical representation of the maximal belief accessibility relation after the pattern shown in Figure 3 is adjusted to comply with the perception report shown in Figure 4 according to the constraint stated in Postulate 3.3

rather dramatic because there are only(!) $(2^2-1)^8 = 6,561$ ways to chose the belief accessibility relation and still comply with her attitude reports. (All worlds have two belief-alternatives and we can choose either one or both.)

The impact of an update with a default

Suppose we also learn that the agent takes for granted that ψ) is the case. The maximal accessibility relation for her defaults is shown in Figure 6, only those four worlds can be default accessible which make ψ true.

Now we can take advantage of Postulate 3.5, which states that $\mathscr{R}^d \cap \mathscr{R}^p \subseteq \mathscr{R}^b$. Figure 7 depicts the intersection $\mathscr{R}^d \cap calR^p$ for the perceptions and defaults considered above. It is easy to see that one further update due to the maximal belief accessibility relation is necessary: the intersection has to be part of the belief accessibility relation.

To conclude: modeling of (typically partial) belief reports does not help us to define all three accessibility relations sharply, but thanks to our axioms we can still characterize three sets of relations that such that all elements of these sets represent consistent possible realizations of the accessibility relations. If the belief reports allow for a single accessibility relation, the definition becomes sharp. If, on the other hand, no element remains in the set of possible accessibility relations, then the agent's beliefs were inconsistent.

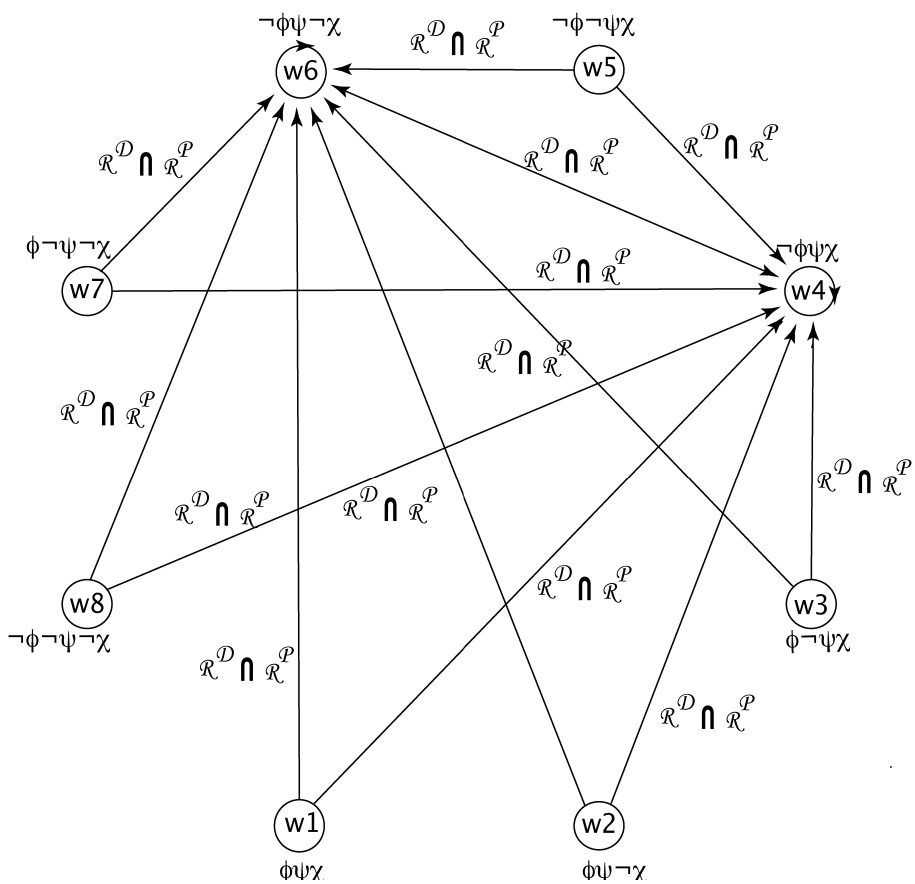

Figure 6: The intersection of the perception and the default accessibility relations from Figs. 5 and 6

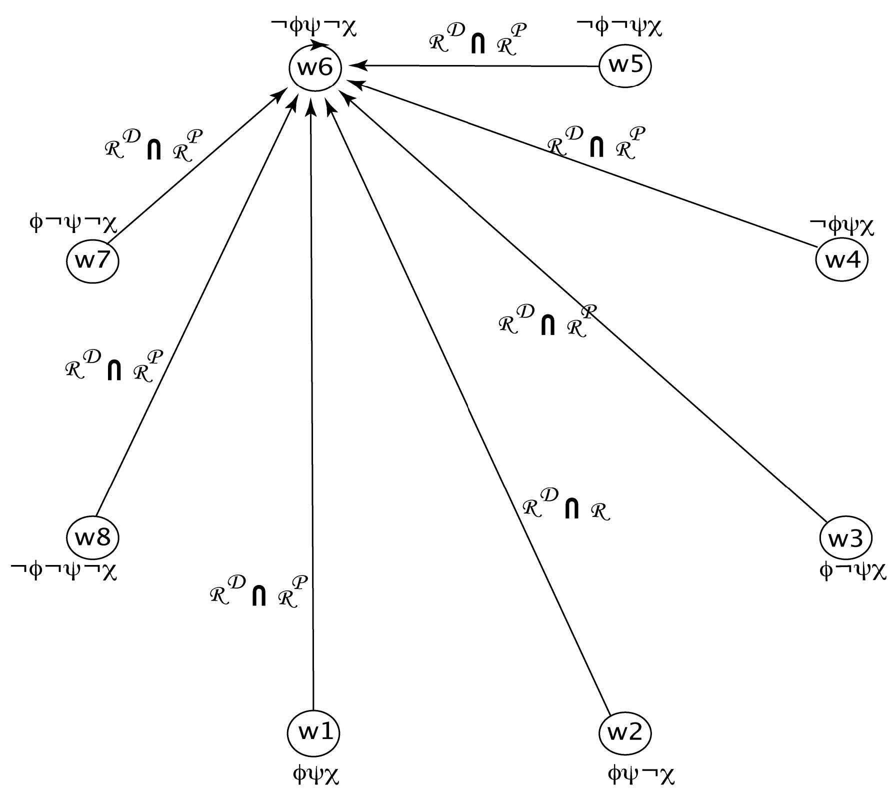

Figure 7: The intersection of the perception and the default accessibility relations after complying with postulate Postulate 3.5

4 A Sociological Application: Legitimation

Now we illustrate the deployment of these theory-building tools in a sociological argument. We build on the idea that the legitimation of a category means a heavy reliance on defaults and develop an extension of arguments based on an analysis of a single category to a multiple-category case. We tune our argument to yield implications about typecasting.

At this point we combine the different types of logical approaches discussed in the opening paragraph. For the (re)construction of the argumentation, we take advantage of the nonmonotonic logic developed by Pólos and Hannan (2002, 2004). We import some "metaphysical" and "ontological" insights from Hannan et al. (2007) who treat audiences, clusters, categories, forms, and populations of organizations (or products) as fuzzy sets and express causal claims (about categorization and legitimation) as rules with exceptions, so that the logic of argumentation is nonmonotonic. On top of these two components, we now employ the language used to describe perceptions, taken-for-granted assumptions, and beliefs.

Our treatment of the role of the audience in generating and legitimating categories builds on the conceptualization developed by Hannan et al. (2007, Chs. 2–5). The theory considers a domain as consisting of a language and a dual role structure: producer (an agent who makes "offerings" in the domain) and audience member (an agent. The basic linguistic objects are labels that audience members apply to producers. This idea is implemented in terms of a labeling function, which maps from triplets of audience members, producers, and time points to the powerset of the set of available labels.[12] We denote the set of labels that y applies to x at time t as $\mathbf{lab}(x, y, t)$. If it happens that the agent applies the label l to x at time t, then $l \in \mathbf{lab}(x, y, t)$.

Fuzziness plays a central role in the development of the language of a domain. When audience members perceive clusters of similar producers, they might apply a label to the cluster. Sometimes they decide that certain producers "deserve" a label to varying degrees, that some fully merit the label, some do not, and some merit it only partially. Following several major lines of work in cognitive psychology and cognitive science, Hannan et al. (2007) treat the extensions (memberships) of labels as fuzzy sets—meaning that membership

[12]The powerset of a set is the collection of all of its subsets. We refer the powerset here because an agent can apply multiple labels to the same object.

can be partial, a matter of degree. The extent to which an audience member regards a label as applicable to a producer is reflected in a grade of membership (GoM) function. In notation, $\mu_{e(l)}(x, y, t)$ is the grade of membership of x in the extension of the label l from y's perspective at time t. Hannan et al. (2007) equate this grade of membership in the extension of a label with the probability over occurrences (possible situations) that the audience member will apply the label to the object.

Highly engaged members of the audience often try to make sense of what accounts for partiality in the applicability of a label; and they often produce abstract representations ("theories") of the basis for their assessments of membership. They take notice that certain configurations of feature values yield full-fledged membership, others only moderate standing as a member, others a low but nonzero standing, and still others a zero grade of membership. A mental representation of such a pattern is generally called a schema. A schema for a label is a cognitive model that explains who is in, who is out, and who lies at various positions between these extremes. In other words, a schema establishes the meaning (or intension) of a label.

We represent schemas for labels as sets of formulas that pick out a set of relevant features (or relations) and distinguish the values of those features (or relations) that are consistent with membership in a label from those that are not. We need some additional notation to define schemata formally. The ordering of elements in a listing of the membership of a set is generally taken to be arbitrary. Now we fix the ordering of elements by expressing the relevant sets of features and of their values as indexed sets.[13] Let $\mathbf{f}_i = \{f_1, f_2, \ldots f_i\}$ be the indexed set of i features that are relevant for a schema. Each feature in the set has a range of possible values. We denote the set of possible values of feature f_j by $\mathbf{r}_j$ and a value for an object at a time point as $f_{j,x,t}$.

Definition 4.1 (Schema)**.** A schema for a label maps pairs of audience members and time points to an n-tuple of nonempty subsets values of the relevant features; this subset contains the schema-conforming feature values.

$$\sigma_l : \mathbf{a} \times \mathbf{t} \longrightarrow \mathscr{P}(\mathbf{r_1}) \times \cdots \times \mathscr{P}(\mathbf{r_I})$$

$$\sigma_l(y, t) = \langle \mathbf{s}_1^l, \ldots, \mathbf{s}_I^l \rangle$$

[13] Suppose we have a set $\mathbf{x} = \{x_1, x_2, \ldots x_n\}$ and a set I containing the first i natural numbers: $\mathbf{i} = \{0, 1, 2, \ldots i\}$. We can express the indexed set $\mathbf{x}_n = \{x_i \mid i \in \mathbf{i}\}$.

where a $\mathbf{s}_i^l$ is the set of all the schema conforming values of the ith feature, and I is the total number of schema-relevant features. $\sigma_l(y,t) = \sigma(l,y,t)$ is defined provided that $l \in \mathbf{lab}(k,y,t)$.

In a parallel to the grade of membership defined for extensions of labels, $\mu_{i(l)}(x,y,t)$ denotes the grade of membership of x in the "meaning" (intension) of the label l from y's perspective at time t. (Hannan et al. (2007) equate the GoM in the meaning of a label to the GoM in the schema.)

From a sociological perspective, formation of schemas matters most when the members of an audience reach a high level of agreement about the meaning of a label, about the schemas that give its meaning. This suggests that a category is a special kind of semantic object in the language of the audience, a label for which an audience achieves a high level of extensional and intensional consensus. Extensional consensus is the level of agreement within the audience about the applicability of a label to the objects in the domain. Intensional consensus is the level of agreement within an audience about the degrees to which the objects in the domain fit their schemas for a label. The predicate $\textsc{cat}(l,t)$ indicates that l is the label of a category for the audience segment at time t.

Sometimes audience members perceive that the set of objects with a high grade of membership in a label generally display feature values that fit the agent's meaning for the label. Hannan et al. (2007) argue that a belief of widespread fit with a schema creates the impression that the reality expressed by a category is natural and perhaps even obligatory. Therefore, generic fit and a low frequency of observed misfits generally cause audience members to take for granted that behaviors fit the relevant schema for those producers who bear the label. In such cases, schemas become default assumptions of everyday life. What changes is that the defaults get used to fill the many gaps in perceptions that arise from incomplete information, unobservability, and ambiguity. That is, absent evidence (perceptions) to the contrary, agents in default mode assume that they will find behavior and structure consistent with the relevant schemata whenever they encounter an organization to which they have reason to apply the label.

Legitimation and Defaults

Defaults only matter in the absence of perception, because (perceived) facts override defaults, as this modality is defined. So we consider situations in which audience members' perceptions of feature values are partial. Partial perception of fit to a schema is common, because only some relevant features can be perceived easily. Sometimes all that one sees is that a producer makes a claim to a label or that some other audience member (perhaps a critic or another kind of gatekeeper) applies the label to the producer. Hannan et al. (2007) argue that such situations of partial perception offer the analytic leverage needed to define legitimation.

The key idea is that of a test code, a partial segment of a schema for a label that an agent uses to make inferences about the fit to a schema for the values of features for which she has no beliefs (perceptions or defaults). If the agent perceives that an object's perceived feature values "pass" the test code (that is, the combination of the perceived feature values lies in the set of schema conforming combinations), then the agent induces that the unperceived/non-default values of schema relevant features also fit the schema. In technical terms, (perceived) satisfaction of a test code triggers the agent to apply the default that the unperceived/non-default feature values also satisfy the schema.

The argument is simplest in the case of "flat" schemas, those in which each feature has equal weight in determining fit. This is because restricting attention to flat schemas allows us to conceptualize legitimation in terms of the number of features with certain properties. If a code system is not flat, then we would need to take account of the weights assigned to features.

We begin by formalizing the idea of a test code for fit with a schema, which we will use to define legitimation. We need to introduce some notation for this task.

Notation. Let $\mathbf{f}_J$ denote an indexed set of values of some subset of the relevant features: $0 \leq J < I$, where I is the number of features coded in the relevant schema. We use the expression $f_{i,x,t} \in \mathbf{s}_i^l$ to represent the fact that the ith feature of the object x has a value that complies with the schema $\sigma(l,x,t)$ at the time point t.

With this notation in hand, we can state formally what it means to induce conformity to a schema based on partial perception. We depart slightly from the formulation offered by Hannan et al. (2007) (in their Definition 4.1), which holds that induction "fills in" all non-perceived feature values when a test code is perceived to be satisfied. By attending only to perception, this definition overlooks the role of existing defaults. Recall that defaults shape beliefs when there is no contrary perception. There does not appear to be any reason to think that agents will override existing defaults based only the passing of a test on other features. So we refine the earlier conception in line with this intuition. Specifically we propose that induction works on features for which the agent does not have any existing beliefs (based either on perception or default) rather than on non-perception.

Definition 4.2 (Induction from a test)**.** An induction from a test is a situation in which an agent's perception that producer bears a category label and that its feature values satisfy a test code triggers the agent to apply the default that the values of features about which there is no belief to the contrary also satisfy the schema.

Let $\sigma(l, y, t) = \langle \mathbf{s}_1^l, \ldots, \mathbf{s}_I^l \rangle$.

$$\begin{aligned} \text{INDUC}(\sigma(l,y,t), \mathbf{tc}_J) \leftrightarrow \forall i, j, x[(l \in \mathbf{lab}(x,y,t)) \wedge (j \in J) \wedge (i \in I \setminus J) \\ \wedge (\boxed{\text{P}}_y (f_{j,x,t} \in \mathbf{tc}_j^l)) \rightarrow (\neg \boxed{\text{B}}_y (f_{i,x,t} \not\in \mathbf{s}_i^l)) \leftrightarrow \overrightarrow{\boxed{\text{D}}_y} (f_{i,x,t} \in \mathbf{s}_i^l)]. \end{aligned}$$

In this case we refer to $\mathbf{tc}_J = \{\mathbf{tc}_j \mid j \in J\}$ as y's test for judging conformity to the schema $\sigma(l, y, t)$, in notation, $\text{TST}(\sigma(l, y, t), \mathbf{tc}_J)$, and we say that the test has J items. An empty test is one with $J = 0$.

How many features would need to be checked before an agent treats as a default that the rest of the features fall in a schema-consistent pattern? If every feature must be checked, then the agent takes nothing for granted. If only a small fraction of the relevant features must be checked, then defaults get used in a powerful way. Hannan et al. (2007) argue that the latter case suggests that schema conformity is taken for granted, given a small amount of positive perception (perhaps only a claim to the label). These comparisons make the most sense when we consider the minimal test for an agent-schema pair, the test that involves the smallest number of features.

Definition 4.3 (Minimal test for induction)**.** The set of feature values $\mathbf{tc}_J$ is y's a minimal test for induction for the schema for l at time t, in notation $\text{MTST}(\sigma(l, y, t), \mathbf{tc}_J)$, if and only if

1. it is one of y's tests for conformity with the schema;
2. it no more test features for the schema than any other of y's tests;
3. y induces satisfaction of the schema $\sigma(l, y, t)$ on the untested features from this test.

In the theory proposed by Hannan et al. (2007), the size of the minimal test for induction for a category relates directly to the degree of taken for grantedness of the category among relevant audiences.

Definition 4.4 (Taken for grantedness)**.** The degree to which an agent takes for granted that the untested feature values of a labeled producer conform to a schema for the label at a time point is the ratio of the untested code to the whole code. The degree to which an agent takes for granted the label at the time point, $G(l, y, t)$, is the average of taken for grantedness over the producers to which the agent assigns the label.

$$g(l,x,y,t) \equiv \begin{cases} (I-J)/I & \text{if } \langle \mathbf{s}_1^l,\ldots,\mathbf{s}_I^l\rangle \wedge \mathrm{MTST}(\sigma(l,y,t),\mathbf{tc}_J) \\ & \quad \wedge (l \in \mathbf{lab}(x,y,t)); \\ 0 & \text{otherwise;} \end{cases}$$

and

$$G(l,y,t) \equiv \textstyle\sum_{x \mid l \in \mathbf{lab}(x,y,t)} \dfrac{g(l,x,y,t)}{|\{x \mid l \in \mathbf{lab}(x,y,t)\}|}.$$

Note that I indicates the (crisp) cardinality of the set of schema-relevant features and J indicates the cardinality of the minimal test. Therefore, this definition sets $g = 0$ if the agent does not apply the label to the object or needs to see every (nonlabel) feature before making an induction (which is no induction at all); nothing is taken as satisfied by default. It sets $g = 1$ if applying the label by itself is enough to shift the agent to defaults about schema-conformity on all of the other relevant features. In this case, the test on feature values is empty, $J = 0$; and the test is passed automatically whenever the label is applied.

We are especially interested in cases in which an audience member takes satisfaction of a schema for granted generally for the bearers of a label who pass a very small test code. If such cases the schema-label pair displays a high degree of taken for grantedness. Making this kind of assessment requires that we consider the levels of taken for grantedness across the relevant producers/products in the domain at a time point. The relevant ones consist of those to whom the audience-segment member applies the label at that time point.

Definition 4.5 (Legitimation of a label for an audience)**.** The degree of legitimation of a label at a time point for an (unspecified) audience, $L(l, t)$, is the average (over the members of the audience) of the taken for grantedness of the label.

$$L(l,t) \equiv \frac{1}{|\mathbf{a}|}\textstyle\sum_{y \in \mathbf{a}} G(l,y,t).$$

Legitimation, as we define it, means that the members of the audience treat the satisfaction of their schemas for the label as defaults for those producers that pass a test code. If the test code is "thin," in the sense that it involves a small subset of the full schema, then much is taken for granted. At the extreme, simple application of a label is enough to trigger defaults on unperceived features. In such a case, a category has the standing of a taken-for-granted element of the social structure, what Hannan and Freeman (1977) called a form.

Conclusion

In this paper, we argue that modal logic can be a valuable tool for sharpening social and behavioral theories. The multi-modal language we describe contains sentential operators that are intensional as well as follow a principle of logical closure. This means the agents we model can have circumscribed (and sometimes mistaken) understandings of the world around them but are expected to follow basic patterns of deduction based on the understandings they hold. This fits well with styles of argumentation needed to express core sociological concepts such as perception, belief, and valuation.

To model perception-based (as opposed to fact-based) behavior, we treat perception as providing a set of possible worlds. By specifying accessibility relations between worlds, researchers can build detailed models of the inferential behavior of belief operators. In our substantive application, we demonstrate how this approach can be used to construct a model of perception-default-belief processes that provides a foundation for representing key processes underlying legitimation.

This application allows us to express formally what seems distinctive about membership in highly legitimated categories. Following the conceptualization developed by Hannan et al. (2007), we proposed that when a category label is highly taken for granted by an agent, the default modality applies. Producers who pass a very small test code will be automatically assigned a full grade of membership in the category even when many schema-relevant feature values are not observed.

This substantive application can be used as a basis for formalizing arguments core to other theory fragments within sociological literature. Elsewhere (Hsu, Hannan, and Pólos 2009) we demonstrate this in a companion paper that develops formal connections between three theories of categorization: typecasting, form emergence, and institutionalization. In that paper, we extend this conceptualization of category membership assignments to a multiple category context to yield dynamics central to each of these theories. In the case of the typecasting dynamic documented by Zuckerman, Kim, Ukanwa, and von Rittman (2003), we show that once an audience member accepts a producer as a full member of a highly taken-for-granted concept, then schema-conforming defaults are treated as facts when considering membership in other types or concepts. Unless and until these defaults are directly overridden by new perceptions that show the producer to violate the schema for its original label, they will continue to hold and define the identity of the producer in the eyes of the agent.

We conclude this paper by noting that, while sociologists have greater familiarity with classical first-order logic, dynamic logics (such as modal logic) can provide more realistic representations of the structure of sociological arguments. The potential benefits for theory-building are considerable. Modal constructions can be used to sharpen understanding of core concepts, as in the case of legitimation. They can be used to deepen understanding of the connections between distinct theory fragments, as in the case of typecasting and form emergence. And they can also be used to reveal new predictions, as in the case of institutionalization. By paying greater attention to the formal language used to express core concepts, researchers can set a better foundation for developing new and substantively meaningful insights in their formalization efforts.

References

Bach, Emmon W. 1986. Natural language metaphysics. In R. B. Marcus, G. J. W. Dorn, and P. Weingartner. (eds.) *Logic Methodology and Philosophy of Sciences VII. Proceeding of the Seventh International Congress.* Amsterdam: North Holland.

Copeland, B. Jack. 2002. The genesis of possible worlds semantics. *J. Phil. Logic* 31:99–137.

Etchemendy, John. 1990. *The concept of logical consequence.* Cambridge, Mass.: Harvard Univ. Press.

Gamut, L. T. F. 1991. *Logic, language, and meaning,* Volume 2: *Intensional logic and intensional grammar.* Chicago: Univ. of Chicago Press. [L. T. F. Gamut is the collective pseudonym for J. F. A. K. van Benthem, J. A. G. Groenendijk, D. H. J. de Jong, M. J. B. Stockhof, and H. J. Verkuyl.]

Hannan, Michael T. 1997. On logical formalization of theories from organizational ecology. *Soc. Methodology* 27:145–50.

——— 1998. Rethinking age dependence in organizational mortality: Logical formalizations. *Am. J. Soc.* 104:85–123.

Hannan, Michael T., László Pólos, and Glenn R. Carroll. 2007. *Logics of organization theory: Audiences, codes, and ecologies.* Princeton: Princeton University Press.

Hsu, Greta, Michael T. Hannan, and László Pólos. 2009. Typecasting and legitimation: A formal theory. Research report XXX, Stanford Graduate School of Business.

Kamps, Jaap and László Pólos. 1999. Reducing uncertainty: A formal theory of organizations in action. *Am. J. Soc.* 104:1776–1812.

Lewis, David K. 1973. *Counterfactuals.* Oxford: Blackwell.

Montgomery, James D. 2005. The logic of role theory: Role conflict and stability of the self-concept. *J. Math. Soc.* 29: 33–71.

Péli, Gábor, Jeroen Bruggeman, Michael Mausch, and Brendan O'Nualláin. 1994. A logical approach to organizational ecology: Formalizing the inertia fragment in first-order logic. *Am. Soc. Rev.* 59:571–93.

Pólos, László and Michael T. Hannan. 2002. Reasoning with partial knowledge. *Soc. Methodology* 32:133–81.

—— 2004. A logic for theories in flux: A model-theoretic approach. *Logique et Analyse* 47:85–121.

van Benthem, Johan. 1996a. Logic and argumentation theory. Pp. 18–31 in F. H. van Eemeren, R. Grootendorst, J. A. Blair, and C. A. Willard (eds.), *Perspectives and approaches: Proceedings of the third ISSA Conference on Argumentation, Vol. 1.* Amsterdam: International Center for the Study of Argumentation.

—— 1996b. *Exploring logical dynamics.* Stanford: CSLI Publications, Stanford.

Zuckerman, Ezra W., Tai-Young Kim, Kalinda Ukanwa, and James von Rittman. 2003. Robust identities or non-entities? Typecasting in the feature film labor market. *Am. J. Soc.* 108:1018–1074.