

# Believing Change and Changing Belief

Peter Haddawy \*

*haddawy@cs.uwm.edu*

Department of Electrical Engineering and Computer Science  
University of Wisconsin-Milwaukee  
Milwaukee, WI 53201

## Abstract

We present a first-order logic of time, chance, and probability that is capable of expressing the relation between subjective probability and objective chance at different times. Using this capability, we show how the logic can distinguish between causal and evidential correlation by distinguishing between conditions, events, and actions that 1) influence the agent's belief in chance and 2) the agent believes to influence chance. Furthermore, the semantics of the logic captures commonsense inferences concerning objective chance and causality. We show that an agent's subjective probability is the expected value of its beliefs concerning objective chance. We also prove that an agent using this representation believes with certainty that the past cannot be causally influenced.

## 1 Introduction

The ability to distinguish evidential from causal correlation is crucial for carrying out a number of different types of problem solving. To perform diagnosis we must be able to identify the factors that caused an observed failure in order to determine how to repair the faulty device. If we cannot distinguish causal from evidential correlation, we may end up treating the symptoms rather than the causes of the fault. When reasoning about plans, an agent may have goals that involve achieving a specified state of the world, or achieving a specified state of knowledge, or a combination of both. In order to effectively reason about such goals, we need to distinguish actions that influence the state of the world from those that only influence our state of knowledge of the world.

In this paper we extend Haddawy's [3] logic of time, chance, and action  $\mathcal{L}_{tca}$  by adding a subjective probability operator. We show how the resulting first-order logic of time, chance, and probability,  $\mathcal{L}_{tcp}$ , can distinguish between causal and evidential correlation by distinguishing between conditions and events that 1) influence the agent's belief in chance and 2) the agent believes to influence chance. Furthermore, the semantics of the logic captures some commonsense inferences

concerning causality and the relation between objective chance and subjective probability. We prove that an agent's subjective probability is the expected value of its beliefs concerning objective chance. We also prove that an agent whose beliefs are represented in this logic believes with certainty that the past cannot be causally influenced. On the other hand, an agent can execute actions that influence its subjective beliefs about the past.

## 2 Ontology

We briefly present the ontology of the logic, which includes the representation of time, facts, events, objective chance, and subjective probability. For simplicity of exposition, we will omit the representation of actions and will treat them as events. For a more detailed development of chance, facts, events, and actions see [3].

Time is modeled as a collection of world-histories, each of which is one possible chronology or history of events throughout time. A totally ordered set of time points provides a common reference to times in the various world-histories.

We represent an agent's beliefs with subjective probabilities. Since beliefs may change with time, subjective probability is taken relative to a point in time. We represent it by defining a probability distribution over the set of world-histories at each point in time. So an agent can have beliefs concerning temporally qualified facts and events.

We represent causal correlation with objective chance. Objectively, actions and events can only affect the state of the world at times after their occurrence. That is to say, at each point in time, the past is fixed—no occurrences in the world will cause it to change; but at each point in time the future might unfold in any number of ways. So relative to any point in time, only one objectively possible past exists, but numerous possible futures exist. Thus we represent objective chance by defining a future-branching tree structure on the world-histories and by defining probabilities over this tree. Like subjective probability, chance is taken relative to a point in time. By defining chance in this way, conditions in the present and past relative to a given time are either certainly true or certainly false. So actions and other events can only affect the chances of future facts and events. This property distinguishes objective chance from subjective probability. Subjectively the past can be uncertain but objectively it is completely determined.

---

\*This work was partially supported by NSF grant #IRI-9207262.

The present characterization of objective chance is not to be confused with the frequentist interpretation of probability [10, 11] which is often called objective probability. Frequentist theories define probability in terms of the limiting relative frequency in an infinite number of trials or events. The current work does not rely on relative frequencies for its semantics. Rather it models objective chance by formalizing the properties that characterize objective chance. Thus while frequentist theories have difficulty assigning meaningful probabilities to unique events like a fish jumping out of the water at a given location and time, our model has no problem in assigning non-trivial probabilities to such events. Our model of objective chance and subjective probability is motivated by the subjectivist theories of objective chance [6, 8, 9], which define chance in terms of properties that one would expect a rational agent to believe objective chance to possess.

This distinction between the frequentist theory of probability and our conception of objective chance puts the present work in sharp contrast with Bacchus’s [1] logic of statistical probabilities which models exactly relative frequency type probabilities. One telling difference between the two logics is that Bacchus’s logic  $L_p$  assigns only probability 0 or 1 to unique events (more precisely, to all closed formulas). The present logic can assign any chance value to unique events in the future, while events in the past are assigned only chance values 0 or 1, as required by our definition of objective chance.

It is reasonable to expect the subjective beliefs of a rational agent concerning objective chance to obey certain constraints. Skyrms [7, Appendix 2] has argued for a constraint he calls Millers’ principle. This asserts that an agent’s subjective belief in a proposition, given that he believes the objective chance to be a certain value, should be equal to that value. Skyrms argues that this is a plausible rule for assimilating information about chance. We will call this relation the subjective/objective Miller’s principle.

The world is described in terms of facts and events. Facts tend to hold and events tend to occur over intervals of time. So facts and events are associated with the time intervals over which they hold or occur in the various world-histories. Facts are distinguished from events on the basis of their temporal properties. A fact may hold over several intervals in any given world-history and if a fact holds over an interval then it holds over all subintervals of that interval.

Events are somewhat more complex than facts. First, one must distinguish between *event types* and *event tokens*. An event type is a general class of events and an event token is a specific instance of an event type. Event tokens are unique individuals – the interval over which an event token occurs is the unique interval associated with the event token and an event token can occur at most once in any world-history. The present work deals with event types, which for brevity are simply referred to as events.

### 3 The Logic of Time, Chance, and Probability

#### 3.1 Syntax

The language of  $\mathcal{L}_{tcp}$  contains two predicates to refer to facts and event types occurring in time:  $HOLDS(FA, t_1, t_2)$  is true if fact  $FA$  holds over the time interval  $t_1$  to  $t_2$ , and  $OCCURS(EV, t_1, t_2)$  is true if event  $EV$  occurs during the interval  $t_1$  to  $t_2$ . Henceforth we will use the symbol  $t$ , possibly subscripted, to denote time points;  $\phi$ ,  $\psi$ , and  $\gamma$  to denote formulas; and  $\alpha$  and  $\beta$  to denote probability values.

In addition to the usual first-order logical operators, the language contains two modal operators to express subjective probability and objective chance. The operators are subscripted with a time since according to the ontology subjective probability and objective chance are taken relative to a point in time. We write  $P_t(\phi)$  to denote the subjective probability of  $\phi$  at time  $t$  and we write  $pr_t(\phi)$  to denote the objective chance of  $\phi$  at time  $t$ . Probability is treated as a sentential operator in the object language. So the probability operators can be arbitrarily nested and combined with one another, allowing us to write complex sentences like: “I believe there was a one in a million chance of my winning the lottery, yet I won.”

$$P_{t_3}(pr_{t_2}(OCCURS(\text{win}, t_1, t_2))) = 10^{-6} \wedge \\ OCCURS(\text{win}, t_1, t_2) = 1,$$

where  $t_1 < t_2 < t_3$ . We also allow conditional probability sentences such as  $P_t(\phi|\psi) = \alpha$ , which is interpreted as shorthand for  $P_t(\phi \wedge \psi) = \alpha \cdot P_t(\psi)$ .

The language of  $\mathcal{L}_{tcp}$  is fully first-order, allowing quantification over time points, probability values, and domain individuals. A formal specification of the syntax is provided in the full paper [2].

#### 3.2 Semantics

We describe only the more interesting aspects of the models of  $\mathcal{L}_{tcp}$ . The models are completely specified in the full paper. A model is a tuple  $\langle W, D, FN, NFN, PFN, FRL, ERL, NRL, FA, EVENTS, EV, R, \mathcal{X}, PR_o, PR_s, F \rangle$ , where:

- $W$  is the set of possible world-histories, called worlds.
- $D$  is the non-empty domain of individuals.
- $FA$  is the set of facts, a subset of  $2^{(\mathbb{R} \times \mathbb{R}) \times W}$ . A fact is a set of  $\langle \text{temporal interval, world} \rangle$  pairs:  $\{\langle \langle t_1, t'_1 \rangle, w_1 \rangle, \dots, \langle \langle t_n, t'_n \rangle, w_n \rangle\}$ . If  $fa$  is a fact and  $\langle \langle t_1, t_2 \rangle, w \rangle \in fa$  then  $fa$  holds throughout interval  $\langle t_1, t_2 \rangle$  in world-history  $w$ .
- $EVENTS$  is the set of event tokens, a subset of  $(\mathbb{R} \times \mathbb{R}) \times W$ . An event token is a single  $\langle \text{temporal interval, world} \rangle$  pair.
- $EV$  is the set of event types, a subset of  $2^{EVENTS}$ . An event type is a set of event tokens:  $\{\langle \langle t_1, t'_1 \rangle, w_1 \rangle, \dots, \langle \langle t_n, t'_n \rangle, w_n \rangle\}$ . If  $ev$  is an event and  $\langle \langle t_1, t_2 \rangle, w \rangle \in ev$  then  $ev$  occurs during interval  $\langle t_1, t_2 \rangle$  in world-history  $w$ .

1.  $\llbracket \text{HOLDS}(rf(trm_1, \dots, trm_n), ttrm_1, ttrm_2) \rrbracket^{M,w,g} = \text{true}$  iff  
 $\langle \langle \llbracket ttrm_1 \rrbracket^{M,w,g}, \llbracket ttrm_2 \rrbracket^{M,w,g} \rangle, w \rangle \in F(rf)(\llbracket trm_1 \rrbracket^{M,w,g}, \dots, \llbracket trm_n \rrbracket^{M,w,g})$ .
2.  $\llbracket \text{OCCURS}(re(trm_1, \dots, trm_n), ttrm_1, ttrm_2) \rrbracket^{M,w,g} = \text{true}$  iff  
 $\langle \langle \llbracket ttrm_1 \rrbracket^{M,w,g}, \llbracket ttrm_2 \rrbracket^{M,w,g} \rangle, w \rangle \in e$  for some  
 $e \in F(re)(\llbracket trm_1 \rrbracket^{M,w,g}, \dots, \llbracket trm_n \rrbracket^{M,w,g})$ .
3.  $\llbracket \text{pr}_{ttrm}(\phi) \rrbracket^{M,w,g} = \mu_{\llbracket ttrm \rrbracket^{M,w,g}}^w(\{w' \in R_{\llbracket ttrm \rrbracket^{M,w,g}}^w : \llbracket \phi \rrbracket^{M,w',g} = \text{true}\})$ .
4.  $\llbracket \text{P}_{ttrm}(\phi) \rrbracket^{M,w,g} = \mu_{\llbracket ttrm \rrbracket^{M,w,g}}^w(\{w' : \llbracket \phi \rrbracket^{M,w',g} = \text{true}\})$ .

Figure 1: Semantic definitions

- $R$  is an accessibility relation defined on  $\mathfrak{R} \times W \times W$ .  $R(t, w_1, w_2)$  means that world-histories  $w_1$  and  $w_2$  are indistinguishable up to and including time  $t$ . If  $R(t, w_1, w_2)$  we say a world-history  $w_2$  is  $R$ -accessible from  $w_1$  at time  $t$ . The set of all world-histories  $R$ -accessible from  $w$  at time  $t$  will be designated  $R_t^w$ . For each time  $t$ , the  $R_t^w$  partition the world-histories into sets of equivalence classes indistinguishable up to  $t$ .
- $\mathcal{X}$  is a  $\sigma$ -algebra over  $W^1$ , containing all the sets corresponding to wff's in the language, as well as all  $R$ -equivalence classes of world-histories.
- $PR_o$  is the objective probability assignment function that assigns to each time  $t \in \mathfrak{R}$  and world-history  $w \in W$  a countably additive probability distribution  $\mu_o^w$  defined over  $\mathcal{X}$ .
- $PR_s$  is the subjective probability assignment function that assigns to each time  $t \in \mathfrak{R}$  and world-history  $w \in W$  a countably additive probability distribution  $\mu_s^w$  defined over  $\mathcal{X}$ .

Given the models described above, the semantic definitions for the well-formed formulas can now be defined. Denotations are assigned to expressions relative to a model, a world-history within the model, and an assignment of individuals in the domain to variables. The denotation of an expression  $\phi$  relative to a model  $M$  and a world-history  $w$ , and a variable assignment  $g$  is designated by  $\llbracket \phi \rrbracket^{M,w,g}$ . Figure 1 shows the less familiar semantic definitions. The remainder are provided in the full paper.

### 3.2.1 Semantic Constraints

In order to obtain the properties discussed in the ontology, we impose eight constraints on the models.

The future-branching temporal tree is defined in terms of the  $R$  relation over world-histories. To capture the property that the tree does not branch into the past, we say that if two world-histories are indistinguishable up to time  $t_2$  then they are indistinguishable up to any earlier time:

- (C1) If  $t_1 \leq t_2$  and  $R(t_2, w_1, w_2)$  then  $R(t_1, w_1, w_2)$ .

<sup>1</sup>A  $\sigma$ -algebra over  $W$  is a class of subsets that contains  $W$  and is closed under complement and countable union.

Since  $R$  just represents the indistinguishability of histories up to a time  $t$ , for a fixed time  $R$  is an equivalence relation, i.e., reflexive, symmetric, and transitive:

- (C2)  $R(t, w, w)$   
 If  $R(t, w_1, w_2)$  then  $R(t, w_2, w_1)$   
 If  $R(t, w_1, w_2)$  and  $R(t, w_2, w_3)$  then  $R(t, w_1, w_3)$

As mentioned earlier, facts and events differ in their temporal properties. This distinction is captured by the following two semantic constraints. If a fact holds over an interval, it holds over all subintervals, except possibly at the endpoints:

- (C3) If  $t_1 \leq t_2 \leq t_3 \leq t_4$ ,  $t_1 \neq t_3$ ,  $t_2 \neq t_4$ ,  $fa \in FA$   
 and  $\langle \langle t_1, t_4 \rangle, w \rangle \in fa$  then  $\langle \langle t_2, t_3 \rangle, w \rangle \in fa$ .

An event token occurs only once in each world-history:

- (C4) If  $evt \in EVENTS$ ,  $\langle \langle t_1, t_2 \rangle, w \rangle \in evt$ , and  
 $\langle \langle t_3, t_4 \rangle, w \rangle \in evt$  then  $t_1 = t_3$  and  $t_2 = t_4$ .

If two worlds are indistinguishable up to a time then they must share a common past up to that time. And if they share a common past up to a given time, they must agree on all facts and events up to that time. To enforce this relationship, we impose the constraint that if two world-histories are  $R$ -accessible at time  $t$ , they must agree on all facts(events) that hold(occur) over intervals ending before or at the same time as  $t$ :

- (C5) If  $t_0 \leq t_1 \leq t_2$  and  $R(t_2, w_1, w_2)$  then  $\langle \langle t_0, t_1 \rangle, w_1 \rangle \in A$  iff  $\langle \langle t_0, t_1 \rangle, w_2 \rangle \in A$ , where  $A$  is a fact or event.

The ontology discussed two desired characteristics of objective chance. The first is that the chance at a time  $t$  be completely determined by the history up to that time. The second desired characteristic is that the chance of the present and past should be either zero or one, depending on whether or not it actually happened. These two properties follow as meta-theorems from the following two constraints:

- (C6) For all  $X \in \mathcal{X}$ ,  $t \leq t'$ , and  $w, w'$  such that  
 $R(t, w, w')$ ,  
 $\mu_t^w(R_t^{w'}) > 0 \rightarrow \mu_t^{w'}(X) = \mu_t^w(X|R_t^{w'})$ .

- (C7)  $\mu_t^w(R_t^w) > 0$ .

**Meta-theorem 1** *The probability of the present and past is either zero or one.*

$$\mu_t^w(R_t^w) = 1$$

1.  $\mu_t^w(R_t^w) > 0$  (C7)
2.  $\mu_t^w(R_t^w) = \mu_t^w(R_t^w|R_t^w)$  Modus Ponens: (C6),1
3.  $\mu_t^w(R_t^w) = 1$  def of c-prob

Defining the probabilities in this way makes good intuitive sense if we look at the meaning of  $R$ .  $R_t^w$  designates the set of world-histories that are objectively possible with respect to  $w$  at time  $t$ . It is natural that the set of world-histories that are objectively likely with respect to  $w$  at time  $t$  should be a subset of the ones that are possible.

**Meta-theorem 2** *If two worlds are indistinguishable up to time  $t$  then they have identical probability distributions at that time.*

*If  $R(t, w, w')$  then  $\mu_t^{w'}(X) = \mu_t^w(X)$*

1.  $\mu_t^w(R_t^{w'}) > 0$  (C2), (C7)
2.  $\mu_t^w(R_t^{w'}) = \mu_t^w(X|R_t^{w'})$  Modus Ponens: (C6),1
3.  $\mu_t^w(X|R_t^{w'}) = \mu_t^w(X|R_t^w)$  (C2)
4.  $\mu_t^w(R_t^w) = 1$  Meta-theorem 1
5.  $\mu_t^{w'}(X) = \mu_t^w(X)$  def of c-prob

In the ontology, we argued that subjective probability and objective chance should be related to one another by Millers' principle. This relation is enforced by the following constraint, which says that the probability of a set of worlds  $X$ , given some  $R$  equivalence class, should just be the objective chance in that equivalence class.

$$(C8) \quad \mu s_t^w(X|R_t^w) = \mu o_t^w(X)$$

## 4 Theorems

We first provide several simple theorems that will be used in later proofs. Then we prove two forms of Miller's principle and provide two associated expected value properties. Proofs not provided here appear in the full paper.

**Theorem 3** *From  $\phi \leftrightarrow \psi$  infer  $\text{pr}_t(\phi) = \text{pr}_t(\psi)$ .*

**Theorem 4** *Stronger sentences have lower probability. From  $\phi \rightarrow \psi$  infer  $P_t(\phi) \leq P_t(\psi)$ .*

**Theorem 5** *Certainty cannot be conditioned away from.*

$$P_t(\phi \wedge \psi) = P_t(\psi) \rightarrow P_t(\phi \wedge \psi \wedge \gamma) = P_t(\psi \wedge \gamma)$$

**Theorem 6** *The present and past are objectively certain.*

*Let  $\Phi$  be a fact or event:*

*HOLDS( $\phi, t_\phi, t'_\phi$ ) or OCCURS( $\phi, t_\phi, t'_\phi$ ) then*

$$\forall t, t_\phi, t'_\phi (t'_\phi \leq t) \rightarrow [\text{pr}_t(\Phi) = 0 \vee \text{pr}_t(\Phi) = 1]$$

The semantic constraints on objective chance give us a version of Miller's principle that relates objective chance at different times. It says that the chance of a sentence  $\phi$  at a time, given that the chance of  $\phi$  at the same or a later time is  $\alpha$ , should be  $\alpha$ .

**Theorem 7 Objective Miller's Principle (OMP)**  
*All instances of the following sentence schema are valid in  $\mathcal{L}_{\text{tep}}$ .*

$$\forall \alpha, t_0, t_1 (t_0 \leq t_1) \rightarrow \text{pr}_{t_0}(\phi | \text{pr}_{t_1}(\phi) = \alpha) = \alpha$$

*Proof:*

We first prove an expected value property and then use it to prove Miller's principle. Let  $t, t'$  be two time points  $t \leq t'$  and consider the  $R$ -equivalence classes of worlds at time  $t'$ . Let the variable  $r$  range over these equivalence classes. The  $r$  form a partition of  $W$ , so the probability of a set  $X$  can be written as the integral over this partition:

$$\mu o_t^w(X) = \int_{r \subset W} \mu o_t^w(X|r) \mu o_t^w(dr)$$

Since the history up to time  $t'$  determines the probability at time  $t'$ , this can be written as

$$\mu o_t^w(X) = \int_{r \subset W} \mu o_{t'}^r(X) \mu o_t^w(dr),$$

where  $\mu o_{t'}^r$  denotes the probability at time  $t'$  in equivalence class  $r$ . Since the probability at a given time is assumed to be constant over all worlds in an  $R$ -equivalence class, the probability at a given time is the expected value of the probability at any future time:

$$\mu o_t^w(X) = \int_W \mu o_{t'}^{w'}(X) \mu o_t^w(dw').$$

Next we show that Miller's principle is valid in the probability models. By the expected value property,

$$\begin{aligned} \mu o_t^w(X \cap \{w' : \mu o_{t'}^{w'}(X) = \alpha\}) = \\ \int_W \mu o_{t'}^{w''}(X \cap \{w' : \mu o_{t'}^{w'}(X) = \alpha\}) \mu o_t^w(dw''). \end{aligned}$$

Now, by semantic constraints (C6) and (C7) it follows that

$$\begin{aligned} \forall w \in \{w' : \mu o_{t'}^{w'}(X) = \alpha\}, \\ \mu o_t^w(\{w' : \mu o_{t'}^{w'}(X) = \alpha\}) = 1 \\ \forall w \notin \{w' : \mu o_{t'}^{w'}(X) = \alpha\}, \\ \mu o_t^w(\{w' : \mu o_{t'}^{w'}(X) = \alpha\}) = 0. \end{aligned}$$

So we can restrict the integral to the set

$$\begin{aligned} \{w' : \mu o_{t'}^{w'}(X) = \alpha\}: \\ = \int_{\{w' : \mu o_{t'}^{w'}(X) = \alpha\}} \mu o_{t'}^{w''}(X \cap \{w' : \mu o_{t'}^{w'}(X) = \alpha\}) \mu o_t^w(dw''). \end{aligned}$$

And by the above property again

$$\begin{aligned} \mu o_{t'}^{w''}(X \cap \{w' : \mu o_{t'}^{w'}(X) = \alpha\}) = \alpha, \text{ so} \\ = \alpha \cdot \int_{\{w' : \mu o_{t'}^{w'}(X) = \alpha\}} \mu o_t^w(dw''). \end{aligned}$$

$$= \alpha \cdot \mu o_t^w(\{w' : \mu o_{t'}^{w'}(X) = \alpha\}).$$

By the semantic definitions it follows that

$$P_t(\phi \wedge P_{t'}(\phi) = \alpha) = \alpha \cdot P_t(P_{t'}(\phi) = \alpha).$$

And by a slight generalization of the proof it follows that

$$\forall (t \leq t') P_t(\phi \wedge P_{t'}(\phi) \geq \alpha) \geq \alpha \cdot P_t(P_{t'}(\phi) \geq \alpha). \quad \square$$

From the Objective Miller's Principle it follows directly that current chance is the expected value of current chance applied to current or future chance.

**Theorem 8 Objective Expected Value Property**  
*All instances of the following sentence schema are valid in  $\mathcal{L}_{tep}$ .*

$$\forall \alpha, \beta, t_1, t_2 (t_1 \leq t_2) \rightarrow \\ [\text{pr}_{t_1}(\text{pr}_{t_2}(\phi) \geq \alpha) \geq \beta \rightarrow \text{pr}_{t_1}(\phi) \geq \alpha \cdot \beta]$$

As discussed in the ontology, the current subjective probability of a sentence, given that the current or future chance is some value should be that value. The following theorem shows that this property follows from the semantics of the logic.

**Theorem 9 Subjective/Objective Miller's Principle (SOMP)**  
*All instances of the following sentence schema are valid in  $\mathcal{L}_{tep}$ .*

$$\forall \alpha, t_0, t_1 (t_0 \leq t_1) \rightarrow \text{P}_{t_0}(\phi | \text{pr}_{t_1}(\phi) = \alpha) = \alpha$$

*Proof:* We first prove an expected value property and then use it to prove Miller's principle. Let  $t, t'$  be two time points  $t \leq t'$  and consider the R-equivalence classes of worlds at time  $t'$ . Let the variable  $r$  range over these equivalence classes. The  $r$  form a partition of  $W$ , so the probability of a set  $X$  can be written as the integral over this partition:

$$\mu s_t^w(X) = \int_{r \subset W} \mu s_t^w(X|r) \mu s_t^w(dr)$$

By semantic constraint (C8), this can be written as

$$\mu s_t^w(X) = \int_{r \subset W} \mu o_t^r(X) \mu s_t^w(dr)$$

where  $\mu o_t^r$  denotes the objective chance at time  $t$  in equivalence class  $r$ . Since the chance at a given time is assumed to be constant over all worlds in an R-equivalence class, the subjective probability at any time is the expected value of the subjective probability applied to the objective chance at that time:

$$\mu s_t^w(x) = \int_W \mu o_t^{w'}(X) \mu s_t^w(dw')$$

Next we show that the Subjective/Objective Miller's principle is valid in the probability models. By the above subjective/objective expected value property,

$$\mu s_t^w(X \cap \{w' : \mu o_{t'}^{w'}(X) = \alpha\}) = \\ \int_W \mu o_t^{w''}(X \cap \{w' : \mu o_{t'}^{w'}(X) = \alpha\}) \mu s_t^w(dw'')$$

By Objective Miller's Principle,

$$\mu s_t^w(X \cap \{w' : \mu o_{t'}^{w'}(X) = \alpha\}) = \\ \alpha \int_W \mu o_t^{w''}(\{w' : \mu o_{t'}^{w'}(X) = \alpha\}) \mu s_t^w(dw'')$$

Finally, by the subjective/objective expected value property,

$$\mu s_t^w(X \cap \{w' : \mu o_{t'}^{w'}(X) = \alpha\}) = \\ \alpha \mu s_t^w(\{w' : \mu o_{t'}^{w'}(X) = \alpha\})$$

So by the semantic definitions it follows that

$$\forall t, t' (t \leq t') \rightarrow \text{P}_t(\phi | \text{pr}_{t'}(\phi) = \alpha) = \alpha$$

And by a slight generalization of the proof it follows that

$$\forall t, t' (t \leq t') \rightarrow \text{P}_t(\phi | \text{pr}_{t'}(\phi) \geq \alpha) \geq \alpha$$

□

From the subjective/objective Miller's principle it follows directly that subjective probability is the expected value of current subjective probability applied to present or future chance.

**Theorem 10 Subjective/Objective Expected Value Property**

*All instances of the following sentence schema are valid in  $\mathcal{L}_{tep}$ .*

$$\forall \alpha, \beta, t_1, t_2 (t_1 \leq t_2) \rightarrow \\ [\text{P}_{t_1}(\text{pr}_{t_2}(\phi) \geq \alpha) \geq \beta \rightarrow \text{P}_{t_1}(\phi) \geq \alpha \cdot \beta]$$

## 5 Distinguishing Evidential and Causal Correlation

We wish to distinguish between two situations in which an agent may believe that two conditions are correlated. An agent may believe that two conditions are correlated because one is simply evidence for another and an agent may believe that they are correlated because one causes the other.

Let  $\Phi$  stand for the formula  $HOLDS(\phi, t_\phi, t'_\phi)$  or  $OCCURS(\phi, t_\phi, t'_\phi)$  and let  $\Psi$  stand for the formula  $HOLDS(\psi, t_\psi, t'_\psi)$  or  $OCCURS(\psi, t_\psi, t'_\psi)$ . We represent evidential correlation as correlation in the subjective probability distribution, which is the standard approach in Bayesian decision theory.

**Definition 11** *We say that  $\Psi$  is evidence for or against  $\Phi$  iff*

$$\text{P}_{now}(\Phi | \Psi) \neq \text{P}_{now}(\Phi) \quad (1)$$

It follows from this definition that  $\Psi$  is not evidence for or against  $\Phi$  iff

$$\text{P}_{now}(\Phi | \Psi) = \text{P}_{now}(\Phi)$$

We represent causal correlation by reference to the objective chance distribution. We represent an agent's belief that  $\Psi$  causally influences  $\Phi$  by saying that there is some value for the objective chance of  $\Phi$  such that the agent's belief in  $\Phi$  given the objective chance of  $\Phi$  just before  $\Psi$  holds or occurs is not the same as the agent's belief given also knowledge of  $\Psi$ . In other words, knowledge of  $\Psi$  overrides knowledge of the objective chance of  $\Phi$ .

**Definition 12** *We say that  $\Psi$  is a cause of  $\Phi$  iff*

$$\exists \alpha \text{P}_{now}(\Phi | \text{pr}_{t_\psi}(\Phi) = \alpha \wedge \Psi) \neq \alpha. \quad (2)$$

Note that this does not necessarily imply that  $\text{P}_{now}(\Phi | \Psi) \neq \text{P}_{now}(\Phi)$ . Thus we may have causal correlation without evidential correlation and, conversely, we may have evidential correlation without causal correlation. It follows from this definition that  $\Psi$  is not a cause of  $\Phi$  iff

$$\forall \alpha \text{P}_{now}(\Phi | \text{pr}_{t_\psi}(\Phi) = \alpha \wedge \Psi) = \alpha.$$

### 5.1 Example

We now present an example demonstrating the use of the definitions and theorems. We wish to describe the following situation. You have a coin that may be biased 3:1 towards heads or 3:1 towards tails. You believe there is an equal probability of each. You can observe the coin. If the coin looks shiny, this increases your belief that the coin is biased towards heads. You also have a magnet that you can use to influence the outcome of the coin toss. Turning on the magnet biases the coin more toward heads. We can describe the situation with the following set of sentences in which “heads” is the event of the coin landing heads, “shiny” is the event of the coin being observed to be shiny, and “magnet” is the fact that the magnet is on.

$$(now < t_0 < t_1 < t_2 < t_3 < t_4)$$

Turning on the magnet influences the chance of heads.

$$P_{now}(OCCURS(Heads, t_2, t_3)) \quad (3)$$

$$\begin{aligned} \text{pr}_{t_1}(OCCURS(Heads, t_2, t_3)) &= 3/4 \wedge \\ HOLDS(Magnet, t_1, t_4)) &= 7/8 \end{aligned}$$

$$P_{now}(OCCURS(Heads, t_2, t_3)) \quad (4)$$

$$\begin{aligned} \text{pr}_{t_1}(OCCURS(Heads, t_2, t_3)) &= 1/4 \wedge \\ HOLDS(Magnet, t_1, t_4)) &= 1/2 \end{aligned}$$

The probability that the coin is biased toward heads and the probability that the coin is biased toward tails are equal.<sup>2</sup>

$$P_{now}(\text{pr}_{t_1}(OCCURS(Heads, t_2, t_3)) = 3/4) = \quad (5)$$

$$P_{now}(\text{pr}_{t_1}(OCCURS(Heads, t_2, t_3)) = 1/4) = 1/2$$

Observing the coin doesn't influence the chance of heads.

$$\forall \alpha, t (t > now) \rightarrow \quad (6)$$

$$\begin{aligned} P_{now}(OCCURS(Heads, t_2, t_3) \mid \\ \text{pr}_t(OCCURS(Heads, t_2, t_3)) = \alpha \wedge \\ OCCURS(Shiny, t_0, t_2)) &= \alpha \end{aligned}$$

Observing the coin gives us knowledge of its bias.

$$P_{now}(\text{pr}_{t_0}(OCCURS(Heads, t_2, t_3)) = 3/4 \mid \quad (7)$$

$$OCCURS(Shiny, t_0, t_2)) = 5/8$$

$$P_{now}(\text{pr}_{t_0}(OCCURS(Heads, t_2, t_3)) = 1/4 \mid \quad (8)$$

$$OCCURS(Shiny, t_0, t_2)) = 3/8$$

Turning on the magnet does not give us knowledge of the coin's bias.

$$\forall \alpha P_{now}(\text{pr}_{t_1}(OCCURS(Heads, t_2, t_3)) = \alpha \mid \quad (9)$$

$$HOLDS(Magnet, t_1, t_4)) =$$

$$P_{now}(\text{pr}_{t_1}(OCCURS(Heads, t_2, t_3)) = \alpha)$$

The coin is either biased toward heads or toward tails.

$$\forall t \text{pr}_t(OCCURS(Heads, t_2, t_3)) = 3/4 \vee \quad (10)$$

$$\text{pr}_t(OCCURS(Heads, t_2, t_3)) = 1/4 \quad (11)$$

<sup>2</sup>It would be more appropriate to say that our belief that the current chance is 3/4 or 1/4 is 1/2 and that in the absence of events that will influence the chance, chance will remain unchanged till time  $t_1$ . Such an inference would require some kind of theory of persistence, which is beyond the scope of this paper.

Using this information, we can make several useful inferences. First we can derive the unconditional probability that the coin will land heads. From (5) by SOMP we have

$$\begin{aligned} P_{now}(OCCURS(Heads, t_2, t_3)) &= \\ (1/2)(3/4) + (1/2)(1/4) &= 1/2 \end{aligned} \quad (12)$$

Next, we can use the above information to derive the probability that the coin will come up heads given that it is observed to be shiny. Instantiating (6) with  $\alpha = 3/4$  and  $t = t_0$  and multiplying the result by (7) we get

$$\begin{aligned} P_{now}(OCCURS(Heads, t_2, t_3) \wedge \\ \text{pr}_{t_0}(OCCURS(Heads, t_2, t_3)) = 3/4 \mid \\ OCCURS(Shiny, t_0, t_2)) &= \\ (5/8)(3/4) &= 15/32 \end{aligned} \quad (13)$$

And instantiating (6) with  $\alpha = 1/4$  and  $t = t_0$  and multiplying the result by (8) we get

$$\begin{aligned} P_{now}(OCCURS(Heads, t_2, t_3) \wedge \\ \text{pr}_{t_0}(OCCURS(Heads, t_2, t_3)) = 1/4 \mid \\ OCCURS(Shiny, t_0, t_2)) &= (3/8)(1/4) = 3/32 \end{aligned} \quad (14)$$

From (10), (13), and (14) by the law of total probability we get

$$\begin{aligned} P_{now}(OCCURS(Heads, t_2, t_3) \mid \\ OCCURS(Shiny, t_0, t_2)) &= 9/16 \end{aligned} \quad (15)$$

We can also derive the probability of heads given that we activate the magnet. From (3), (5), and (9) we get

$$\begin{aligned} P_{now}(OCCURS(Heads, t_2, t_3) \wedge \\ \text{pr}_{t_2}(OCCURS(Heads, t_2, t_3)) = 3/4 \mid \\ HOLDS(Magnet, t_1, t_4)) &= (1/2)(7/8) = 7/16 \end{aligned} \quad (16)$$

From (4), (5), and (9) we get

$$\begin{aligned} P_{now}(OCCURS(Heads, t_2, t_3) \wedge \\ \text{pr}_{t_2}(OCCURS(Heads, t_2, t_3)) = 1/4 \mid \\ HOLDS(Magnet, t_1, t_4)) &= (1/2)(1/2) = 1/4 \end{aligned} \quad (17)$$

From (10), (16), and (17) by the law of total probability we get

$$\begin{aligned} P_{now}(OCCURS(Heads, t_2, t_3) \mid \\ HOLDS(Magnet, t_1, t_4)) &= 11/16 \end{aligned} \quad (18)$$

### 5.2 The temporal flow of causality

Using our definition of causal influence and SOMP we can now show that an agent whose beliefs are represented with  $\mathcal{L}_{tcp}$  believes that the past cannot be influenced.

**Theorem 13** *Let  $\Phi$  be a fact or event:  $HOLDS(\phi, t_\phi, t'_\phi)$  or  $OCCURS(\phi, t_\phi, t'_\phi)$  and let  $\Psi$  be a fact or event:  $HOLDS(\psi, t_\psi, t'_\psi)$  or  $OCCURS(\psi, t_\psi, t'_\psi)$ .*

*Then all instances of the following sentence schema are valid in  $\mathcal{L}_{tcp}$ .*

$$\begin{aligned} \forall \alpha, t, t_\phi, t'_\phi, t_\psi, t'_\psi (t'_\phi \leq t_\psi) \wedge (t \leq t_\psi) \rightarrow \\ P_t(\Phi \mid \text{pr}_{t_\psi}(\Psi) = \alpha \wedge \Psi) = \alpha \end{aligned}$$

*Proof:* We prove a slightly more general result of which the above sentence is an instance. By the Subjective/Objective Miller's Principle,

$$\forall \alpha, t, t', t_\phi, t'_\phi (t'_\phi \leq t') \wedge (t \leq t') \rightarrow \quad (19)$$

$$P_t(\Phi \wedge pr_{t'}(\Phi) = \alpha) = \alpha \cdot P_t(pr_{t'}(\Phi) = \alpha)$$

Since valid formulas have probability one, it follows by Theorem 6 that,

$$\forall \alpha, t, t', t_\phi, t'_\phi, t_\psi, t'_\psi (t'_\phi \leq t') \wedge (t \leq t') \rightarrow \quad (20)$$

$$P_t(\Phi \wedge pr_{t'}(\Phi) = \alpha \wedge [pr_{t'}(\Phi) = 0 \vee pr_{t'}(\Phi) = 1]) =$$

$$\alpha \cdot P_t(pr_{t'}(\Phi) = \alpha \wedge [pr_{t'}(\Phi) = 0 \vee pr_{t'}(\Phi) = 1])$$

Since  $pr_{t'}(\Phi) = 0$  and  $pr_{t'}(\Phi) = 1$  are mutually exclusive, we have

$$\forall \alpha, t, t', t_\phi, t'_\phi (t'_\phi \leq t') \wedge (t \leq t') \rightarrow \quad (21)$$

$$P_t(\Phi \wedge pr_{t'}(\Phi) = \alpha \wedge pr_{t'}(\Phi) = 0) +$$

$$P_t(\Phi \wedge pr_{t'}(\Phi) = \alpha \wedge pr_{t'}(\Phi) = 1) =$$

$$\alpha \cdot P_t(pr_{t'}(\Phi) = \alpha \wedge pr_{t'}(\Phi) = 0) +$$

$$\alpha \cdot P_t(pr_{t'}(\Phi) = \alpha \wedge pr_{t'}(\Phi) = 1)$$

Now we have three cases to consider: *i)*  $\alpha = 0$ , *ii)*  $\alpha = 1$ , *iii)*  $0 < \alpha < 1$ .

#### Case i)

Expression (21) reduces to

$$\forall t, t', t_\phi, t'_\phi, t_\psi, t'_\psi (t'_\phi \leq t') \wedge (t \leq t') \rightarrow \quad (22)$$

$$P_t(\Phi \wedge pr_{t'}(\Phi) = 0) = 0 \cdot P_t(pr_{t'}(\Phi) = 0)$$

So by Theorem 4 and universal generalization,

$$\forall t, t', t_\phi, t'_\phi, t_\psi, t'_\psi (t'_\phi \leq t') \wedge (t \leq t') \rightarrow \quad (23)$$

$$P_t(\Phi \wedge pr_{t'}(\Phi) = 0 \wedge \Psi) = 0 \cdot P_t(pr_{t'}(\Phi) = 0 \wedge \Psi)$$

#### Case ii)

Expression (21) reduces to

$$\forall t, t', t_\phi, t'_\phi (t'_\phi \leq t') \wedge (t \leq t') \rightarrow \quad (24)$$

$$P_t(\Phi \wedge pr_{t'}(\Phi) = 1) = P_t(pr_{t'}(\Phi) = 1)$$

So by Theorem 5 and universal generalization,

$$\forall t, t', t_\phi, t'_\phi, t_\psi, t'_\psi (t'_\phi \leq t') \wedge (t \leq t') \rightarrow \quad (25)$$

$$P_t(\Phi \wedge pr_{t'}(\Phi) = 1 \wedge \Psi) = P_t(pr_{t'}(\Phi) = 1 \wedge \Psi)$$

#### Case iii)

For  $0 < \alpha < 1$ ,  $P_t(pr_{t'}(\Phi) = \alpha) = 0$ . So by Theorem 4 and universal generalization,

$$\forall \alpha, t, t', t_\phi, t'_\phi, t_\psi, t'_\psi \quad (26)$$

$$(t'_\phi \leq t') \wedge (t \leq t') \wedge (0 < \alpha < 1) \rightarrow$$

$$P_t(\Phi \wedge pr_{t'}(\Phi) = \alpha \wedge \Psi) = P_t(pr_{t'}(\Phi) = \alpha \wedge \Psi)$$

Therefore we have proven that the following sentence is valid

$$\forall \alpha, t, t', t_\phi, t'_\phi, t_\psi, t'_\psi (t'_\phi \leq t') \wedge (t \leq t') \rightarrow \quad (27)$$

$$P_t(\Phi | pr_{t'}(\Phi) = \alpha \wedge \Psi) = \alpha$$

from which it follows that the past cannot be influenced.  $\square$

## 6 Related Work

Three outstanding subjective theories of objective chance from the philosophical literature are those of van Fraassen [9], Lewis [6], and Skyrms [7]. van Fraassen's model of objective chance is more constrained than Lewis's model which is more constrained than Skyrms's model. Thus, in van Fraassen's model, chance has more inherent properties than in either Lewis's or Skyrms's models. van Fraassen's theory is the only one of the three that is cast in a temporal framework. All three are semantic theories and do not provide logical languages.

The model of objective chance used in  $\mathcal{L}_{tcp}$  is based on van Fraassen's [9] model of objective chance. He presents a semantic theory that models subjective probability and objective chance, using a future-branching model of time points. van Fraassen places two constraints on objective chance:

1. The chance of a past is either 0 or 1, depending on whether or not it actually occurred.
2. Chance at a time is completely determined by history of the world up to that time.

From these assumptions, he shows the following relation between subjective probability and objective chance

$$P_t(X|Y) = E_Y[C_t(X)],$$

where  $P_t$  is the subjective probability at time  $t$ ,  $C_t$  is the objective chance at time  $t$ ,  $E_Y$  is the expected value given  $Y$ , and provided the truth of  $Y$  depends only on the history up to  $t$ . This relation entails both Miller's principle and Lewis's principal principle, discussed below. Note that van Fraassen does not show that a similar relation holds between objective chances at different times. In van Fraassen's models, objective chance can change with time but truth values cannot.

Lewis's [6] theory of objective chance is based on his assertion that

... we have some very firm and definite opinions concerning reasonable credence (*subjective probability*) about chance (*objective chance*). These opinions seem to me to afford the best grip we have on the concept of chance.

He describes a number of intuitive relationships between subjective probability and objective chance and shows that these are captured by his principal principle:

$$Pr(A|pr_t(A) = \alpha \wedge E) = \alpha,$$

where  $Pr$  is subjective probability,  $pr$  is objective chance, and  $E$  is any proposition compatible with  $pr_t(A) = \alpha$  and admissible at time  $t$ .

The interesting thing here is the proposition  $E$ . The constraint that  $E$  be compatible with  $pr_t(A) = \alpha$  means that  $Pr(E \wedge pr_t(A) = \alpha) > 0$ . Admissibility is less readily defined. Lewis does not give a definition of admissibility but he does characterize admissible propositions as "the sort of information whose impact on credence about outcomes comes entirely by way of credence about the chances of those outcomes." So objective chance is invariant with respect to conditioning on

admissible propositions. This concept of invariance under conditioning is the central notion of Brian Skyrms's theory of objective chance.

Skyrms [7] works with the notion of resiliency. A probability value is resilient if it is relatively invariant under conditionalization over a set of sentences. The resiliency of  $Pr(q)$  being  $\alpha$  is defined as 1 minus the amplitude of the wiggle about  $\alpha$ :

The resiliency of  $Pr(q)$  being  $\alpha$  is  $1 - \max |\alpha - Pr_j(q)|$  over  $p_1, \dots, p_n$ , where the  $Pr_j$ 's are gotten by conditionalizing on some Boolean combination of the  $p_i$ 's which is logically consistent with  $q$ .

Skyrms then defines propensity (objective chance) as a highly resilient subjective probability.

Independent of his resiliency notion, Skyrms requires that propensities and subjective probabilities be related by Miller's principle:

$$Pr(A|pr(A) = \alpha) = \alpha,$$

where  $Pr$  is a subjective probability and  $pr$  is a propensity. He shows that Miller's principle entails that subjective probabilities are equal to the expectation of the subjective probabilities applied to the objective probabilities. But Skyrms [7, p158] points out that, counter to intuition, independence in every possible objective distribution does not imply independence in the subjective distribution. This observation provided the motivation for our use of the two probabilities to distinguish causal from evidential correlation.

Halpern [4, 5] presents a probability logic that can represent both statistical and subjective probabilities. Statistical probabilities represent proportions over the domain of individuals, while propositional probabilities represent degrees of belief. The two probability operators in the language can be nested and combined freely with other logical operators. So the language is capable of representing sentences like "The probability is .95 that more than 75% of all birds can fly." The models for the language contain a domain of individuals, a set of possible worlds, a single discrete probability function over the individuals, and a single discrete probability function over the possible worlds. The first probability function is used to assign meaning to the statistical probability operator, while the second is used to assign meaning to the propositional probability operator. Although he does not place constraints within the logic on the relation between the two probabilities, he does discuss a form of Miller's principle that relates subjective and objective probabilities. His version of the principle states that "for any real number  $r_0$  the conditional probability of  $\phi(a)$ , given that the probability of a randomly chosen  $x$  satisfies  $\phi$  is  $r_0$ , is itself  $r_0$ ." He points out that this could be used as a rule for inferring degrees of belief from statistical information.

Bacchus [1] presents a logic essentially identical to that of Halpern. He goes further than Halpern in exploring the inference of degrees of belief from statistical probabilities. According to his principle of direct inference, an agent's belief in a formula is the expected

value with respect to the agent's beliefs of the statistical probability of that formula, given the agent's set of accepted objective assertions.

## References

- [1] F. Bacchus. *Representing and Reasoning With Probabilistic Knowledge*. MIT Press, Cambridge, Mass, 1990.
- [2] P. Haddawy. Believing change and changing belief. Technical Report TR-94-02-01, Dept. of Elect. Eng. & Computer Science, University of Wisconsin-Milwaukee, February 1994. Available via anonymous FTP from `pub/tech_reports` at `ftp.cs.uwm.edu`.
- [3] P. Haddawy. *Representing Plans Under Uncertainty: A Logic of Time, Chance, and Action*, volume 770 of *Lecture Notes in Artificial Intelligence*. Springer-Verlag, Berlin, 1994.
- [4] J.Y. Halpern. An analysis of first-order logics of probability. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 1375-1381, 1989.
- [5] J.Y. Halpern. An analysis of first-order logics of probability. *Artificial Intelligence*, 46:311-350, 1991.
- [6] D. Lewis. A subjectivist's guide to objective chance. In W. Harper, R. Stalnaker, and G. Pearce, editors, *Ifs*, pages 267-298. D. Reidel, Dordrecht, 1980.
- [7] B. Skyrms. *Causal Necessity*. Yale Univ. Press, New Haven, 1980.
- [8] B. Skyrms. Higher order degrees of belief. In D.H. Mellor, editor, *Prospects for Pragmatism*, chapter 6, pages 109-137. Cambridge Univ. Press, Cambridge, 1980.
- [9] B.C. van Fraassen. A temporal framework for conditionals and chance. In W. Harper, R. Stalnaker, and G. Pearce, editors, *Ifs*, pages 323-340. D. Reidel, Dordrecht, 1980.
- [10] J. Venn. *The Logic of Chance*. MacMillan, London, 1866. (new paperback edition, Chelsea, 1962).
- [11] R. von Mises. *Probability, Statistics and Truth*. Allen and Unwin, London, 1957.