Semantical Considerations on Dialectical and Practical Commitments

Munindar P. Singh
Department of Computer Science
North Carolina State University
Raleigh, NC 27695-8206, USA
singh@ncsu.edu

Abstract

This paper studies commitments in multiagent systems. A *dialectical* commitment corresponds to an agent taking a position about a putative fact, including for the sake of argument. A *practical* commitment corresponds to an agent being obliged to another to bring about a condition. Although commitments have been used in many works, an adequate formal semantics and axiomatization for them does not yet exist.

This paper presents a logic of commitments that illustrates the commonalities and differences of the two kinds of commitments. In this manner, it generalizes the developments of previous papers, precisely delineates the meanings of commitments, and identifies important postulates used informally or semiformally in previous work.

This paper considers "social" commitments as introduced in (Singh 1991): by one agent to another, not of an agent to itself. Commitments help formalize a variety of interactive, loosely contractual, settings especially including argumentation and business protocols. Despite several formalizations that use commitments, there is surprisingly little work treating them as an abstraction in their own right. With few exceptions (reviewed in the last section), existing work has generally not emphasized the *model-theoretic semantics* of commitments as such, concentrating on ways of reasoning with or using them. It was a sensible research strategy to first establish that commitments were a useful concept. However, now that the case for commitments has been made well, further progress is hampered by the lack of a clear model-theoretic semantics. For example, tools for designing correct protocols or verifying the interoperability or compliance of agents would rely upon a precise notion of what it means for an agent to be committed, which unfortunately is lacking.

Analyses of commitments range in complexity from obligations to extensive conglomerates of social expectations and obligations. Following (Singh 1999), this paper takes a middle ground, erring perhaps toward simplicity. A commitment here is somewhat like a directed obligation, but one that arises in a context, and which can be manipulated in standardized ways. This paper doesn't discuss context and manipulation, but its semantics provides a basis for specifying them precisely. Richer notions are readily accommodated in this approach. For example, Castelfranchi (1995) requires that a commitment be explicitly accepted by its creditor. This is reasonable in some applications but not others. We can easily define additional concepts that combine multiple instances of the more basic kinds of commitments studied here to achieve the various intuitive requirements of Castelfranchi and other researchers.

Dialectical and Practical Commitments

Commitments fall into two main varieties. Dialectical or *dialogical* commitments (Norman *et al.* 2004) reflect positions taken in dialogue or argumentation. By contrast, practical commitments reflect promises made during negotiation or trade. Thus dialectical commitments are about what holds and practical commitments about what is to be done. For example, a stock quote may be a dialectical commitment about the price; or a practical commitment to sell at the specified price. The two commitments may go together but not necessarily so. This paper lays the groundwork for formulating any such constraints as needed for different applications.

Some key patterns of reasoning arise in dealing with commitments. Examples of natural reasoning patterns during modeling include (**Ex 1**) if a pharmacy commits to delivering medicines if the customer pays and shows a prescription, then once the customer shows the prescription, the pharmacy is committed to delivering medicines if the customer pays; (**Ex 2**) if a merchant commits to a customer to ship goods and commits to the same customer to send warranty paperwork, then the merchant commits to the customer to ship goods and send warranty paperwork; (**Ex 3**) a commitment that if the light is on, the light will be on would not be meaningful. We capture these patterns below as postulates B_2, B_5, and B_8, respectively (along with several other postulates).

How commitments relate to time is important. Dialectical commitments are about claims staked now (even if about future conditions) whereas practical commitments are about actions to be performed or conditions to be brought about in the future. Thus, postulates that hold for one kind of commitment can fail for the other kind. Examples **Ex 1** and **Ex 2** above would hold for dialectical but fail for practical commitments unless we impose additional constraints: because conjunction means that the conditions must hold simultaneously. Clearly, the above examples illustrate practical commitments. Therefore, we identify constraints under which various postulates hold for both kinds of commitments.

Contributions This paper presents the first substantial model-theoretic semantics of commitments, treating dialectical and practical commitments in a uniform light. This paper (1) motivates several reasoning postulates for commitments; (2) states corresponding constraints on models; and (3) makes an extensive comparison with the relevant literature and describes key directions for future research.

Modeling Active Conditional Commitments

Singh (1999) and Verdicchio & Colombetti (V&C) (2003) operationalize commitments as being in one of a small number of states. For simplicity, we only distinguish whether a commitment is *active* or not, and our modalities, i.e., modal operators, express a dialectical or practical commitment being active. Other states can be modeled via other modalities.

Commitments are quite naturally framed conditionally. Examples are "if the interest rate has fallen, (I assert) the bond yield has gone up" (dialectical) or "if you pay, I will deliver the goods" (practical). Winikoff *et al.* (WLH) (2002) and Yolum & Singh (Y&S) (2002) model conditional commitments as unconditional commitments combined with strict (nonmaterial) implication. By contrast, we model conditional commitments as fundamental and unconditional commitments as special cases where the antecedent is true, thereby avoiding modeling a strict implication operator.

Below, $\mathsf{C}^{\mathsf{p}}_{x,y}(r,u)$ denotes a practical commitment meaning that debtor x is committed to creditor y that if r holds, it will bring about u. $\mathsf{C}^{\mathsf{d}}_{x,y}(r,u)$ similarly denotes a dialectical commitment. When r equals $\mathtt{true}$, each type of commitment becomes unconditional.

In some cases, it helps to distinguish from inactive commitments that are discharged from those that haven't yet been activated. Inactive but discharged commitments can sometimes be acceptable alternatives to active commitments. For example, a purchase protocol may state that an offer from a merchant constitutes a commitment to deliver goods. But (if and) when a merchant provides the goods along with its offer, its commitment to deliver the goods is discharged. If we desire this interpretation, we would specify the meaning of the communicative act *offer* not as $\mathsf{C}^{\mathsf{p}}_{x,y}(\mathit{pay}, \mathit{deliver})$, but as $\mathit{deliver} \vee \mathsf{C}^{\mathsf{p}}_{x,y}(\mathit{pay}, \mathit{deliver})$. The former requires the specified commitment to become active, but the commitment would not exist when it has already been discharged. The latter handles this case. Informally, anyone would be happy if a debtor discharges its commitment: the second formulation, but not the first, captures this intuition. A value of the proposed semantics is in helping to choose among such formulations.

Like any other modality used for modeling agents, commitments are often naturally evaluated with respect to a conceptual frame, such as in a dialogue. Thus the formal model relates not necessarily to objective reality but possibly to a virtual or hypothetical state of affairs. Thus if $\mathsf{C}(r,u)$ arises within an argument and u is established within the same argument, then $\mathsf{C}(r,u)$ would be discharged.

Syntax and Formal Model

$\mathcal{L}$, our formal language, enhances a linear-time logic (Emerson 1990) with modalities C^{p} for practical and C^{d} for dialectical commitments. Below, Φ is a set of atomic propositions and $\mathcal{X}$ is a set of agent names. L and X are nonterminals corresponding to $\mathcal{L}$ and $\mathcal{X}$, respectively.

L_1. $L \longrightarrow \mathit{Commit} \mid \Phi \mid L \wedge L \mid \neg L \mid \mathsf{P}L \mid L\mathsf{U}L$

L_2. $\mathit{Commit} \longrightarrow \mathsf{C}^{\mathsf{d}}_{X,X}(L,L) \mid \mathsf{C}^{\mathsf{p}}_{X,X}(L,L)$

We use the following conventions: x, etc. are agents, ψ, etc. are atomic propositions, p, q, r, etc. are formulae in $\mathcal{L}$, t, etc. are moments, and P, etc. are paths. We drop agent subscripts when they are understood. A model for $\mathcal{L}$ is a tuple, $M = \langle \mathbb{T}, \mathbb{I}, <, \mathbb{D}, \mathbb{C} \rangle$:

- $\mathbb{T}$ is a set of possible moments, each a possible snapshot (i.e., a state) of the world.
- $<\subseteq \mathbb{T} \times \mathbb{T}$ is a discrete linear order on $\mathbb{T}$, which induces paths (contiguous sets of moments) beginning at each moment. Two paths are either disjoint or one is a subset of the other. $[P;t,t']$ denotes a period on path P from t to t'. Formally, $[P;t,t']$ is the intersection of P with the set of moments between t and t', both inclusive. Thus it is possible that $[P;t,t'] = [P';t,t']$ even though $P \neq P'$. $\mathbb{P}$ is the set of all periods and $\mathbb{P}_t$ of periods that begin at t ($\mathbb{P}_t \neq \emptyset$).
- The interpretation, $\mathbb{I}$, of an atomic proposition is the set of moments at which it is true. That is, $\mathbb{I} : \Phi \mapsto \wp(\mathbb{T})$. We show below, through the definition of moment-intension (which, in essence, lifts $\mathbb{I}$ to all propositions), that the denotations of all propositions are sets of moments.
- At each moment, for each proposition, $\mathbb{D}$ assigns a set of propositions to each debtor-creditor (ordered) pair of agents. Each proposition is itself a set of moments. That is, $\mathbb{D} : \mathbb{T} \times \mathcal{X} \times \mathcal{X} \times \wp(\mathbb{T}) \mapsto \wp(\wp(\mathbb{T}))$ yields a set of moments for each moment and proposition.
- At each moment, for each proposition, $\mathbb{C}$ assigns a set of sets of periods to each debtor-creditor (ordered) pair of agents. That is, $\mathbb{C} : \mathbb{T} \times \mathcal{X} \times \mathcal{X} \times \wp(\mathbb{T}) \mapsto \wp(\wp(\mathbb{P}))$ yields a set of periods for each moment and proposition.

Models for modal logics are commonly based on Kripke structures, which define a set of possible worlds along with an accessibility relation that maps each world to a set of worlds. The semantics of the given modal operator tests for *inclusion* in that set of worlds. The models proposed here are *not* Kripke structures and do not involve an accessibility relation. Instead they are based on Segerberg's idea to define a "standard" of correctness by mapping each world to a set of set of worlds. The semantics of the given modal operator tests for *membership* in the set of set of worlds. Segerberg's approach offers greater flexibility in allowing or denying some inferences that the Kripke approach requires.

$\mathbb{D}$ and $\mathbb{C}$ capture the dialectical and practical commitment standards, respectively, for each moment and debtor-creditor pair. Given an antecedent proposition, $\mathbb{D}$ yields a set each of whose members is a consequent proposition (what the debtor would be committed to if the antecedent is met). Likewise, $\mathbb{C}$ yields a set each of whose members is a set of periods, each of which culminates in the consequent proposition. These are the propositions the debtor would bring about. As in many logics of intention and obligation, we don't need to model actions explicitly: $\mathbb{C}$ is simply understood as describing what the debtor would bring about.

Semantics

The semantics of $\mathcal{L}$ is given relative to a model and a moment. $M \models_t p$ expresses "M satisfies p at t." The expression p is *satisfiable* iff for some M and t, $M \models_t p$; p is *valid* iff it is satisfied for all M and t. Formally, we have:

M$_1$. $M \models_t \psi$ iff $t \in \mathbb{I}(\psi)$, where $\psi \in \Phi$

M$_2$. $M \models_t p \wedge q$ iff $M \models_t p$ and $M \models_t q$

M$_3$. $M \models_t \neg p$ iff $M \not\models_t p$

M$_4$. $M \models_t \mathsf{P}q$ iff $(\exists t' : t' \leq t$ and $M \models_{t'} q)$

M$_5$. $M \models_t p\mathsf{U}q$ iff $(\exists t'' : t \leq t''$ and $M \models_{t''} q$ and $(\forall t' : t \leq t' \leq t'' \Rightarrow M \models_{t'} p))$

Disjunction ($\vee$), implication ($\rightarrow$), equivalence ($\equiv$), false, and true are the usual abbreviations. $p\mathsf{U}q$ means "p holds until q": thus $\mathsf{true}\mathsf{U}q$ (abbreviated $\mathsf{F}q$) means "eventually q." $\mathsf{P}q$ means "in the past q."

We define the *moment-intension* of formula p as the set of moments where it is true: $[\![p]\!] = \{t | M \models_t p\}$. We define *period-intension* of formula p as the set of periods culminating in its becoming true: $\langle\!\langle p \rangle\!\rangle = \{[P; t, t'] | M \models_{t'} p\}$. In other words, in these periods, p occurs in the last moment but may possibly occur earlier as well. Thus these are all possible ways in which p may be brought about. Based on these, we can now specify the semantics of commitments.

M$_6$. $M \models_t \mathsf{C}^{\mathsf{d}}_{x,y}(r, u)$ iff $[\![u]\!] \in \mathbb{D}_{x,y}(t, [\![r]\!])$

M$_7$. $M \models_t \mathsf{C}^{\mathsf{p}}_{x,y}(r, u)$ iff $\langle\!\langle u \rangle\!\rangle \in \mathbb{C}_{x,y}(t, [\![r]\!])$

Menu of Reasoning Postulates

Let's now consider several postulates that reflect common reasoning patterns that apply uniformly to practical and dialectical commitments. The idea is to see how they relate to each other, how they are addressed in the literature, and how they are supported by our model theory.

Each postulate below uses C, which can be uniformly substituted by C^{d} or C^{p}. Since these postulates use debtor x and creditor y, we write $\mathsf{C}(r, u)$ instead of $\mathsf{C}_{x,y}(r, u)$. In reading the postulates below, it helps to keep in mind (as the previous section explains) that C refers to an active commitment.

B$_1$. DISCHARGE. $u \rightarrow \neg\mathsf{C}(r, u)$

When u holds, the commitment to bring about u is discharged and is, therefore, no longer *active*. Likewise, when u holds, a dialectical commitment for u is also discharged. Notice that this yields $\neg\mathsf{C}(r, \mathsf{true})$ for any r. Chopra & Singh (C&S) (2004) and WLH incorporate this postulate.

B$_2$. PARTIALLY DETACH. $\mathsf{C}(r \wedge s, u) \wedge r \rightarrow \mathsf{C}(s, u)$

An example is **Ex 1** in the introduction. When "part of" the antecedent of a commitment holds, a commitment for the "remainder" of the antecedent and with the original consequent comes into being.

B$_2$ generalizes over $\mathsf{C}(r, u) \wedge r \rightarrow \mathsf{C}(\mathsf{true}, u)$, i.e., detaching in one shot, which C&S, WLH, and others support.

B$_3$. MONOTONICITY. From $\mathsf{C}(r, u), s \vdash r$ infer $\mathsf{C}(s, u)$

Any commitment that holds for a weaker antecedent also holds for a stronger antecedent. Some useful consequences are $\mathsf{C}(r \vee s, u) \rightarrow \mathsf{C}(r, u)$, $\mathsf{C}(r, u) \rightarrow \mathsf{C}(r \wedge s, u)$, and $\mathsf{C}(\mathsf{true}, u) \rightarrow \mathsf{C}(r, u)$.

Here $p \vdash q$ means we can prove q from p: this is stronger than implication $p \rightarrow q$, which holds merely if p is false. Clearly, $\mathsf{C}(r, u) \wedge \neg s \rightarrow \mathsf{C}(s, u)$ is bogus, i.e., we would not conclude $\mathsf{C}(s, u)$ simply because s happens to be false.

B$_4$. L-DISJOIN. $\mathsf{C}(r, u) \wedge \mathsf{C}(s, u) \rightarrow \mathsf{C}(r \vee s, u)$

The expressions to the left of the $\rightarrow$ mean that x would become committed to u if r or if s hold, which is the meaning of the expression on the right.

B$_5$. R-CONJOIN. $\mathsf{C}(r, u) \wedge \mathsf{C}(r, v) \rightarrow \mathsf{C}(r, u \wedge v)$

An example is **Ex 2** in the introduction. The debtor would become committed to u and to v if r holds, which is the meaning of the commitment on the right.

B$_6$. CONSISTENCY. $\neg\mathsf{C}(r, \mathsf{false})$

An agent cannot commit to false.

B$_7$. STRONG CONSISTENCY. $\mathsf{C}(r, u) \rightarrow \neg\mathsf{C}(r, \neg u)$

This is stronger than B$_6$ in spirit but, as is readily verified, coincides with B$_6$ if we assume B$_5$.

B$_8$. NONVACUITY. From $r \vdash u$ infer $\neg\mathsf{C}(r, u)$

An example is **Ex 3** in the introduction. If r holds so does u. Thus $\mathsf{C}(r, u)$ is discharged as soon as it detaches. In effect, there is no constraint on the debtor. Thus $\mathsf{C}(r, u)$ is vacuous. In particular, because $r \vdash r$, we have $\neg\mathsf{C}(r, r)$.

B$_9$. CHAIN. From $\mathsf{C}(r, u), u \vdash s, \mathsf{C}(s, v)$ infer $\mathsf{C}(r, v)$

This states that commitments are closed in a manner analogous to strict implications.

B$_{10}$. WEAKEN. $\mathsf{C}(r, u \wedge v) \wedge \neg u \rightarrow \mathsf{C}(r, u)$

In general, if you are committed to two propositions you are committed to each one. However, the obvious formulation $\mathsf{C}(r, u \wedge v) \rightarrow \mathsf{C}(r, u)$ is inconsistent with B$_1$, because if u holds, B$_1$ would eliminate $\mathsf{C}(r, u)$. Since B$_1$ is fundamental to capturing discharge, we include $\neg u$ on the left in B$_{10}$.

For example, if the merchant commits to sending both goods and a warranty, then the merchant commits to sending goods unless the goods are already sent. Or, if I am dialectically committed that short-term bond yields are up and interest rates are low, then I am dialectically committed that interest rates are low (unless that is discharged).

B$_{11}$. ELIMINATE ON DETACH. $r \rightarrow \neg\mathsf{C}(r, u)$

$\mathsf{C}(r, u)$ is eliminated when its antecedent (r) is brought about. Although B$_{11}$ might sound natural, it precludes unconditional commitments. Specifically, because true always holds, no unconditional commitment can exist. B$_{11}$ is thus inconsistent with B$_2$: given r and $\mathsf{C}(r, u)$, B$_2$ yields $\mathsf{C}(\mathsf{true}, u)$, but B$_{11}$ entails $\neg\mathsf{C}(\mathsf{true}, u)$.

C&S and WLH capture B$_{11}$ in their operationalizations of commitments. Their intuition is that once a commitment is detached we need not reason about it further. Although we find B_{11} unnatural, we can capture C&S and WLH's operational intuition in an interesting manner. Notice that B_3 yields that $\mathsf{C}(\mathsf{true}, u)$ entails $\mathsf{C}(r, u)$. Thus when B_2 yields $\mathsf{C}(\mathsf{true}, u)$, we can reason further with $\mathsf{C}(\mathsf{true}, u)$ and no longer need to *state* $\mathsf{C}(r, u)$ although it continues to hold.

Corresponding Semantic Constraints

The above model of commitments must be constrained further to support the postulates introduced above. Following van Benthem's (1984) notion of *correspondence theory* for modal logic, we present the constraints separately from the initial framework and in a modular manner. In general, a postulate B corresponds to a constraint C iff B is satisfied in precisely those models that respect C. Thus correspondence clarifies and modularizes the relationship between the semantics and the reasoning. Further, the existence of the correspondence yields a large set of soundness and completeness theorems: one for each subset of the postulates.

Our semantics treats antecedents for both kinds of commitments as sets of moments, the consequents of dialectical commitments as sets of moments, and the consequents of practical commitments as sets of periods. Consequently, several semantic constraints can be stated uniformly in terms of sets and apply equally to both kinds of commitments. However, some constraints are specific to each variety. We include a proof sketch below for one constraint of each type (each proof assumes the usual canonical models).

Below, R and S are sets understood as antecedents; U and V are sets understood as consequents; $\mathbb{B}$ is either $\mathbb{D}$ or $\mathbb{C}$.

Common Constraints

C_1. Upward closure of consequents at a covered moment. Corresponds to B_2. In essence, if $t \in R$, this lifts the constraint imposed by R on the antecedent. Notice that the special case where S equals the universe $\mathbb{T}$ (i.e., $[\![\mathsf{true}]\!]$) yields the truth of the conventional one-shot detach postulate.
$(\forall R, S, t \in R \Rightarrow \mathbb{B}_{x,y}(t, R \cap S) \subseteq \mathbb{B}_{x,y}(t, S))$

Proof of correspondence. Let B_2 hold at moment t. That is, $M \models_t \neg\mathsf{C}(r \wedge s, u) \vee \neg r \vee \mathsf{C}(s, u)$. This holds iff $[\![u]\!] \notin \mathbb{B}_{x,y}(t, [\![r]\!] \cap [\![s]\!])$ or $t \notin [\![r]\!]$ or $[\![u]\!] \in \mathbb{B}_{x,y}(t, [\![s]\!])$. This equals: if $t \in [\![r]\!]$ then if $[\![u]\!] \in \mathbb{B}_{x,y}(t, [\![r]\!] \cap [\![s]\!])$ then $[\![u]\!] \in \mathbb{B}_{x,y}(t, [\![s]\!])$, which means if $t \in [\![r]\!]$ then $\mathbb{B}_{x,y}(t, [\![r]\!] \cap [\![s]\!]) \subseteq \mathbb{B}_{x,y}(t, [\![s]\!])$. Which is equivalent to C_1, as desired.

C_2. Downward closure of consequents. Whatever is committed at a broader antecedent is committed at a narrower antecedent. Corresponds to B_3. In contrast with C_1, C_2 applies even when the current moment does not belong to the antecedent proposition.
$(\forall R, S : \mathbb{B}_{x,y}(t, R \cup S) \subseteq \mathbb{B}_{x,y}(t, R))$

C_3. Lower bound on union of antecedents. Add up the consequents as the antecedents expand.
Corresponds to B_4.
$(\forall R, S : \mathbb{B}_{x,y}(t, R) \cap \mathbb{B}_{x,y}(t, S) \subseteq \mathbb{D}_{x,y}(t, R \cup S))$

C_4. Nonemptiness of consequents. Corresponds to B_6.
$(\forall t, R : \{\,\} \notin \mathbb{B}_{x,y}(t, R))$

C_5. Nondisjointness of consequents. Corresponds to B_7.
$(\forall R, U, V : U, V \in \mathbb{B}_{x,y}(t, R) \Rightarrow U \cap V \neq \{\,\})$

C_6. Closure under chaining through preimage of superset. Corresponds to B_9.
$(\forall t, R, U, V : U \in \mathbb{B}_{x,y}(t, R)$ and $U \subseteq S$ and $V \in \mathbb{B}_{x,y}(t, S) \Rightarrow V \in \mathbb{B}_{x,y}(t, R))$

C_7. No commitments whose antecedents hold. Corresponds to B_{11}. As remarked above, this constraint is not natural.
$(\forall t, R : t \in R \Rightarrow \mathbb{B}_{x,y}(t, R) = \emptyset)$

Dialectical Constraints

D_1. Repulsion. The current moment is not included in any of the propositions selected by $\mathbb{D}$. Thus cannot commit to any currently true consequent. Corresponds to B_1.
$(\forall t, R, U : U \in \mathbb{D}_{x,y}(t, R) \Rightarrow t \notin U)$

Proof of correspondence. Let B_1 hold at moment t. That is, $M \models_t \neg u \vee \neg\mathsf{C}^{\mathsf{d}}_{x,y}(r, u)$. This holds iff $[\![u]\!] \notin \mathbb{D}_{x,y}(t, [\![r]\!])$ or $t \notin [\![u]\!]$. Which is equivalent to D_1, as desired.

D_2. Closure under intersection of consequents. Combine the consequents. Corresponds to B_5.
$(\forall R, U, V : U, V \in \mathbb{D}_{x,y}(t, R) \Rightarrow U \cap V \in \mathbb{D}_{x,y}(t, R))$

D_3. The antecedent is not included in, i.e., does not entail, the consequent. Corresponds to B_8.
$(\forall R, U \in \mathbb{D}_{x,y}(t, R) : R \not\subseteq U)$

D_4. Closure under supersets of moments that don't include the current moment. Corresponds to B_{10}.
$(\forall R, U, V : U \subseteq V, U \in \mathbb{D}_{x,y}(t, R) \Rightarrow V \setminus \{t\} \in \mathbb{D}_{x,y}(t, R))$

Practical Constraints

Define the culmination of a period as its last moment: $\rho_{[P;t,t']} = t'$. Define the culmination ρ_U of a set of periods U as the set of culminations of each period in U. Thus $[\![u]\!] = \rho_{\langle\!|u|\!\rangle}$. Let Π_t be the set of periods culminating in t.

P_1. Repulsion. The current moment does not culminate any of a set of periods selected by $\mathbb{C}$. Corresponds to B_1.
$(\forall S, \forall U \in \mathbb{C}_{x,y}(t, S) : t \notin \rho_U)$

Proof of correspondence. Let B_1 hold at moment t. That is, $M \models_t \neg u \vee \neg\mathsf{C}^{\mathsf{p}}_{x,y}(r, u)$. This holds iff $\langle\!|u|\!\rangle \notin \mathbb{C}_{x,y}(t, [\![r]\!])$ or $t \notin [\![u]\!]$. This holds iff $\langle\!|u|\!\rangle \notin \mathbb{C}_{x,y}(t, [\![r]\!])$ or $t \notin \rho_u$. Which is equivalent to P_1, as desired.

P_2. Closure of consequents under intersection of culminations. Sets of periods culminating in conjunctions of all pairs of consequents exist. Corresponds to B_5.
$(\forall R, U, V : U, V \in \mathbb{C}_{x,y}(t, R) \Rightarrow (\exists W : \rho_W = \rho_U \cap \rho_V$ and $W \in \mathbb{C}_{x,y}(t, R)))$

P_3. The antecedent does not culminate any consequent period, i.e., doesn't entail any consequent proposition. Corresponds to B_8.
$(\forall R, U \in \mathbb{C}_{x,y}(t, R) : R \not\subseteq \rho_U)$

P_4. Closure under supersets of periods that do not culminate in the current moment. Corresponds to B_{10}.
$(\forall R, U, V : U \subseteq V, U \in \mathbb{C}_{x,y}(t, R) \Rightarrow V \setminus \Pi_t \in \mathbb{C}_{x,y}(t, R))$

Soundness and Completeness

As remarked above, the soundness and completeness results fall out from the correspondences presented above. The same postulates apply to C^d and C^p. The common constraints, of course, apply to both C^d and C^p. The theorems below assume substitution instances of propositional and temporal tautologies. The proofs of these theorems follow the representative proof sketches given in the foregoing.

Theorem 1 For C^d, the logic generated by any subset of postulates $\{B_1–B_{11}\}$ is sound and complete with respect to models that satisfy the corresponding constraints: $\{C_1–C_7, D_1–D_4\}$ ∎

Theorem 2 For C^p, the logic generated by any subset of postulates $\{B_1–B_{11}\}$ is sound and complete with respect to models that satisfy the corresponding constraints: $\{C_1–C_7, P_1–P_4\}$ ∎

Discussion

Of the above reasoning postulates, we would select B_1–B_8 and B_{10} as the most natural for the modeling situations we have encountered. We are not convinced of the naturalness of B_9 in general, but it can make sense in some settings. We reject B_{11} because, as explained above, it would eliminate unconditional commitments.

The above formal development enables us to contemplate additional questions regarding commitments. For example, would it be conceivable that we have $\mathsf{C}_{x,y}(r, u)$ and $\mathsf{C}_{y,x}(r, u)$ at the same time? In a competitive setting, this might appear odd. However, it would be a natural situation in teamwork where r corresponds to some externally caused proposition and u represents the team's response. A more typical case of complementary commitments is when $\mathsf{C}_{x,y}(r, u)$ and $\mathsf{C}_{y,x}(u, r)$ hold at the same time. These arise commonly in business protocols, in terms of C^p.

Remarks on Modeling with Commitments

We could formalize commitments with explicit temporal operators. For example, in **Ex 2** we could say ship goods "in the future" and send warranty "in the future." Then their conjunction would mean that goods will be shipped and a warranty will be sent without implying that these occur simultaneously. Requiring a future temporal operator in all cases has three shortcomings. First, it complicates commitments even where "in the future" is implied and modelers would not ordinarily include time in their informal statements. Second, temporal operators would disrupt the illusion—valuable during early modeling—that dialectical and practical commitments are alike. For example, in designing a financial trade protocol, we can first state that a quote creates a commitment and only later determine if it is indicative of the current price (dialectical) or an offer to sell (practical). Third, explicit temporal operators are unnecessary because *multiagent* systems are inherently distributed. Consequently, specifications that rely upon transient conditions require synchronizing the agents and are difficult to implement reliably.

An easy way to make commitments insensitive to arbitrary ordering or delays of messages would be to limit ourselves to *stable* propositions. A proposition p is stable provided that if p becomes true, it stays true. That is, $q \equiv \mathsf{P}q$. Past propositions such as "the payment has been made" and future-past propositions such as "the package will have been delivered" are stable: because $\mathsf{P}q \equiv \mathsf{PP}q$ and $\mathsf{FP}q \equiv \mathsf{PFP}q$, respectively. (In English grammar, these exemplify the *perfect* aspect.) We have observed that modelers, especially those with a distributed computing background as in the finance industry, often assume they are dealing with stable propositions. Our approach is not limited to stable propositions, however.

Relevant Work on Commitments and Semantics

Some papers base communication semantics on commitments but don't give a semantics to commitments. Amgoud *et al.* (2002) use a commitment store as an abstraction to operationally specify the meanings of communicative acts as effects on the store, but don't formalize commitments. Likewise, Flores *et al.* (2007) characterize the operational semantics for conversations in Object Z. These approaches show the potential payoff to understanding communications from formalizing commitments.

Y&S treat unconditional commitments as basic and employ a strict implication for conditional (practical) commitments. Y&S do not offer an axiomatization. WLH expand on Y&S's formalization of protocols but use a similar definition of commitments. These and other works adopt B_1, a one-shot form of B_2 (see the discussion of B_2), and B_{11}. As explained above, B_{11} is not consistent with B_2 or its one-shot version. Thus, the above works ought to have sacrificed B_{11} or framed it as a reasoning convenience, not as a fundamental postulate.

V&C formalize commitments in temporal logic. They treat the various commitment states (fulfilled, violated, ...) as predicates. V&C give neither a model theory nor a sound and complete set of reasoning postulates (i.e., axiomatization or inference rules). They argue informally for some postulates (most clearly for one that resembles B_{10}).

Bentahar *et al.* (2005) develop a modal approach for dialectical and practical commitments. Their main emphasis is on how commitments are updated in conversation. They state several reasoning postulates but do not pursue soundness and completeness.

Khan and Lespérance (K&L) (2006) treat commitments as intentions. This is not quite correct because commitments are a social artifact whereas intentions are a mental artifact. However, the logical structure of the two is similar enough.

K&L criticize Y&S's approach as being counterintuitive, because an agent who is conditionally committed could in principle try to make the antecedent stay false. In their example, an agent who conditionally commits to paying upon receiving the goods may not be committed to receiving the goods. This is not counterintuitive, however, because it does not excessively constrain the debtor: the agent would commit to the purchase only if it desires the goods. Conversely, there may be a commitment such as "if fail then compensate," where trying to avoid failure is indeed desirable. K&L

seem to require that a conditional commitment include a commitment to the antecedent, which is clearly not appropriate for failure.

K&L's definition does not support $\mathsf{C}^{\mathsf{p}}(\mathsf{true}, u) \rightarrow \mathsf{C}^{\mathsf{p}}u$ ($\mathsf{C}^{\mathsf{p}}u$ is an unconditional commitment for u). This is unintuitive, because it suggests that a conditional commitment whose antecedent is necessarily met is not as strong as a corresponding unconditional commitment. By contrast, WLH and Y&S support, and we define, $\mathsf{C}^{\mathsf{p}}(\mathsf{true}, u) \equiv \mathsf{C}^{\mathsf{p}}u$.

Further, K&L assume that if the agent intends that ϕ never holds, then ϕ never holds: the agent must exercise perfect control over ϕ. Consider a vendor who creates a commitment meaning that if it fails (in its main contracted task), then it would issue a refund. Formally, we would have $\phi = \mathit{failure}$ and $\psi = \mathit{refund}$. The above definition requires that the vendor cannot intend not to fail unless it can guarantee success. In the latter case, the conditional commitment would be moot. In general, under the K&L semantics, there would be no need for a conditional commitment.

Directions: Conceptual, Theoretical, Practical

Conceptually. A deeper study of the kinds of reasoning postulates that would be beneficial in a wider range of applications than have been studied to date. In particular, additional postulates would be valuable that involve combinations of dialectical and practical commitments, and composite patterns of usage based on the operations on commitments. Also, the logical aspects of commitments studied here would need to be augmented with specific timestamps and durations, as needed in many applications.

Theoretically. Stronger results on the reasoning and models of commitments would enable advanced applications. As an example of a useful result, consider possible invariants on commitments over a class of reasoning postulates, including those supporting actions. Informally, define a commitment as yielding another commitment if the first leads to the second via a sequence of zero or more reasoning steps. Say the discharge of the latter entails the discharge of the former. Can we then show that two commitments that yield each other are semantically equivalent? Solving that problem would open up the way to understanding both dialogs and business protocols at a deeper level, and would support emerging notions of interoperability and compliance (Chopra & Singh 2008; Desai & Singh 2008).

Practically. A rigorous semantics opens up the way for improved verification and design tools. An immediate step in this regard would be to map the commitment semantics introduced here to conventional verification technologies. Efforts, e.g., by Giunchiglia *et al.* (2002), in applying SAT solvers to classical modal logics are promising. The postulates of our logic are different from those considered by Giunchiglia *et al.*, especially B_1, but there is no fundamental reason why an approach analogous to theirs could not be developed for commitments.

Our semantics is not based on Kripke structures. Currently popular model checking techniques work best for logics whose semantics is given via Kripke structures (or analogously, as interpreted systems) (Lomuscio & Penczek 2007). Two complementary directions suggest how this lacuna might be addressed. One, model checking techniques for our logic can be developed. Two, when the postulates such as those given at the beginning of this section are assumed, an equivalent Kripke semantics could be formulated.

Thanks to Amit Chopra, Nirmit Desai, Pınar Yolum, and the reviewers for helpful comments.

References

Amgoud, L., Maudet, N., and Parsons, S. 2002. An argumentation-based semantics for agent communication languages. *Proc. ECAI*, 38–42.

Bentahar, J., Moulin, B., Meyer, J.-J. C., and Chaib-draa, B. 2005. A modal semantics for an argumentation-based pragmatics for agent communication. *Proc. ArgMAS*.

Castelfranchi, C. 1995. Commitments: From individual intentions to groups and organizations. *ICMAS*, 41–48.

Chopra, A. K. and Singh, M. P. 2004. Nonmonotonic commitment machines. Proc. ACL, *LNAI* 2922, 183–200.

Chopra, A. K. and Singh, M. P. 2008. Constitutive interoperability. Proc. AAMAS. To appear.

Desai, N., Chopra, A. K., Arrott, M., Specht, B., and Singh, M. P. 2007. Engineering foreign exchange processes via commitment protocols. *Proc. SCC*, 514–521.

Desai, N. and Singh, M. P. 2008. On the enactability of business protocols. Proc. AAAI. To appear.

Emerson, E. A. 1990. Temporal and modal logic. *Handbook of Theoretical Computer Science*, vol. B. 995–1072.

Flores, R. A., Pasquier, P., and Chaib-draa, B. 2007. Conversational semantics sustained by commitments. *Autonomous Agents and Multi-Agent Systems* 14(2):165–186.

Giunchiglia, E., Tacchella, A., and Giunchiglia, F. 2002. SAT-based decision procedures for classical modal logics. *Journal of Automated Reasoning* 28(2):143–171.

Khan, S. M. and Lespérance, Y. 2006. On the semantics of conditional commitment. *Proc. AAMAS*, 1337–1344.

Lomuscio, A. and Penczek, W. 2007. Symbolic model checking for temporal-epistemic logics. *ACM SIGACT News* 38(3):77–99.

Singh, M. P. 1991. Social and psychological commitments in multiagent systems. *AAAI Fall Symp. Knowledge and Action at Social and Organizational Levels* 104–106.

Singh, M. P. 1999. An ontology for commitments in multiagent systems. *Artificial Intelligence & Law* 7:97–113.

van Benthem, J. F. A. K. 1984. Correspondence theory. *Handbook of Philosophical Logic*, volume II. 167–247.

Verdicchio, M. and Colombetti, M. 2003. A logical model of social commitment for agent communication. *AAMAS*.

Norman, T., Carbogim, D., Krabbe, E., and Walton, D. 2004. Argument and multi-agent systems. Reed C. and Norman T., eds., *Argumentation Machines*, ch. 2. Kluwer.

Winikoff, M., Liu, W., and Harland, J. 2005. Enhancing commitment machines. *Proc. DALT*, *LNAI* 3476, 198–220.

Yolum, P. and Singh, M. P. 2002. Commitment machines. *Proc. ATAL*, *LNAI* 2333, 235–247.