A Causal Theory of Speech Acts

Chiaki Sakama

Department of Computer and Communication Sciences
Wakayama University
930 Sakaedani, Wakayama 640-8510, Japan
sakama@sys.wakayama-u.ac.jp

Abstract. In speech acts, a speaker utters sentences that might affect the belief state of a hearer. To formulate causal effects in assertive speech acts, we introduce a logical theory that encodes causal relations between speech acts, belief states of agents, and truth values of sentences. We distinguish trustful and untrustful speech acts depending on the truth value of an utterance, and distinguish truthful and untruthful speech acts depending on the belief state of a speaker. Different types of speech acts cause different effects on the belief state of a hearer, which are represented by the set of models of a causal theory. Causal theories of speech acts are also translated into logic programs, which enables one to represent and reason about speech acts in answer set programming.

1 Introduction

An *assertive speech act* commits a speaker to the truth of the expressed proposition [5]. Although assertive sentences are either true or false, a speaker generally does not have complete knowledge of the world. It may happen that a speaker utters a sentence that is believed to be true by herself while it is actually false. In this case, a speaker acts truthfully but a hearer would consider the speaker untrustful. On the other hand, there is a case that a speaker utters a sentence disbelieved by himself while the statement happens to be true. In this case, a speaker acts untruthfully but a hearer would consider the speaker trustful. Whether a speech act is truthful or not depends on the belief state of a speaker, while whether a speech act is trustful or not is judged by the truth of information conveyed by the utterance. At this point, four different combinations of speech acts are considered—(i) both truthful and trustful, (ii) truthful but untrustful, (iii) untruthful but trustful, and (iv) both untruthful and untrustful.

In this paper, we distinguish truthful/untruthful and trustful/untrustful speech acts and consider their performative effects on hearers. Different from previous studies based on modal logic [2, 3] or dynamic epistemic logic [6], we formulate assertive speech acts using a *causal logic* introduced in [4]. The logic can simply encode causal relations in speech acts and a causal theory is implemented by logic programming.

2 Causal Theory

We first review a causal logic of [4]. Let $\mathcal{L}$ be a language of propositional logic and $\mathcal{A}$ a finite set of *atoms* in the language. *Formulas* (or *sentences*) in $\mathcal{L}$ are defined as

follows: (i) If $p \in \mathcal{A}$, then p is a formula. (ii) If φ and ψ are formulas, then $\neg\varphi,\ \varphi \wedge \psi,\ \varphi \vee \psi,\ \varphi \supset \psi,$ and $\varphi \equiv \psi$ are all formulas. In particular, $\top$ and $\bot$ represent valid and contradictory formulas, respectively. We often use parentheses "()" in a formula as usual. A finite set T of formulas is identified with the conjunction of all formulas in T. The set of all formulas in $\mathcal{L}$ is represented by $\mathcal{F}$. A *literal* is an atom A or its negation $\neg A$. An *interpretation* I is a complete and consistent (finite) set of literals, i.e., $L \in I$ iff $\neg L \notin I$ for any literal L appearing in a theory. A literal L is *true* in an interpretation I iff $L \in I$. The truth value of a formula φ in I is defined based on the usual truth tables of propositional connectives. An interpretation I *satisfies* a formula φ (written $I \models \varphi$) iff φ is true in I. Given formulas φ and ψ, $\varphi \Rightarrow \psi$ is called a *causal rule* meaning that "ψ is caused if φ is true." In particular, the rule $(\top \Rightarrow \psi)$ is a *fact* representing that ψ is true, and $(\varphi \Rightarrow \bot)$ is a *constraint* representing that φ cannot be true.

A *causal theory* is a finite set of causal rules. Given a causal theory T and an interpretation I, define $T^I = \{\, \psi \mid (\varphi \Rightarrow \psi) \in T \text{ for some } \varphi \text{ and } I \models \varphi \,\}$. Then I is a *model* of T if and only if $I = \{\, L \mid T^I \models L \,\}$ where $T^I \models L$ means that T^I entails L in classical logic. We say that I *satisfies* every rule in T if I is a model of T. By definition, I is a model of T iff I is the unique model of T^I. A causal theory T is *consistent* if it has a model; otherwise, T is *inconsistent*. Actions and their effects are represented by a causal theory as follows. First, atoms of the language are expressions of the forms: a_t and f_t where a, f and t are action, fluent, and time names, respectively. a_t means that an action a occurs at time t, and f_t means that a fluent f holds at time t. In this paper we consider actions as assertive speech acts by an agent. An utterance of a sentence σ by an agent x at time t is represented by the atom $Utter_t(x, \sigma)$.[1] The truth value of a sentence is represented by a fluent. When a sentence σ is true (resp. false) at time t, we write $Hold_t(\sigma)$ (resp. $\neg Hold_t(\sigma)$). A belief state of an agent is also represented by a fluent. When an agent x believes (resp. disbelieves) a sentence σ at time t, it is represented by the literal $Bel_t(x, \sigma)$ (resp. $\neg Bel_t(x, \sigma)$). We define $Hold_t(\top) \equiv Bel_t(x, \top) \equiv \top$, $Hold_t(\bot) \equiv Bel_t(x, \bot) \equiv \bot$ and $Hold_t(\neg\sigma) \equiv \neg Hold_t(\sigma)$ for any x, σ and t.

A causal theory must specify conditions that are sufficient for every fact to be caused. To this end, a causal theory of action contains *action rules*: $(a_t \Rightarrow a_t)$ and $(\neg a_t \Rightarrow \neg a_t)$ which represent that the occurrence (resp. non-occurrence) of an action a at a time t is caused whenever a occurs (resp. does not occur) at t. Likewise, to explain facts at the moment t when a fluent f comes into existence, a causal theory of action contains *fluent rules*: $(f_t \Rightarrow f_t)$ and $(\neg f_t \Rightarrow \neg f_t)$ which represent that the state of a fluent f at a time t is determined outside the theory. Finally, fluents that do not change by an action are represented by *inertia rules*: $(f_t \wedge f_{t+1} \Rightarrow f_{t+1})$ and $(\neg f_t \wedge \neg f_{t+1} \Rightarrow \neg f_{t+1})$ which represent that if the truth value of f at time t is identical with the value at time $t+1$ then the truth value at time $t+1$ is caused by virtue of its persistence. Note that $\Rightarrow$ is not identical to material implication in classical logic. In fact, the above rules become tautologies if $\Rightarrow$ is replaced by $\supset$. Those rules are *not* tautologies in causal logic. For notational convenience, we use $L^{\pm}$ meaning L or $\neg L$. For instance, $(a_t^{\pm} \Rightarrow a_t^{\pm})$ means $(a_t \Rightarrow a_t)$ and $(\neg a_t \Rightarrow \neg a_t)$, and $(f_t^{\pm} \wedge f_{t+1}^{\pm} \Rightarrow f_{t+1}^{\pm})$ means $(f_t \wedge f_{t+1} \Rightarrow f_{t+1})$ and $(\neg f_t \wedge \neg f_{t+1} \Rightarrow \neg f_{t+1})$.

[1] $Utter_t(x, \sigma)$ is represented by the proposition $utter\_x\_\sigma_t$ where $utter\_x\_\sigma$ is an action name. We write $utter\_x\_\sigma_t$ by $Utter_t(x, \sigma)$ for notational convenience.

3 Causal Theories of Speech Acts

Definition 3.1 (causal theory of speech acts) Let x be an agent, $\sigma \in \mathcal{F}$, and t a parameter representing time. A *causal theory of speech acts* $\mathcal{CT}^t_{x\sigma}$ consists of rules:

$$Utter^{\pm}_t(x,\sigma) \Rightarrow Utter^{\pm}_t(x,\sigma), \tag{1}$$
$$Hold^{\pm}_t(\sigma) \Rightarrow Hold^{\pm}_t(\sigma), \tag{2}$$
$$Hold^{\pm}_t(\sigma) \wedge Hold^{\pm}_{t+1}(\sigma) \Rightarrow Hold^{\pm}_{t+1}(\sigma), \tag{3}$$
$$Bel^{\pm}_t(x,\sigma) \Rightarrow Bel^{\pm}_t(x,\sigma), \tag{4}$$
$$Bel^{\pm}_t(x,\sigma) \wedge Bel^{\pm}_{t+1}(x,\sigma) \Rightarrow Bel^{\pm}_{t+1}(x,\sigma). \tag{5}$$

The rules (1) are action rules, the rules (2) and (4) are fluent rules, and the rules (3) and (5) are inertia rules. The rules (1) represent that an agent x utters (or does not utter) a sentence σ at time t. (2) represent that a sentence σ is true (or false) at time t. (4) represent that an agent x believes (or disbelieves) a sentence σ at time t.

A speech act by an agent is *trustful* (resp. *untrustful*) if the agent utters a true (resp. *false*) sentence. They are represented by causal theories as follows.

Definition 3.2 (trustful/untrustful speech acts) A *trustful* (or *untrustful*) *speech act* of a sentence σ by an agent a at time t is defined as follows.

$$\textbf{Trustful}(a,\sigma,t) := \mathcal{CT}^t_{a\sigma} \cup \{\ Utter_t(a,\sigma) \wedge \neg\, Hold_t(\sigma) \Rightarrow \bot\ \}.$$
$$\textbf{Untrustful}(a,\sigma,t) := \mathcal{CT}^t_{a\sigma} \cup \{\ Utter_t(a,\sigma) \wedge\ Hold_t(\sigma) \Rightarrow \bot\ \}.$$

By contrast, a speech act by an agent is *truthful* (resp. *untruthful*) if the agent utters a sentence believed (resp. *disbelieved*) to be true.

Definition 3.3 (truthful/untruthful speech acts) A *truthful* (or *untruthful*) *speech act* of a sentence σ by an agent a at time t is defined as follows.

$$\textbf{Truthful}(a,\sigma,t) := \mathcal{CT}^t_{a\sigma} \cup \{\ Utter_t(a,\sigma) \wedge \neg\, Bel_t(a,\sigma) \Rightarrow \bot\ \}.$$
$$\textbf{Untruthful}(a,\sigma,t) := \mathcal{CT}^t_{a\sigma} \cup \{\ Utter_t(a,\sigma) \wedge Bel_t(a,\sigma) \Rightarrow \bot\ \}.$$

Whether a speech act is trustful or not is determined by the truth of an utterance, while whether a speech act is truthful or not is determined by the belief state of a speaker. Any logical combination of trustfulness and truthfulness is consistent, for instance, $\textbf{Trustful}(a,\sigma,t) \wedge \textbf{Untruthful}(a,\sigma,t)$ has the model:[2] $\{\ U_t(a,\sigma),\ \neg B_t(a,\sigma),\ \neg B_{t+1}(a,\sigma),\ H_t(\sigma),\ H_{t+1}(\sigma)\ \}$ which represents that a speaker a utters a disbelieved sentence σ that happens to be true. $\textbf{Untrustful}(a,\sigma,t) \wedge \textbf{Truthful}(a,\sigma,t)$ has the model $\{\ U_t(a,\sigma),\ B_t(a,\sigma),\ B_{t+1}(a,\sigma),\ \neg H_t(\sigma),\ \neg H_{t+1}(\sigma)\ \}$ which represents that a speaker a utters a believed-true sentence σ that is in fact false.

Next we consider the effect of a speech act on a hearer. Suppose that a speaker a utters a sentence σ at time t, which brings about a hearer b's believing σ at time $t+1$. It is represented by the causal rule:

$$Utter_t(a,\sigma) \Rightarrow Bel_{t+1}(b,\sigma). \tag{6}$$

[2] U means *Utter*, B means *Bel*, and H means *Hold*.

On the hearer's side, she would believe an utterance only when it is consistent with her own belief. The situation is represented by the constraint:

$$Bel_t(b, \neg\sigma) \wedge Bel_t(b, \sigma) \Rightarrow \bot. \tag{7}$$

Prepare rules (4) and (5) for b and $\neg\sigma$, and put them together with (6) and (7). Let $\Delta^t_{ab\sigma} = \{\, Bel^{\pm}_t(b,\sigma) \Rightarrow Bel^{\pm}_t(b,\sigma),\ \ Bel^{\pm}_t(b,\sigma) \wedge Bel^{\pm}_{t+1}(b,\sigma) \Rightarrow Bel^{\pm}_{t+1}(b,\sigma),$
$Bel^{\pm}_t(b,\neg\sigma) \Rightarrow Bel^{\pm}_t(b,\neg\sigma),\ \ Bel^{\pm}_t(b,\neg\sigma) \wedge Bel^{\pm}_{t+1}(b,\neg\sigma) \Rightarrow Bel^{\pm}_{t+1}(b,\neg\sigma),$
$Utter_t(a,\sigma) \Rightarrow Bel_{t+1}(b,\sigma),\ \ Bel_\tau(b,\neg\sigma) \wedge Bel_\tau(b,\sigma) \Rightarrow \bot$ (for $\tau = t,\, t+1$) $\}$.

Definition 3.4 ((mis)inform / (in)sincere) Let a and b be two agents, $\sigma \in \mathcal{F}$, and t a parameter representing time. Then define

$$\textbf{Inform}(a,b,\sigma,t) := \textbf{Trustful}(a,\sigma,t) \ \cup\ \Delta^t_{ab\sigma}\,.$$
$$\textbf{Misinform}(a,b,\sigma,t) := \textbf{Untrustful}(a,\sigma,t) \ \cup\ \Delta^t_{ab\sigma}\,.$$
$$\textbf{Sincere}(a,b,\sigma,t) := \textbf{Truthful}(a,\sigma,t) \ \cup\ \Delta^t_{ab\sigma}\,.$$
$$\textbf{Insincere}(a,b,\sigma,t) := \textbf{Untruthful}(a,\sigma,t) \ \cup\ \Delta^t_{ab\sigma}\,.$$

If the speech act is trustful (resp. untrustful), the speaker brings true (resp. false) information to the hearer. In this case, we say that a *informs* (resp. *misinforms*) b of σ. On the other hand, if the speech act is truthful (resp. untruthful) a speaker a communicates a believed-true (resp. disbelieved) sentence σ to a hearer b. In this case, we say that a *sincerely* (resp. *insincerely*) *communicates* σ to b. The effect of an utterance is observed, for instance, by comparing two models: $M_1 = \{\, U_t(a,\sigma),\ \neg B_t(b,\sigma),\ B_{t+1}(b,\sigma)\} \cup N$ and $M_2 = \{\, \neg U_t(a,\sigma),\ \neg B_t(b,\sigma),\ \neg B_{t+1}(b,\sigma)\} \cup N$ of **Inform**(a,b,σ,t) where $N = \{B_t(a,\sigma),\ B_{t+1}(a,\sigma),\ \neg B_t(b,\neg\sigma),\ \neg B_{t+1}(b,\neg\sigma),\ H_t(\sigma),\ H_{t+1}(\sigma)\,\}$. When a hearer b disbelieves the sentence σ at time t, an utterance of σ changes the belief state of the hearer at time $t+1$ as far as b disbelieves $\neg\sigma$. Such a belief change does not happen if b believes $\neg\sigma$ at t. When a speaker a utters a believed-true sentence σ that is actually false, it would *mislead* a hearer b's acquiring the false belief σ. The situation is represented by the model $\{\, U_t(a,\sigma),\ B_t(a,\sigma),\ \neg B_t(b,\sigma),\ \neg B_t(b,\neg\sigma),\ \neg H_t(\sigma),\ B_{t+1}(a,\sigma),$ $B_{t+1}(b,\sigma),\ \neg B_{t+1}(b,\neg\sigma),\ \neg H_{t+1}(\sigma)\,\}$ of **Sincere**(a,b,σ,t). On the other hand, if a speaker a utters a disbelieved sentence σ that is actually false and it causes a hearer b's acquiring the false belief σ, the speaker *deceives* the hearer. The situation is represented by the model $\{\, U_t(a,\sigma),\ \neg B_t(a,\sigma),\ \neg B_t(b,\sigma),\ \neg B_t(b,\neg\sigma),\ \neg H_t(\sigma),\ \neg B_{t+1}(a,\sigma),$ $B_{t+1}(b,\sigma),\ \neg B_{t+1}(b,\neg\sigma),\ \neg H_{t+1}(\sigma)\,\}$ of **Insincere**(a,b,σ,t).

Some formal properties are addressed as follows.

Proposition 3.1 *Let a and b be two agents, $\sigma \in \mathcal{F}$, and t a parameter representing time. Also let* **Comm** *be either* **Inform**, **Misinform**, **Sincere**, *or* **Insincere**. *It holds that*

(i) **Trustful**$(a,\sigma,t) \wedge$ **Untrustful**$(a,\sigma,t) \supset \neg Utter_t(a,\sigma)$.
(ii) **Truthful**$(a,\sigma,t) \wedge$ **Untruthful**$(a,\sigma,t) \supset \neg Utter_t(a,\sigma)$.
(iii) **Trustful**$(a,\sigma,t) \wedge$ **Truthful**$(a,\sigma,t) \supset (\,Utter_t(a,\sigma) \supset Hold_t(a,\sigma) \wedge Bel_t(a,\sigma))$.
(iv) **Trustful**$(a,\sigma,t) \wedge$ **Untruthful**(a,σ,t)
$\supset (\,Utter_t(a,\sigma) \supset Hold_t(a,\sigma) \wedge \neg Bel_t(a,\sigma))$.
(v) **Untrustful**$(a,\sigma,t) \wedge$ **Truthful**(a,σ,t)
$\supset (\,Utter_t(a,\sigma) \supset \neg Hold_t(a,\sigma) \wedge Bel_t(a,\sigma))$.

(vi) **Untrustful**$(a, \sigma, t) \wedge$ **Untruthful**(a, σ, t)
$$\supset (Utter_t(a, \sigma) \supset \neg Hold_t(a, \sigma) \wedge \neg Bel_t(a, \sigma)).$$
(vii) **Comm**$(a, b, \sigma, t) \supset (Utter_t(a, \sigma) \wedge Bel_{t+1}(b, \sigma) \supset \neg Bel_t(b, \neg\sigma))$.
(viii) **Comm**$(a, b, \sigma, t) \supset (Utter_t(a, \sigma) \wedge \neg Bel_{t+1}(b, \sigma) \supset Bel_t(b, \neg\sigma))$.

4 Encoding in Logic Programs

A causal rule $L_1 \wedge \cdots \wedge L_n \Rightarrow L_0$ where L_i $(0 \leq i \leq n)$ is a literal, is translated into the logic programming rule: $L_0 \leftarrow not\, \neg L_1, \ldots, not\, \neg L_n$ where not represents *negation as failure*. Let Π_T be the logic program that is obtained from a causal theory T by translating each causal rule in T into the logic programming rule in Π_T. Then an interpretation I is a model of T iff I is a consistent and complete *answer set* of Π_T [4]. By this fact, a causal theory of speech acts is represented by a logic program as follows.

Definition 4.1 (logic program of speech acts) Let $\mathcal{CT}^t_{x\sigma}$ be a causal theory of speech acts of Def. 3.1. Then a *logic program $\Pi^t_{x\sigma}$ associated with $\mathcal{CT}^t_{x\sigma}$* consists of rules:

$$Utter_t(x, \sigma) \leftarrow not\, \neg Utter_t(x, \sigma), \qquad \neg Utter_t(x, \sigma) \leftarrow not\, Utter_t(x, \sigma),$$
$$Hold_t(\sigma) \leftarrow not\, \neg Hold_t(\sigma), \qquad \neg Hold_t(\sigma) \leftarrow not\, Hold_t(\sigma),$$
$$Hold_{t+1}(\sigma) \leftarrow not\, \neg Hold_t(\sigma),\ not\, \neg Hold_{t+1}(\sigma),$$
$$\neg Hold_{t+1}(\sigma) \leftarrow not\, Hold_t(\sigma),\ not\, Hold_{t+1}(\sigma),$$
$$Bel_t(x, \sigma) \leftarrow not\, \neg Bel_t(x, \sigma), \qquad \neg Bel_t(x, \sigma) \leftarrow not\, Bel_t(x, \sigma),$$
$$Bel_{t+1}(x, \sigma) \leftarrow not\, \neg Bel_t(x, \sigma),\ not\, \neg Bel_{t+1}(x, \sigma),$$
$$\neg Bel_{t+1}(x, \sigma) \leftarrow not\, Bel_t(x, \sigma),\ not\, Bel_{t+1}(x, \sigma).$$

By the correspondence between a causal theory T and its logic programming translation Π_T, we have the next result.

Proposition 4.1 *Let $\Pi^t_{x\sigma}$ be a logic program associated with a causal theory $\mathcal{CT}^t_{x\sigma}$. Then, I is a model of $\mathcal{CT}^t_{x\sigma}$ iff I is an answer set of $\Pi^t_{x\sigma}$.*

(Un)trustful or (un)truthful speech acts, and (mis)inform or (in)sincere communication are also represented by logic programs. In this way, we can compute the effect of assertive speech acts in *answer set programming* [1].

References

1. Brewka, G., Eiter, T., Truszczyński, M.: Answer set programming at a glance. CACM 54, pp. 92–103 (2011)
2. Cohen, P. R., Levesque, H.: Speech acts and rationality. In: Proc. 23rd Annual Meeting of ACL, pp. 49–59 (1985)
3. Demolombe, R.: Reasoning about trust: a formal logic framework. In: Proc. 2nd Int. Conf. Trust Management, LNCS, vol. 2995, pp. 291–303, Springer (2004)
4. Giunchiglia, E., Lee, J., Lifschitz, V. McCain, N., Turner, H.: Nonmonotonic causal theories. Artif. Intell. **153**:49–104 (2004)
5. Searle, J. R.: Expression and Meaning. Cambridge University Press (1979)
6. Yamada, T.: Logical dynamics of some speech acts that affect obligations and preferences. Synthese **165**:295–315 (2008)