A Logic of Trust and Reputation

Emiliano Lorini
(joint work with A. Herzig, J. F. Hübner, L. Vercouter)

IRIT-CNRS, Toulouse, France

Workshop on Trust, Advice, and Reputation
Toulouse, 22-24 October 2009
Motivation: formal analysis of trust and reputation

- Many computational models [Sabater & Sierra 02; Huynh et al. 04]
  - Quantitative
  - No qualitative analysis
- Few logical models
  - Mostly focused on trust in information sources [Liau 03; Jones & Firozabadi 01; Dastani et al. 04]
- Cognitive model of trust [Castelfranchi & Falcone 98, 01]
  - Informal definitions

⇒ Formal definitions?
⇒ Relations between trust and reputation?
Contribution: definitions in a BDI logic


- ‘top-down’ qualitative analysis: reduction of trust to more primitive concepts
  - trust \( \Rightarrow \) belief, goal, intention, action
  - analysis in terms of logics of action and time, BDI logics
- two versions of trust
  - ‘occurrent trust’ vs. ‘dispositional trust’
  - refinement of Castelfranchi & Falcone’s definition
- from dispositional trust to reputation
Trust: the ingredients [Castelfranchi & Falcone]

“i trusts j to do $\alpha$ in order to achieve $\varphi$”

- truster $i$,
- trustee $j$,
- action $\alpha$ of $j$,
- goal $\varphi$ of $i$.

- important: trust $\neq$ thinking and foreseeing
Informal definition [Castelfranchi & Falcone]

⇒ trust ≈ evaluation of the truster about certain properties of trustee (that are relevant for a goal/task \( \varphi \))

“\( i \) trusts \( j \) to do \( \alpha \) in order to achieve \( \varphi \)” if and only if:

1. \( i \) has the goal \( \varphi \);
2. \( i \) believes that
   1. \( j \) is capable to do \( \alpha \);
   2. \( j \) has the power to achieve \( \varphi \) by doing \( \alpha \);
   3. \( j \) intends to do \( \alpha \).

⇒ trust defined from four more primitive concepts

⇒ logic of goal, ability, power and intention?
1. Two kinds of trust: occurrent vs. dispositional

2. Logical formalization

3. From trust to reputation

4. Conclusion
Two kinds of trust: occurrent vs. dispositional
Occurrent trust vs. dispositional trust

two perspectives on trustee’s action $\alpha$:

- truster believes trustee is going to do $\alpha$ here and now
  $\Rightarrow$ **occurrent trust**

- truster believes trustee is going to do $\alpha$ whenever some conditions are satisfied.
  $\Rightarrow$ **dispositional trust**

(cf. occurrent belief vs. dispositional belief [Searle 92])
Occurrent trust

$i$ wants $\varphi$ to be true and believes $j$ is going to perform $\alpha$ here and now so that $\varphi$ will be ensured

$$\text{OccTrust}(i, j, \alpha, \varphi) \overset{\text{def}}{=} \text{OccGoal}(i, \varphi) \land \text{Belief}(i, \text{OccCap}(j, \alpha) \land \text{OccPower}(j, \alpha, \varphi) \land \text{OccIntends}(j, \alpha))$$

$\Rightarrow$ predicates to be defined: $\text{Belief, OccGoal, OccCap, OccPower, OccIntends}$
Example

1’s *occurrent trust* in 2 to send a certain product in view of satisfying 1’s goal of possessing the product:

- 1 wants to possess the product,
- 1 believes that
  - 2 is capable to send the product,
  - 2’s act of sending the product will result in 1 possessing it,
  - 2 has the intention to send the product.
Dispositional trust

\(i\) expects that there will be some situations \(\kappa\) in which he will have the goal of achieving \(\varphi\) (potential goal), and believes that in all these situations \(j\) will ensure \(\varphi\) by doing action \(\alpha\).

\[
\text{DispTrust}(i, j, \alpha, \varphi, \kappa) \overset{\text{def}}{=} \text{PotGoal}(i, \varphi, \kappa) \land \\
\text{Belief}(i, \text{Henceforth}) \land \\
((\kappa \land \text{OccGoal}(i, \varphi)) \rightarrow \\
(\text{OccCap}(j, \alpha) \land \\
\text{OccPower}(j, \alpha, \varphi) \land \\
\text{OccIntends}(j, \alpha)))
\]

\(\Rightarrow\) predicates to be defined: \textit{PotGoal, Belief, Henceforth, OccGoal, OccCap, OccPower, OccIntends}
Example

1’s *dispositional trust* in 2 (a mechanic) to repair his car so that the car will be in order, in the circumstances in which 1 will ask the mechanic to repair his car.
Logical formalization
Occurrent trust and dispositional trust: components

- **definiendum:**
  - belief,
  - occurrent goal,
  - occurrent capability,
  - occurrent power,
  - occurrent intention.

- **definiens:** formulas of a modal logic of time, belief, preference, and action
  - ‘spell out’ predicates $\text{Belief}(i, \varphi)$, $\text{OccGoal}(i, \varphi)$, $\text{PotGoal}(i, \varphi, \kappa)$, $\text{OccCap}(j, \alpha)$, $\text{OccPower}(j, \alpha, \varphi)$, $\text{OccIntends}(j, \alpha)$ using modal operators of time, belief, preference, and action

**convention:**
- $\text{Bel}_i$, etc. = logical operators
Temporal operators

Henceforth $\varphi$ = “$\varphi$ henceforth holds”

Eventually $\varphi$ $\overset{\text{def}}{=} \neg$Henceforth $\neg \varphi$

Eventually $\varphi$ = “$\varphi$ eventually holds”

will be used to define $OccGoal(i, \varphi)$, etc.
Belief operators

Bel_i \varphi = "agent i believes that \varphi"

Poss_i \varphi \overset{\text{def}}{=} \neg \text{Bel}_i \neg \varphi

Poss_i \varphi = "agent i thinks that \varphi is possible"

- epistemic and doxastic logic [Hintikka 62]
- express truster’s beliefs about trustee’s properties

\text{Belief}(i, \varphi) \overset{\text{def}}{=} \text{Bel}_i \varphi
Goals as preferences about the future

\( \text{Pref}_i \varphi = \text{“agent } i \text{ wants } \varphi \text{ to be true”} \)

- binary preferences [Cohen & Levesque 90]
- positive introspection: \( \vdash \text{Pref}_i \varphi \rightarrow \text{Bel}_i \text{Pref}_i \varphi \)
- negative introspection: \( \vdash \neg \text{Pref}_i \varphi \rightarrow \text{Bel}_i \neg \text{Pref}_i \varphi \)
- realism: \( \vdash \text{Bel}_i \varphi \rightarrow \text{Pref}_i \varphi \)

\( \text{OccGoal}(i, \varphi) \overset{\text{def}}{=} \text{Pref}_i \text{ Eventually } \varphi \)
Capability in dynamic logic

\( \text{After}_{i:\alpha} \varphi = \text{“} \varphi \text{ will be true after every possible execution of action } \alpha \text{ by agent } i \text{”} \)

- propositional dynamic logic (PDL)
- formulas \( = \varphi, \psi, \ldots \) = state descriptions
- actions \( = \alpha, \beta, \ldots \) = state transition descriptions
  - action \( \neq \) formula
  - \( i : \alpha = \text{action } \alpha \text{ with author } i \)
- \( \text{After}_{i:\alpha} \bot = \text{“} i \text{ cannot do } \alpha \text{”} \)

\( \text{OccCap}(j, \alpha) \overset{\text{def}}{=} \neg \text{After}_{j:\alpha} \bot \)

- “\( j \) is capable to perform \( \alpha \)” = “\( j \) can do \( \alpha \)”
Power in dynamic logic

\[
\text{OccPower}(j, \alpha, \varphi) \overset{\text{def}}{=} \text{After}_{j: \alpha} \varphi
\]

- relates \(j\)’s action \(\alpha\) with \(i\)’s goal \(\varphi\)
- missing: epistemic aspect of power (⇒ ‘knowing how to play’)
Intention-to-do as preferred action

\[ \text{Does}_{i:\alpha} \varphi = " \text{agent } i \text{ is going to do } \alpha \text{ and } \varphi \text{ will be true afterwards} \]

- \[ \text{Does}_{i:\alpha} \top = i \text{ does } \alpha \]

\[ \text{OcclIntends}(j, \alpha) \overset{\text{def}}{=} \text{Pref}_j \text{Does}_{j:\alpha} \top \]

Some axioms:
1. \[ \text{Does}_{i:\alpha} \varphi \rightarrow \neg \text{After}_{i:\alpha} \neg \varphi \ (\vdash \text{Does}_{i:\alpha} \top \rightarrow \neg \text{After}_{i:\alpha} \bot) \]
2. \[ (\text{Pref}_i \text{Does}_{i:\alpha} \top \land \neg \text{After}_{i:\alpha} \bot) \rightarrow \text{Does}_{i:\alpha} \top \]
3. \[ \text{Does}_{i:\alpha} \top \rightarrow \text{Pref}_i \text{Does}_{i:\alpha} \top \]

Theorem
\[ \vdash (\text{Pref}_i \text{Does}_{i:\alpha} \top \land \neg \text{After}_{i:\alpha} \bot) \leftrightarrow \text{Does}_{i:\alpha} \top \]
Formal definition of occurrent trust

\[ \text{OccTrust}(i, j, \alpha, \varphi) \overset{\text{def}}{=} \text{Pref}_i \text{Eventually } \varphi \land \text{Bel}_i (\neg \text{After}_{j;\alpha} \bot \land \text{After}_{j;\alpha} \varphi \land \text{Pref}_j \text{Does}_{j;\alpha} \top) \]
Example

1’s *occurrent trust* in 2 to send a certain product in view of satisfying 1’s goal of possessing the product

\[
\text{OccTrust}(1, 2, \text{sendP1}, \text{hasP1}) \overset{\text{def}}{=} \\
\text{Pref}_1 \text{ Eventually hasP1} \land \\
\text{Bel}_1 (\neg \text{After}_2 : \text{sendP1} \perp \land \\
\text{After}_2 : \text{sendP1} \text{ hasP1} \land \\
\text{Pref}_2 \text{ Does}_2 : \text{sendP1} \top)
\]
Some properties of occurrent trust

\[ \vdash \text{OccTrust}(i, j, \alpha, \varphi) \leftrightarrow (\text{Pref}_i \text{ Eventually } \varphi \land \text{Bel}_i (\text{Does}_{i: \alpha} \top \land \text{After}_{j: \alpha} \varphi)) \]

\[ \vdash \text{OccTrust}(i, j, \alpha, \varphi) \leftrightarrow \text{Bel}_i \text{ OccTrust}(i, j, \alpha, \varphi) \]

\[ \vdash \text{OccTrust}(i, j, \alpha, \varphi) \rightarrow \text{Bel}_i \text{ Eventually } \varphi \]
Potential goal

\[ PotGoal(i, \varphi, \kappa) \overset{\text{def}}{=} \]

\[ \text{Poss}_i \text{Eventually} (\kappa \land \text{Pref}_i \text{Eventually} \varphi) \]
Formal definition of dispositional trust

\( \text{DispTrust}(i,j, \alpha, \varphi, \kappa) \overset{\text{def}}{=} \)

\[
\begin{align*}
\text{Poss}_i \text{ Eventually} (\kappa \land \text{Pref}_j \text{ Eventually} \varphi) \land \\
\text{Bel}_i \text{ Henceforth} ((\kappa \land \text{Pref}_j \text{ Eventually} \varphi) \rightarrow \\
(\lnot \text{After}_{j: \alpha} \bot \land \text{After}_{j: \alpha} \varphi \land \text{Pref}_j \text{Does}_{j: \alpha} \top))
\end{align*}
\]
Example

1’s *dispositional trust* in 2 (a mechanic) to repair his car so that the car will be in order, in the circumstances in which 1 will ask the mechanic to repair his car

\[
\text{DispTrust}(1, 2, \text{rep}1c, \text{OK}1c, \text{ask}12) \overset{\text{def}}{=} \\
\text{Poss}_1 \text{ Eventually} (\text{ask}12 \land \text{Pref}_1 \text{ Eventually} \text{OK}1c) \land \\
\text{Bel}_1 \text{ Henceforth} ((\text{ask}12 \land \text{Pref}_1 \text{ Eventually} \text{OK}1c) \rightarrow \\
(\neg \text{After}_{2:\text{rep}1c} \bot \land \text{After}_{2:\text{rep}1c} \text{OK}1c \land \text{Pref}_2 \text{ Does}_{2:\text{rep}1c} \top))
\]
Some properties of dispositional trust

**Theorem**

\[ \vdash \text{DispTrust}(i, j, \alpha, \varphi, \kappa) \iff \]
\[ (\text{Poss}_i \text{ Eventually } (\kappa \land \text{Pref}_i \text{ Eventually } \varphi) \land \]
\[ \text{Bel}_i \text{ Henceforth } ((\kappa \land \text{Pref}_i \text{ Eventually } \varphi) \rightarrow \]
\[ (\text{Does}_{i:\alpha} \top \land \text{After}_{j:\alpha} \varphi))) \]

\[ \vdash \text{DispTrust}(i, j, \alpha, \varphi, \kappa) \iff \text{Bel}_i \text{ DispTrust}(i, j, \alpha, \varphi, \kappa) \]
From dispositional to occurrent trust

**Theorem**

\[ \vdash (\text{DispTrust}(i, j, \alpha, \varphi, \kappa) \land \text{Pref}_i \text{ Eventually } \varphi \land \text{Bel}_i \kappa) \rightarrow \text{OccTrust}(i, j, \alpha, \varphi) \]
Discussion: other definition of trust

- Deutsch’s definition (1958): trust involves risk perception (trusting is doing a real bet on the trustee)
  - Uncertain beliefs are needed to capture it
- Genuine trust (Trust game, Baier 1986, Holton 1994): i’s belief that j has goodwill towards him (j is willing to sacrifice his wellbeing for i)
  - It is just a special case of our definition
- Jones’s definition (2002): rule belief + conformity belief
  - Trust ≠ thinking or foreseeing
Two kinds of trust: occurrent vs. dispositional
Logical formalization
From trust to reputation
Conclusion

From trust to reputation
Reputation: the building blocks

\[ \text{Rep}(I, j, \alpha, \varphi, \kappa) = "j \text{ has reputation in group } I \text{ to ensure } \varphi \text{ by doing } \alpha \text{ in the circumstances } \kappa" \]

- Reputation as the collective counterpart of trust
- TRUST = agent \( i \)'s individual evaluation (belief) about some properties of agent \( j \) that are relevant for a goal of \( i \)
- REPUTATION = group \( I \)'s collective evaluation about some properties of agent \( j \) that are relevant for a (group) goal of \( I \)
- to be defined: collective evaluation, group goals
Public facts

For every $I$ such that $|I| > 2$,

\[ \text{Public}_I \varphi = \text{“$\varphi$ is public in the group of agents $I$” (”$\varphi$ is said in $I$”)} \]

- “$\varphi$ is public in $I$” $\neq$ common belief in $I$ that $\varphi$
- “$\varphi$ is public in $I$” does not imply “every agent in $I$ believes $\varphi$”
Logic of public facts: some principles

1. $$\neg (\text{Public}_I \varphi \land \text{Public}_I \neg \varphi)$$
2. $$\text{Public}_I \varphi \rightarrow \text{Public}_J \text{Public}_I \varphi$$
   if $$J \subseteq I$$
3. $$\neg \text{Public}_I \varphi \rightarrow \text{Public}_J \neg \text{Public}_I \varphi$$
   if $$J \subseteq I$$
4. $$\text{Public}_I \varphi \rightarrow \text{Bel}_i \text{Public}_I \varphi$$
   if $$i \in I$$
5. $$\neg \text{Public}_I \varphi \rightarrow \text{Bel}_i \neg \text{Public}_I \varphi$$
   if $$i \in I$$
6. $$\text{Public}_I (\bigwedge_{i \in I} \text{Bel}_i \varphi \rightarrow \varphi)$$
Group goals

\[
\text{GroupPref}_I \varphi = \text{“the agents in } I \text{ want that } \varphi \text{”}
\]

- group goals: broad sense (individual preference aggregation)
- weaker than joint goals and joint intentions [Grosz & Kraus 96]

Some examples of \text{GroupPref}_I \varphi definition:

- \( \exists \)-group goal: \( \bigvee_{i \in I} \text{Pref}_i \varphi \)
- \( \forall \)-group goal: \( \bigwedge_{i \in I} \text{Pref}_i \varphi \)
- Group goal based on majority: \( \bigvee_{J \subseteq I, |J| > |I|} \bigwedge_{i \in J} \text{Pref}_i \varphi \)
Formal definition of reputation

\[ \text{Rep}(I, j, \alpha, \varphi, \kappa) \overset{\text{def}}{=} \]

\[ \text{PotGoal}(I, \varphi) \land \]
\[ \text{Poss}_I \text{Henceforth}((\kappa \land \text{GroupPref}_I \text{ Eventually } \varphi) \rightarrow (\neg \text{After}_{j: \alpha} \bot \land \text{After}_{j: \alpha} \varphi \land \text{Pref}_j \text{ Does}_{j: \alpha} \top)) \]

- \[ \text{Poss}_I \varphi \overset{\text{def}}{=} \neg \text{Public}_I \neg \varphi \]
  - Poss \_ I \ varphi = “according to the group I, \ varphi is possible”

- \[ \text{PotGoal}(I, \varphi) \overset{\text{def}}{=} \]
  \[ \text{Poss}_I \text{ Eventually} (\kappa \land \text{GroupPref}_I \text{ Eventually } \varphi) \]
Conclusion
Conclusion

- Contribution.
  - formal definitions of occurrent trust and dispositional trust
  - formal definition of reputation and its relation with trust

- Our related works
  - semantics for the different modalities
  - mathematical properties of the logic of belief, preference, action and time (soundness, completeness)
  - trust in information sources and in communication systems
  - from binary trust to graded trust
  - trust in agents vs. trust in roles
  - Implementation of a BDI agent reasoning about trust (see Vercouter’s presentation)
THANK YOU!