Distrust and Mistrust for Privatively and Modally qualified Information Channels

Giuseppe Primiero
(joint work with Laszlo Kosolosky)

FWO - Research Foundation Flanders
Centre for Logic and Philosophy of Science, Ghent University

Giuseppe.Primiero@Ugent.be
http://www.philosophy.ugent.be/giuseppeprimiero/

5th Workshop in the Philosophy of Information

The Debate on Trust

- Psychology & Sociology: social and work context [Rotter, 1971, Lewis and Weigert, 1985, Shapiro, 1987].

- Epistemology & Philosophy of Science: practical value, testimony, expertise, integrity, philosophical debate on knowledge, see e.g. [Dalton, 2001, Faulkner, 2012, J. and L., 2012, Kosolosky, pear, J. and Kosolosky, 2013, Hardwig, 1991, Audi, 1997]

- Formal: first-order relation among agents 'agent *A* trusts agent *B*' vs. second-order property 'relation *X* between agent *A* and agent *B* is trustworthy', see e.g. [Castelfranchi, 2004, Demolombe, 2004, Dastani et al., 2004, Herzig et al., 2010, Kramer et al., 2012, Primiero and Taddeo, 2012]

- Applications: propagation algorithms in networks [Beth et al., 1994, Kamvar et al., 2003, Guha et al., 2004]

This Work

Distrustful and mistrustful relations for channels of information transmission:

- distrust vs. mistrust
- based on transmission of dis/misinformation (SSTI)
- analyse rational behaviour on assessment of untrust
- reasons for intentionality assessment postponed to study forms of propagation

An example

Bob: I checked the timetable, the train to Brussels leaves at 5pm.

Alice realizes that Bob does not remember that today the Railway Network updates to the Spring time. He is transmitting unintentionally false information, or misinformation. Alice decides to mistrust Bob's transmission. What then?

An example

Bob: I regularly go to St. Pancras, the best way is to take a cab.

Bob wants in fact Alice to miss her train and he wants her to stay in London. He is telling her that the cab would be faster, while the Tube would be. He is transmitting intentionally false information, or disinformation. Alice decides to distrust Bob's transmission. What then?

Outline

Procedures and Goals

A message $\underrightarrow{\langle \mathcal{P}, \mathcal{G} \rangle}$:

$\mathcal{P}$: *I checked the timetable*

$\mathcal{G}$: *the train to Brussels leaves at 5pm*

$\mathcal{P}$: *I regularly go to St. Pancras*

$\mathcal{G}$: *The best way to go to the station is to take a cab.*

2 ways of generating Untrustworthiness

- Transmission with Errors:
  - it contains a *non-executable* $\mathcal{P}$
  - it contains a *non-attainable* $\mathcal{G}$;
- Incomplete Transmission:
  - it misses the procedure: $\underrightarrow{\langle \emptyset, \mathcal{G} \rangle}$
  - it misses the goal: $\underrightarrow{\langle \mathcal{P}, \emptyset \rangle}$

Mistrust as contingency on complete transmissions

- m_R as a modal modifier
- unintentionality induces a contingent truth for element(s) in the pair;
- Formally: negation over the state, non-distributing directly over the pair;

$$\frac{\neg \circ_R \langle \mathcal{P}, \mathcal{G} \rangle}{\circ_R \langle \mathcal{P}', \mathcal{G} \rangle} \qquad \frac{\neg \circ_R \langle \mathcal{P}, \mathcal{G} \rangle}{\circ_R \langle \mathcal{P}, \mathcal{G}' \rangle} \qquad \frac{\neg \circ_R \langle \mathcal{P}, \mathcal{G} \rangle}{\circ_R \langle \mathcal{P}', \mathcal{G}' \rangle}$$

Mistrust as contigency on complete transmissions

Bob transmits in $\langle \mathcal{P}, \mathcal{G} \rangle$ unintentionally false information. Then Alice should infer that:

1. *$\mathcal{P}$ is a procedure correct for $\mathcal{G}'$, or*
2. *$\mathcal{G}$ is valid through procedure $\mathcal{P}'$, or*
3. *there is a valid pair $\mathcal{P}', \mathcal{G}'$*

Back to the example: Correct Procedure, Invalid Goal.

Bob: I checked the timetable, the train to Brussels leaves at 5pm.

Alice: Bob checked before the update. The train might leave at some other time.

Mistrust as contingency on incomplete transmission

- m_R as a modal modifier
- unintentionality induces a contingent truth for the missing element in the pair;
- Formally: negation over the state, non-distributing directly over the pair;

$$\frac{\circ_S\langle\mathcal{P},\emptyset\rangle \qquad m_R(\circ_S\langle\mathcal{P},\emptyset\rangle)}{\neg\circ_R(\langle\mathcal{P},\emptyset\rangle)} \qquad \frac{\circ_S\langle\emptyset,\mathcal{G}\rangle \qquad m_R(\circ_S\langle\emptyset,\mathcal{G}\rangle)}{\neg\circ_R(\langle\emptyset,\mathcal{G}\rangle)}$$

$$\frac{\neg\circ_R(\langle\mathcal{P},\emptyset\rangle)}{\circ_R\langle\mathcal{P},\exists\mathcal{G}\rangle} \qquad \frac{\neg\circ_R(\langle\emptyset,\mathcal{G}\rangle)}{\circ_R\langle\exists\mathcal{P},\mathcal{G}\rangle}$$

Mistrust as contigency on incomplete transmissions

Bob transmits in $\langle \emptyset, \mathcal{G} \rangle$ or $\langle \mathcal{P}, \emptyset \rangle$ unintentionally false information. Then Alice should infer that:

1. *there might be a correct procedure $\mathcal{P}$ for the transmitted goal; or*
2. *there might be a corresponding valid goal $\mathcal{G}$ for the transmitted procedure;*

Back to the example: No Procedure, Invalid Goal.

Bob: [No Procedure]. The train to Brussels leaves at 5pm.

Alice: I do not know whether Bob has checked. I should check. The train might leave at another time.

Distrust as privative operator

- d_R as a privative modifier
- intentionality induces the complement of the set generated by the pair
- Formal rules: Negation Introduction and Distribution over the Content

$$\frac{\circ_R \neg\langle \mathcal{P}, \mathcal{G}\rangle}{\circ_R\langle \neg\mathcal{P}, \mathcal{G}\rangle} \qquad \frac{\circ_R \neg\langle \mathcal{P}, \mathcal{G}\rangle}{\circ_R\langle \mathcal{P}, \neg\mathcal{G}\rangle} \qquad \frac{\circ_R \neg\langle \mathcal{P}, \mathcal{G}\rangle}{\circ_R\langle \neg\mathcal{P}, \neg\mathcal{G}\rangle}$$

Distrust as content privation on complete transmissions

Bob transmits in $\langle \mathcal{P}, \mathcal{G} \rangle$ intentionally false information. Then Alice should infer that:

1. *$\mathcal{P}$ is a procedure correct for $\neg\mathcal{G}$, or*
2. *$\mathcal{G}$ is valid through procedure $\neg\mathcal{P}$, or*
3. *there is a valid pair $\langle \neg\mathcal{P}, \neg\mathcal{G} \rangle$*

Back to the example: Correct Procedure, Invalid Goal.

Bob: I regularly go to St. Pancras. The best way is to take a cab.

Alice: Bob regularly goes to St. Pancras, he knows the best way is not the cab.

Distrust on incomplete transmission

- d_R as a privative modifier
- intentionality induces the complement of the pair (for any fulfilling)
- Formal rules: Negation Introduction and Distribution over the Content

$$\dfrac{\dfrac{\circ_S\langle\mathcal{P},\emptyset\rangle \qquad d_R(\circ_S\langle\mathcal{P},\emptyset\rangle)}{\circ_R\neg(\langle\mathcal{P},\emptyset\rangle)} \qquad \circ_R(\langle\neg\mathcal{P},\forall\mathcal{G}(\neg\mathcal{G})\rangle)}{\circ_R\langle\neg\mathcal{P},\neg\mathcal{G}\rangle}$$

$$\dfrac{\dfrac{\circ_S\langle\emptyset,\mathcal{G}\rangle \qquad d_R(\circ_S\langle\emptyset,\mathcal{G}\rangle}{\circ_R\neg(\langle\emptyset,\mathcal{G}\rangle)} \qquad \circ_R(\langle\forall\mathcal{P}(\neg\mathcal{P}),\neg\mathcal{G}\rangle)}{\circ_R\langle\neg\mathcal{P},\neg\mathcal{G}\rangle}$$

Distrust as content privation on incomplete transmissions

Bob transmits in $\langle \mathcal{P}, \emptyset \rangle$ or $\langle \emptyset, \mathcal{G} \rangle$, intentionally false information. Then Alice should infer that:

1. *procedure $\mathcal{P}$ should not be considered correct; or*
2. *goal $\mathcal{G}$ should not be considered valid;*

Back to the example: No Procedure, Invalid Goal.

Bob: [No Procedure]. The best way to the station is to take a cab.

Alice: He offers no reason, I should trust none. The best way to the station is not the cab.

Some interesting properties

- Reflexivity: agents can consider their own assumptions untrustworthy;
- Limited Transitivity: transitivity fails in general for untrustworthy complete and incomplete transmissions;
- Symmetricity: symmetricty fails for untrustworthy complete transmission (with no double-games and public assessments) and is restricted for incomplete transmissions;
- Expertise: a layperson with no specific competence is but able to mistrust an expert; on the other hand, only an expert (to some sufficient degree) can distrust the content of a transmission.

Related Work

1. Definition of Distrust and Mistrust Propagation by Error Production in information channels;
2. Definition of Distrust and Mistrust for Cases of Expertise in Scientific Contexts;
3. Future Work: Risk, Doubt, Consensus reaching strategies by error resolution.

Thanks!

Questions? Remarks?

References I


Audi, R. (1997).
The place of testimony in the fabric of justification and knowledge.
American Philosophical Quarterly, 34:405–422.


Beth, T., Borcherding, M., and Klein, B. (1994).
Valuation of trust in open networks.
In *Proceedings of the Third European Symposium on Research in Computer Security*, ESORICS '94, pages 3–18, London, UK, UK. Springer-Verlag.


Castelfranchi, C. (2004).
Trust Mediation in Knowledge Management and Sharing, volume 2995 of *Lectures Notes in Computer Science*, pages 304–318.
Springer Verlag.


Dalton, R. (2001).
Peers under pressure.
Nature, 413:102–104.

References II


Dastani, M., Herzig, A., Hulstijn, J., and Torre, L. V. D. (2004).
L.: Inferring trust.
In *In: LNAI n. 3487: Procs. of Fifth Workshop on Computational Logic in Multi-agent Systems (CLIMA V)*, pages 144–160. Springer Verlag.


Demolombe, R. (2004).
Reasoning about Trust: a formal logic framework, volume 2995 of *Lectures Notes in Computer Science*, pages 291–303. Springer Verlag.


Faulkner, P. (2012).
The practical rationality of trust.
Synthese, pages 1–15.


Guha, R., Kumar, R., Raghavan, P., and Tomkins, A. (2004).
Propagation of trust and distrust.
In *Proceedings of the 13th international conference on World Wide Web*, WWW '04, pages 403–412, New York, NY, USA. ACM.

References III


Hardwig, J. (1991).
The role of trust in knowledge.
The Journal of Philosophy, 88:693–708.


Herzig, A., Lorini, E., Hübner, J. F., and Vercouter, L. (2010).
A logic of trust and reputation.
Logic Journal of the IGPL, 18(1):214–244.


J., D. W. and Kosolosky, L. (2013).
The epistemic integrity of NASA practices in the space shuttle program.
Accountability in Research, 20(2):72–92.


J., D. W. and L., K. (2012).
The epistemic integrity of scientific research.
Science and Engineering Ethics, pages 1–18.

References IV


Kamvar, S. D., Schlosser, M. T., and Garcia-Molina, H. (2003).
The eigentrust algorithm for reputation management in p2p networks.
In *Proceedings of the 12th international conference on World Wide Web*, WWW '03, pages 640–651, New York, NY, USA. ACM.


Kosolosky, L. (to appear).
'Peer review is melting our glaciers': Exploring how and why the intergovernmental panel on climate change (ipcc) went astray.
Journal for General Philosophy of Science, Special Issue on Climate Change.


Kramer, S., Goré, R., and Okamoto, E. (2012).
Computer-aided decision-making with trust relations and trust domains (cryptographic applications).
Journal of Logic and Computation.

References V


Lewis, J. and Weigert, A. (1985).
Trust as a social reality.
Social Forces, 63:967–985.


Primiero, G. and Taddeo, M. (2012).
A modal type theory for formalizing trusted communications.
Journal of Applied Logic, 10:92–114.
Manuscript. Submitted.


Rotter, J. B. (1971).
Generalized expectancies for interpersonal trust.
American Psychologist, 26:443–452.


Shapiro, S. P. (1987).
The social control of impersonal trust.
American Journal of Sociology, 93(3):623–658.