

A Perspective on Trust, Security and Autonomous Systems

W. T. Harwood
Will.Harwood@cs.york.ac.uk

J. A. Clark
jac@cs.york.ac.uk

J. L. Jacob
Jeremy.Jacob@cs.york.ac.uk

Department of Computer Science
University of York

ABSTRACT

We argue that the advent of autonomous systems profoundly changes the nature of security arguments. Whereas security in socio-economic systems fundamentally relies on human beings to be the ultimate resolvers of trust decisions, this can no longer be the case for truly autonomous systems. In order to understand how to build such autonomous systems there is a need to create a logical theory of trust and trustworthiness. This paper attempts to provide a logical interpretation of some of the early sociological insights into the nature of trust and trustworthiness and sets out a preliminary framework for discussing reasoning about trust and trustworthiness.

1. INTRODUCTION

Trust is an underlying concept implicit in most, if not all, approaches to system security. In the normal, socioeconomic, setting for system security, systems mediate human interactions, and trust judgements are the responsibility of the people rather than the system. For example, Olga Pacheco[22] formally models trust as existing if and only if it is possible to trace the trust relation back through a system to an individual that provides the explicit warrant for trust.

However, we are in the process of moving from systems that simply mediate human interactions to systems that replace one or both parties in an interaction. Clearly, in dealing with such systems, we will have to make judgements about whether or not we trust such systems. However, the systems themselves will also have to make judgements about whether or not they trust us and, indeed, about whether or not they trust each other.

Trust is one of a number of anthropomorphic judgements that autonomous systems will have to make to interact with us and to interact with each other in a *social* way. The question of moral interaction has already been raised by Wendell Wallach and Colin Allen in the book *Moral Machines* [27] and in the articles of the special issue of IEEE Intelligent Sys-

tems on Machine Ethics [2]. It is hoped that the logical study of trust and trustworthiness will provide another approach to computational principles behind these anthropomorphic judgements.

This paper examines the kinds of reasoning associated with trust and trustworthiness. Trust is part of a network of concepts that support cooperation and collaboration between people. Other concepts in this network include risk management, insurance, regret minimization, contractual obligation, guarantees, statutory obligations, responsibilities and rights (and many more). The common theme of these concepts is that they are all associated with mechanisms that allow transactions between individuals to be spread out in time and space i.e. I can give you goods here and now and I reasonably expect payment/exchange goods in return at another time and/or at another place. Such distributed transactions are subject to uncertainty and each collaboration concept offers a reason to accept that the uncertainty has been reduced or eliminated. A major division in these concepts can be drawn using Knight's[15] distinction between risk, or *probabilistic uncertainty*, and *true uncertainty* i.e. uncertainty that cannot be quantified. Following both Barber [4] and Luhmann [19] we take trust to be a mechanism for eliminating *true* uncertainty and thus enabling decision and action. The belief that the other participants in a transaction are trustworthy is sufficient grounds to believe that the uncertainty has been eliminated and trust is the belief that the uncertainty has been eliminated.

One might ask why trust is so important. The brief answer provided by Niklas Luhmann is that without it we would be prevented from "*even getting up in the morning*" because of the uncertainties in every action and every decision[19, page 4]. Trust provides the social glue required to mediate the uncertainties of life, and in particular, the uncertainties of relying on others. Trust is a mechanism for making assumptions when we have insufficient information to otherwise proceed.

We might also add that trust has an important *economic* advantage over risk based cooperation. When one uses insurance, hedging or some related risk based mechanism there is an associated cost to one or both parties in a transaction that occurs, either because the parties must settle for an average outcome over good and bad transactions, or because they must pay a premium for the required level of protection from bad transactions. Trust permits transactions without

these associated costs.

The security of autonomous systems will depend at least as much on their abilities to make trust judgements as it will on the technical infrastructure supporting their operation. It is with this idea in mind that we attempt to answer the question: What is a logical model of trust and trustworthiness?

Below we attempt to illustrate, by example, that a defining feature of the logic of trust and trustworthiness is the need to reason about the absence of information and the need to draw plausible conclusions under conflicts of information.

We adopt a simple framework for discussing trust in which some individual, *the trustee*, makes statements about some topic to another, *the truster*. If the truster, as the name implies, actually trusts the trustee to tell the truth about the topic then, when the trustee makes a statement about that topic, the truster will believe the statement, provided, that is, there is no evidence to the contrary. That is, our notion of trust is not blind. If we trust someone we do not believe things they say if they are obviously contrary to fact. That is, what they say must be consistent with what we know to be true. Our notion of trust then is one of *trust within reason* rather than one of blind faith.

Once we have analyzed the notion of trust we move on to consider how trust may be justified i.e. when is it reasonable to regard someone as trustworthy. The problem is that of (scientific) induction. Given a pattern of observations, when is it reasonable to generalize it to future behaviour?

2. SOME NON-CLASSICAL LOGIC

We set out a minimal amount of technical apparatus to underpin our discussion below which uses non-monotonic and para-consistent logic.

Our departure from standard logic is to use both deductive reasoning and consistency based reasoning.

Let $\Gamma \vdash \phi$ stand for the standard notion that the set of formula Γ entails the formula ϕ and let $\Gamma \Vdash \phi$ stand for ϕ is consistent with Γ and let $maximal(\Gamma)$ stand for the set of maximally consistent subsets of Γ . We will limit ourselves to a base logic of classical propositional logic and define these notions in terms of entailment:

$$\Gamma \Vdash \phi \equiv \Gamma \not\vdash \neg\phi$$

$$maximal(\Gamma) = \{C \subseteq \Gamma \mid C \not\vdash false \wedge \forall \gamma \in \Gamma \setminus C. C \cup \{\gamma\} \vdash false\}$$

One natural logic of consistency conditions is that of Reiter's non-monotonic default logic[24]. Others have used non-monotonic logic as a particular mechanism of computation (e.g. []). Here, however, we regard non-monotonicity as the underlying logic of trust captured in the slogan "*to trust within reason is to trust non-monotonically*". In the next section we will use Non-monotonic default logic to model trust. Non-monotonic default logic introduces rules whose

meaning is given by combining \vdash and \Vdash . The non-monotonic rule $A : B \implies \phi$ means that ϕ follows *non-monotonically* from some set of propositions Γ if Γ entails each of A and each of B is consistent with Γ i.e. in some context Γ :

$$A : B \implies \phi \text{ means if } \forall \alpha \in A. \Gamma \vdash \alpha \wedge \forall \beta \in B. \Gamma \Vdash \beta \text{ then } \phi$$

a common special case of such a rule is a *normal default rule*, written $A : B \ni \phi$, which imposes the extra condition that the conclusion ϕ must be consistent with Γ before the rule can be applied, and that the set of propositions B must hold after the rule is applied i.e.

$$A : B \ni \phi \text{ means} \\ \text{if } \forall \alpha \in A. \Gamma \vdash \alpha \wedge \forall \beta \in B. \Gamma \Vdash \beta \wedge \Gamma \vdash \phi \text{ then } \phi \wedge \bigwedge B$$

Non-monotonicity is not enough for reasoning about trustworthiness. The problem is that we must often derive conclusions about trustworthiness given conflicting data. It therefore becomes necessary to adopt some level of para-consistent reasoning i.e. reasoning that can draw useful conclusions in the face of inconsistent assertions from potentially unreliable sources of information. The approach we take is essentially that of Rescher and Manor's approach to para-consistent reasoning[25] in which different consistent combinations of statements from a set of contradictory statements give rise to alternative belief states. Each belief state representing a distinct consistent alternative belief about possible states of the world. We take the maximal consistent sets as defined by the operator *maximal* given above and select from that set a 'best' set. If we assume information comes from different sources and that we have no apriori reason to believe one source over another then the problem is how to choose one maximal set over another. Our approach is to assume that there is some additional, possibly domain specific, principle that provides a preference ordering over maximally consistent sets such that the ordering defines a choice function, let us call it C . An inconsistent set of statements, Γ , then leads to the 'most plausible theory' $plausible(\Gamma)$ defined by:

$$plausible(\Gamma) = C(maximal(\Gamma))$$

3. MODELING TRUST AS CERTAINTY

We may now use the notion of a normal default rule to capture the idea that trust-within-reason "eliminates uncertainty". We will model from the perspective of the truster towards the trustee. The trustee makes certain statements to the truster, which we write as trustee $\blacktriangleright_{truster} \phi$. To be regarded as trustworthy statements the statements must be *on topic*, written $\phi \in \text{topic}$; made in an appropriate context (so that, for example, the truster knows they are not a joke); must not be contrary to fact (handled by the default rule) and, finally, that whatever conditions guarantee the continuity of trust have not been violated. This last point needs a little explanation. A truster trusts the trustee for some reason. Usually the reason is not something that is checked every time trust is placed in the trustee. Rather,

non-monotonically, the reason for trustworthiness is assumed until evidence to the contrary is manifest.

To take a concrete example: assume Bob is a doctor, Alice trusts Bob for medical information if she is visiting him at a clinic and if Bob has not been struck off the medical register. Generally not being struck off is a continuity condition that is assumed true unless there is explicit evidence to the contrary. That is, Alice, like most people, does not regularly check up on their doctors to see if they are still allowed to practice. Rather she assumes that she will hear if they are removed from practice via friends, newspapers and general gossip. The resultant rule will look like:

$$\text{Bob} \blacktriangleright_{\text{Alice}} \phi, \phi \in \text{Medicine}, \text{at-clinic} : \text{registered}(\text{Bob}) \Rightarrow \phi$$

Note, because we are using a normal default rule, the conclusion, ϕ , is required to be consistent with the facts as known to Alice before the rule is applied and the justification, $\text{registered}(\text{Bob})$, is a conclusion after applying the rule.

That is Alice's rule about Bob says, that if Bob says ϕ , and ϕ is about Medicine and Bob and Alice are at the clinic and Alice has no reason to think that ϕ is false or that Bob has ceased to be a registered practitioner, then Alice will believe that ϕ is true and that Bob is a registered practitioner.

Applying this logic to the classic trust management situation of certificates elucidates how trust composes. In this it yields a two step process in which we must consider both the trust in the assertion (key binding) carried by the certificate and the trust in the issuing authority. Consider Alice trusts an identity certificate C for asserting Clara's public key. Let $\text{bind}(K, \text{Clara})$ be the assertion that K is Clara's public key where $\text{bind}(K, \text{Clara})$ is an *Identity* assertion, let $\text{valid}(C)$ mean C was validly issued (meaning that the form of the certificate, its signing etc., is valid which is verified by checking the certificate) and $\text{revoked}(C)$ meaning that C has been revoked (verified by checking a revocation list). Then Alice trusts the certificate C to make identity assertions if Alice has the rule:

$$C \blacktriangleright_{\text{Alice}} \phi, \phi \in \text{Identity}, \text{valid}(C) : \neg \text{revoked}(C) \Rightarrow \phi$$

where ϕ is a propositional variable.

Instantiating to $\text{bind}(K, \text{Clara})$:

$$C \blacktriangleright_{\text{Alice}} \text{bind}(K, \text{Clara}), \text{bind}(K, \text{Clara}) \in \text{Identity}, \text{valid}(C) : \neg \text{revoked}(C) \Rightarrow \text{bind}(K, \text{Clara})$$

That is, Alice is willing to use the binding unless she knows that the binding is invalid or she knows that C has become untrustworthy because it has been revoked.

Continuing the example, Alice trusts the certificate authority which issued C , let us call him Dave, if she trusts Dave's

assertions on the topic of *Issuing* certificates (for concreteness sake the assertion $\text{cert}(c, x)$ means that c is a certificate for x), provided Dave is a registered certificate authority in the appropriate legal domain, i.e. if:

$$\text{Dave} \blacktriangleright_{\text{Alice}} \phi, \phi \in \text{Issuing} : \text{registered_CA}(\text{Dave}) \Rightarrow \phi$$

where $\text{registered_CA}(y)$ means that y is a legally registered certificate authority fulfilling the legal requirements in the appropriate jurisdiction. One should note that the assumption that Alice makes here is that Dave is a registered certificate authority i.e. Alice does not check this to be so on each issuing, rather she assumes it to be true unless contradicted by other evidence. Although it is evident that such a rule is widely adopted by people in practice, strictly speaking it is insecure. Instantiating $\text{cert}(c, x)$ for ϕ Alice has the rule:

$$\text{Dave} \blacktriangleright_{\text{Alice}} \text{cert}(c, x), \text{cert}(c, x) \in \text{Issuing} : \text{registered_CA}(\text{Dave}) \Rightarrow \text{cert}(c, x)$$

Putting the two steps together, instantiating x to Clara and c to C : Alice trusts Dave as a certificate issuing authority and if Dave issues certificate C to that says that K is the key for Clara then Alice will treat this as true provided that she it is consistent that Dave is a legally registered certificate authority, that it is consistent that the certificate C has not been revoked (i.e. it is not contrary to Alice's knowledge), that it is consistent that the certificate C is for Clara and that K is Clara key (i.e. Alice does not have information that would contradict these bindings).

This chaining generalizes, allowing trust chains to be built up that incorporate both explicit and default conditions for each step in the chain. The trust chain breaks if either, there is an explicit failure of an explicit condition, or if belief in the default condition would be inconsistent with the trusters current state of knowledge.

4. REASONING ABOUT TRUSTWORTHINESS

To trust someone we first decide if they are trustworthy. The determination of trustworthiness is, however, fraught with difficulties. At the very least it requires us to reason about how the "hidden state" underlying another's behaviour is connected to what they say and do.

We will consider a person as trustworthy about some particular topic if, in relation to that topic, they are:

- Honest, i.e. they believe what they say
- Disclosing, i.e. they disclose all (relevant) beliefs;
- Knowledgeable i.e. if something is true it will be part of their beliefs;
- Competent, i.e. what they believe to be true is actually true

Various logical models have been constructed based on these or similar notions[8, 17, 16, 11, 12, 13, 18] and provide a means for reasoning about these properties.

Our point here is, that to reason about whether or not an individual is, e.g. honest, corresponds to reasoning about the consistency of statements and actions of that individual with the the hypothesis that they are honest. That is to say that all that can really be done is to treat the issue as one of induction where the best one can do is seek out contradictions to the hypothesis.

For example, if an individual says that ϕ is true, then we would regard them as being honest about this, if they made statements that are entailed by the truth of ϕ and never made statements that contradicted ϕ . In terms of actions, we would expect them to avoid actions that would be costly, futile or redundant if ϕ is true and to take actions which are profitable, beneficial or necessary if ϕ is true (and conversely to avoid actions which are only profitable, beneficial or necessary if ϕ is false etc.).

More briefly: we believe that they are competent if the ϕ is consistent with our understanding of the state of the world; we take them to be knowledgeable if their assertions encompass our knowledge (or beliefs) about the particular topic (i.e. they know at least as much as we do); and we take them to be disclosing if there are no relevant facts already known to ourselves that are not disclosed by them.

To take a simple example, consider Bob reasoning about the honesty of Alice. Alice buys antiques and recommends certain dealers to Bob. Alice has said that Carol is the cheapest reputable supplier of antiques in the area ¹. Bob knows that Alice is thrifty and never pays more than she needs to for an antique and that Alice attempts to protect herself by using reputable sellers. He also knows that Alice rarely buys antiques elsewhere and only does so when Carol cannot supply the item. When Alice does buy antiques elsewhere she never goes to a competitor of Carol that competes with Carol on the basis of price, suggesting that Alice does not regard these sellers as reputable. Bob's own knowledge of antiques is limited, but he has compared prices and Carol's prices seem lower than many other sellers in the area. He also believes, based on information supplied by others, that many of the antiques being sold by sellers with lower prices have questionable provenance.

Bob also knows some information about the other sellers in the area. Let us call them a, b, c, d and e . He knows that a, b and c are consistently cheaper than Carol. He knows from other sources that both a and b have supplied goods of questionable provenance in the past.

What then can Bob do to reason about Alice's honesty? If Alice is honest then her statements will coincide with her beliefs and so Bob can ask whether or not Alice's behaviour

¹As always there is some question as to exactly what a such a claim means. Is it to be taken literally? Or should Bob takes this to be a typical default claim i.e. one which should be proceeded by "Unless there is evidence to the contrary ...". Here we will formalize it as the literal claim but the alternative claim could be formalized using a default rule

coincides with her statement? Secondly, if Bob were to assume that Alice was competent, then her statement should reflect the state of the world and should not contradict the state of the world known to Bob. Appendix A sketches the formalization of Bob's understanding of the situation.

Let the collection of statements made by Alice to Bob be F_1 and the collection of facts gathered by Bob independently be F_2 respectively, and Alice's assertion A . Then Bob accepts the hypothesis that Alice is being honest if $F_1 \Vdash A$ and $F_2 \Vdash A$. That is, Bob believes the hypothesis that Alice is honest if Alice is self consistent and Alice is consistent with what Bob knows about the world from other sources. Moreover this is a temporary state of affairs that may be overturned by new facts becoming available.

Bob has used information from other sources in his assessment. Below we discuss the process of corroboration by multiple sources in more detail.

5. EVIDENCE, CORROBORATION AND THE TRUSTWORTHINESS OF INFORMATION

The real world problem of evidence is that evidence from multiple sources is conflicting. In order to use evidence from multiple sources we seek corroboration between sources. One source corroborates another either by corroborating what is being said as independent witnesses or by asserting that other sources are telling the truth or lying.

The reasoning problem is to find the most plausible theory. Given that sources may contradict one another the reasoning is inevitably para-consistent.

Continuing the above example suppose that Efran says Alice is trustworthy, Freya says Alice is not trustworthy, Gabbo says that Efran is telling the truth and Hebe says that Freya is lying. What is Bob to decide about the trustworthiness of Alice? Clearly the statements are not consistent with one another. There are many possibilities of how Bob could proceed depending on his state of knowledge about his informants and his experience of Alice.

We will model the assertions made by Bob's informants directly. We will let $\mathbf{TT}(X)$ stand for X is a truth teller and $X \blacktriangleright \phi$ for " X says ϕ ". The assertions are then:

- Efran \blacktriangleright Trustworthy Alice
- Freya \blacktriangleright \neg Trustworthy Alice,
- Gabbo \blacktriangleright $\mathbf{TT}(\text{Efran})$,
- Hebe \blacktriangleright $\neg\mathbf{TT}(\text{Freya})$

The meaning of being a truth teller is given by the axiom:

$$A \blacktriangleright \phi \wedge \mathbf{TT}(A) \implies \phi$$

Since Bob has no apriori reason for not treating all sources equally his initial assumption is that each informant is a truth teller i.e. $\mathbf{TT}(\text{Efran})$, $\mathbf{TT}(\text{Freya})$, $\mathbf{TT}(\text{Gabbo})$ and $\mathbf{TT}(\text{Freya})$.

This leads to to *maximal* sets as maximally consistent theories: $\{\mathbf{TT}(\text{Efran}), \mathbf{TT}(\text{Gabbo}), \mathbf{TT}(\text{Hebe})\}$ and $\{\mathbf{TT}(\text{Freya})\}$ i.e. either Freya is lying or everyone else is lying.

Bob's preference function is to believe that conspiracies are less likely than individual attempts to deceive, and that large conspiracies are more difficult to maintain than small conspiracies. He therefore adopts the principle that minimizing the number of individuals lying is a reasonable strategy. Therefore, for Bob, *plausible* yields the solution that Freya is lying.

6. CONCLUSION

The motivation for this work is to look to the future of autonomous systems where the systems themselves take on the judgement tasks that today are performed by humans. Such systems will have profound impact on our lives. They will eventually play a significant role in medicine, finance, community care, policing and, inevitably, warfare. When we create such systems we will not want them to be overly rule bound. To function they must neither be overly naive nor overly paranoid in assuming trust. Moreover, to have confidence in these systems, we will need clear understanding of the principles that lie behind their trust reasoning and trust judgements. To us, this means having a clear logical account of that reasoning.

Here, we have only scratched the surface of the logical nature of trust reasoning, using very elementary logical tools. There is an extensive literature of both non-monotonic (see, for, example, [6, 7, 20]) and para-consistent logic (see for example [14, 3, 9, 10, 26, 23, 1, 5]) that provides significantly different interpretations of these notions. Moreover, we have only briefly touched on the philosophical, sociological and psychological analysis of trust (see, for example, the survey paper by McKnight and Chervany [21]). It is a matter of research to determined which, if any, of the existing systems of logic can be used to accurately reflect the notions trust and trustworthiness arising out of such analysis .

7. ACKNOWLEDGEMENTS

This research was sponsored by the U.S. Army Research Laboratory and the U.K. Ministry of Defence and was accomplished under Agreement Number W911NF-06-3-0001. The views and conclusions contained in this document are those of the author(s) and should not be interpreted as representing the ofPcial policies, either expressed or implied, of the U.S. Army Research Laboratory, the U.S. Government, the U.K. Ministry of Defence or the U.K. Government. The U.S. and U.K. Governments are authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

APPENDIX

A. FORMALISING BOB'S REASONING

Let $\text{cheaper}(x, y)$ be the statement that x consistently sells antiques cheaper than y and $\text{reputable } x$ be the statement that x is reputable. Then we can formalize Alice's claim by:

$$\bigwedge_i \text{cheaper}(i, \text{Carol}) \implies \neg \text{reputable } i$$

We can formalize what Bob knows about Alice's behaviour and about reputation in general as:

- Carol can supply the goods \implies Alice buys from Carol
- \neg Carol can supply the goods $\wedge \bigwedge_i \neg \text{cheaper}(i, \text{Carol}) \implies \bigvee_i$ Alice buys from i
- \bigwedge_i Alice buys from $i \implies \text{reputable } i \wedge (\bigwedge_j j \text{ can supply the goods} \wedge \text{cheaper}(j, i) \implies \neg \text{reputable } j)$
- $\bigwedge_i \neg \text{reputable } i \implies \neg$ Alice buys from i
- $\bigwedge_i i$ supplies goods of questionable provenance $\implies \neg \text{reputable } i$
- $\bigwedge_i \text{cheaper}(i, \text{Carol}) \implies i$ supplies goods of questionable provenance
- Alice buys from d

and what Bob knows about the other sellers in the area as:

- $\text{cheaper}(a, \text{Carol})$
- $\text{cheaper}(b, \text{Carol})$
- $\text{cheaper}(c, \text{Carol})$
- a supplies goods of questionable provenance
- b supplies goods of questionable provenance

References

- [1] Seiki Akama. Nelson's paraconsistent logics. *Logic and Logical Philosophy*, 7:101–115, 1999.
- [2] M. Anderson and S.L. Anderson. Special issue on machine ethics. *Intelligent Systems, IEEE*, 21(4), 2006.
- [3] F. G. Asenjo. A calculus of antinomies. *Notre Dame Journal of Formal Logic*, VII(1):103–105, 1966.
- [4] Bernard Barber. *The Logic and Limits of Trust*. Rutgers University Press, 1983. Ref'd by gambetta-1988a near key area of interest.
- [5] Diderik Batens. A universal logic approach to adaptive logics. *Logica universalis*, 1:221–242, 2007.
- [6] Gerhard Brewka. *Nonmonotonic Reasoning: Logical Foundations of Commonsense*. Cambridge University Press, 1991.
- [7] Gerhard Brewka, Jurgen Dix, and Kurt Konolige. *Nonmonotonic Reasoning*. CSLI Publications, 1997.
- [8] Bruce Christianson and William S. Harbison. Why isn't trust transitive? In *Proceedings of the International Workshop on Security Protocols*, pages 171–176, London, UK, 1997. Springer-Verlag.
- [9] Newton C. A. da Costa. On the theory of inconsistent formal systems. *Notre Dame Journal of Formal Logic*, XV(4):497–510, 1974.

- [10] Newton C. A. da Costa and E. H. Alves. A semantical analysis of the calculi c_n . *Notre Dame Journal of Formal Logic*, XVIII(4):621–630, 1977.
- [11] Mehdi Dastani, Andreas Herzig, Joris Hulstijn, and Leendert W. N. van der Torre. Inferring trust. In João Alexandre Leite and Paolo Torroni, editors, *CLIMA V*, volume 3487 of *Lecture Notes in Computer Science*, pages 144–160. Springer, 2004.
- [12] Robert Demolombe. Reasoning about trust: A formal logical framework. In Christian Damsgaard Jensen, Stefan Poslad, and Theodosios Dimitrakos, editors, *iTrust*, volume 2995 of *Lecture Notes in Computer Science*, pages 291–303. Springer, 2004.
- [13] Anders Moen Hagalisletto and Olaf Owe. Local deduction of trust. In *Security and Rewriting Techniques SecReT 2007 Pre-conference proceedings*, pages 45–58, Paris, France, 2008.
- [14] Stanislaw Jaskowski. A propositional calculus for inconsistent deductive systems. *Logic and Logical Philosophy*, 7:35–56, 1999.
- [15] Frank H. Knight. *Risk, Uncertainty, and Profit*. Houghton Mifflin, 1921.
- [16] Churn-Jung Liau. Belief, information acquisition, and trust in multi-agent systems—A modal logic formulation. *Artificial Intelligence*, 149:31–60, 2003.
- [17] Chuchang Liu and Maris Ozols. Trust in secure communication systems - the concept, representations, and reasoning techniques. *AI 2002: Advances in Artificial Intelligence*, pages 60–70, 2002.
- [18] Emiliano Lorini and Robert Demolombe. Trust and norms in the context of computer security: A logical formalization. In Ron van der Meyden and Leendert van der Torre, editors, *DEON*, volume 5076 of *Lecture Notes in Computer Science*, pages 50–64. Springer, 2008.
- [19] Niklas Luhmann. *Trust and Power*. John Wiley & Sons, 1979.
- [20] David Makinson. *Bridges from Classical to Nonmonotonic Logic*. College Publications, 2005.
- [21] D. H. McKnight and N. L. Chervany. Trust and distrust definitions: One bite at a time. In R. Falcone, M. Singh, and Y.-H. Tan (Eds.): *Trust in Cyber-societies, LNAI 2246*, pp. 27 - 54, 2001, 2001.
- [22] Olga Pacheco. Normative specification: A tool for trust and security. In Theodosios Dimitrakos, Fabio Martinelli, Peter Y. A. Ryan, and Steve A. Schneider, editors, *Formal Aspects in Security and Trust*, volume 3866 of *Lecture Notes in Computer Science*, pages 187–202. Springer, 2005.
- [23] Graham Priest. Minimally inconsistent lp. *Studia Logica: An International Journal for Symbolic Logic*, 50(2):321–331, 1991.
- [24] Raymond Reiter. A logic for default reasoning. *Artificial Intelligence*, 13:81–132, 1980.
- [25] N. Rescher and R. Manor. On inference from inconsistent premises. *Theory and Decision*, pages 179–217, 1970.
- [26] Nicholas Rescher and Robert Brandom. *Logic of Inconsistency: A Study in Nonstandard Possible-World Semantics and Ontology*. Blackwell, 1980.
- [27] Wendall Wallach and Colin Allen. *Moral Machines: Teaching Robots Right from Wrong*. Oxford University Press, 2009.