A Trust Logic for the Varieties of Trust

Mirko Tagliaferri[1][0000−1111−2222−3333] and Alessandro Aldini[1][1111−2222−3333−4444]

University of Urbino, Urbino 61029, Italy
mirko.tagliaferri@uniurb.it
alessandro.aldini@uniurb.it

Abstract. In his paper *Varieties of Trust*, Eric Uslaner presents a conceptual analysis of trust with the aim of capturing the multiple dimensions that can characterize various notions of trust. While Uslaner's analysis is theoretically very useful to better understand the phenomenon of trust, his account is rarely considered when formal conceptions of trust are built. This is often due to the fact that formal frameworks concentrate on specific aspects of phenomena rather than general features and, thus, there is little space for omni-comprehensive considerations about concepts. However, building formal languages that can describe trust generally are extremely important, since they can provide basic accounts employable as starting points for further investigations on trust. This paper addresses exactly this issue by providing a logical language expressive enough to describe all the varieties of trust derivable from Uslaner's conceptual analysis. Specifically, Uslaner's analysis is transformed into a conceptual map of trust, by strengthening his analysis with further reflections on the nature of trust. Then, a logical language for trust is introduced and it is shown how the validity classes of such language can characterize all the varieties of trust derivable from the conceptual map previously built.

Keywords: Computational Trust · Trust Logic · Conceptual Analysis of Trust.

1 Introduction

The social and economic research on trust conducted over the last few decades have created an abundance of different theoretical notions of trust [8, 23, 26]. Each of those theoretical notions can be employed to model trusting behaviours in various contexts and for different purposes. However, the existence of various approaches with their specific technical languages and their subject-oriented goals produced an ever increasing number of different and often incompatible definitions for trust. This makes the task of providing a proper and omnicomprehensive definition of trust hardly achievable, if not straightforward impossible. Moreover, despite what might be expected, moving to formal evaluations of the notion of trust made the matter even worse; various and distinct formal notions of trust have been developed in the last few decades to cope with the ever increasing necessity of implementing soft-security mechanisms in digital environments [16]. As a final concern, little attention is paid to crossover analyses of trust between socio-economical studies, on the one side, and computer science on the other. The lack of those crossover analyses is explained by two phenomena that characterize the literature on trust in computer science. First, computer scientists, given the highly complex nature of trust, prefer analyse and employ reputation systems rather than pure trust systems (often, and mistakenly, conflating the two), where reputation is a property possessed by a specific individual/object that determines how the individual/object is perceived by the whole community of which the individual/object is part of. On the other hand, trust is generally seen as an attitude of an individual towards another individual/object [17], thus a private and subjective phenomenon. Second, the few authors that deal directly with trust [15], build systems focused more on trust manipulation rather than trust computing, i.e., they build formal frameworks that can produce new trust values starting from previously computed trust values, but seldom provide tools to compute initial primitive trust values that can be fed into their models. Those phenomena lead to the fact that the various formal notions of trust employed in computer science have little resemblance to the ones that are typical of social or economical environments (either because reputation is modeled instead of trust or because the model doesn't provide any insight on how to generate trust in the first place). Thus, not only there seem to be a failure of both classical and formal analyses to provide unified accounts of trust, but there is also little affinity between the two typologies of analysis. This is highly problematic, since it is thought that formal notions of trust are useful in digital environments to the extent that they can produce benefits similar to the ones trust produces in ordinary society. It is thus necessary to recognize the importance of the socio-economical analyses of trust first and then employ those analyses to guide the evaluation of trust models employed in formal frameworks. This paper is an attempt to provide a partial solution to the problem of bridging socio-economical analyses of trust and formal ones. In order to achieve this goal the paper is structured as follows: in section two, Uslaner's analysis of trust is introduced and additional criteria employable to conceptualize trust are investigated. Those new criteria are taken from [25] and are assessed by looking at customary forms of trust that can be found in the philosophical literature on trust. Thanks to those criteria (Uslaner's and the addedd ones) a conceptual map for trust is built; in section three, a logical language for trust, dubbed Modal Logic for Trust (MLT), is introduced through the definition of its syntax and semantics. The language introduced in this paper is inspired by an already existing modal logic for trust presented in [29, 30, 28]. Differently from those previous versions, the language here presented: i) provides a slightly cleaner semantical structure; ii) eliminates some redundant functions; iii) introduces some theoretical clarifications on the functions employed to compute the trust values; iv) adds the definitions of the validity classes for the language. Finally, in section four, the validity classes for the language are discussed with reference to the conceptual map introduced in section two. Concluding remarks will follow.

2 The Conceptual Map of Trust

Navigating through the various definitions of trust given in the different disciplines can be a burdensome task. First of all, disciplines as diverse as sociology [1, 5, 8, 20], economics [6, 7, 27, 35], political science [10, 11, 19] and evolutionary biology [2, 31, 32] dedicated some of their attention to trust, obviously prioritizing their specific needs and using their typical examination techniques. This produced many theoretical definitions of trust which diverge on the technical language employed to express their definitions and the principal features that are highlighted about the phenomenon. This section is aimed at producing a conceptual map which can help the novice reader in his navigation of the diverse literatures on trust. The map (which can be seen in figure 1) is constructed around three dimensions which characterize trust and it is claimed that all definitions of trust (at least already existing ones) fall under a specific quadrant of the map. The conceptual idea of the map is taken from [25] which is a theoretical improvement of the ideas given by Uslaner in [34].

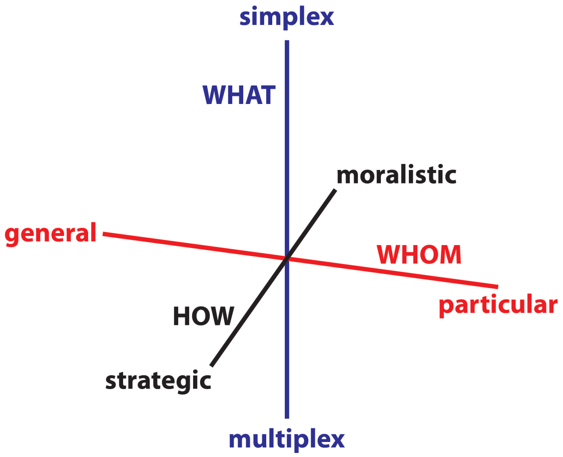

Fig. 1. Conceptual map of trust dimensions.

In his paper *Varieties of Trust* [34], Uslaner identifies two core dimensions which can characterize trust. To the dimensions he identifies, another will be added, to take into consideration also aspects of the situations in which trust arises. The first dimension characterizes the core nature of trust and distinguishes between *strategic* and *moralistic* notions of trust. The second dimension characterizes the nature of the trustees, distinguishing between trust directed towards individuals and trust directed towards institutions or larger groups of individuals. The third, and final dimension, characterizes the nature of the situation in which trust must be assessed. All the dimensions of the map will be discussed in order to provide a clear understanding of their role in possible definitions of trust. Specifically, the three dimensions regard *the nature of the actual trust relation*; *who is trusted by the trustor* and, finally, *what is the context in which to trust*.

The first dimension, indicated in [25] as the *how* dimension, characterizes the core nature of trust and distinguishes between trust definitions that are *strategic* and those that are *moralistic*. A *strategic* definition [5, 10, 11] of trust identifies the phenomenon of trusting as one depending on explicit knowledge and explicit computations about the interacting party's trustworthiness, intentions and capacities. On the other hand, a *moralistic* definition [21, 33] of trust identifies the phenomenon of trusting as a by-product of an agent's moral and ethical upbringing and consequently it depends on his psychological predispositions as defined by social norms and the values of the agent's culture. Where strategic trust can be described by the motto: *Agent A trusts agent B to do X, because of Y*; moralistic trust is simply described by saying that: *agent A trusts agent B to do X*. This dimension of trust is absolutely important to discussions concerning the notion, insofar as strategic definitions of trust presuppose that, for agent A to trust agent B, repeated encounters between the agents are necessary and, moreover, agent A must posses the computational powers to compute trustworthiness values based on information acquired during those encounters. Even though plausible, those assumptions are suited only for small communities and apply to a small number of situations and thus, strategic trust can't account for *all* the transactions and collaborations that occur in ordinary life. Moralistic versions of trust are designed to overcome this downside of strategic trust. If trust is produced as a moral commandment (similar in spirit to Kant's *categorical imperative* [18]), then even complete strangers might initiate a trust relationship. In the case of moralistic trust, it is the culture of the trustor that determines whether or not he will trust someone else and past experiences with the trustee are neither required nor important. The fact that this dimension really captures the core ideas behind the nature of trust is supported by the fact that all major accounts of trust are instances of either the strategic view of trust or the moralistic view of trust. In particular, *risk-assessment views* [10, 22, 6] and *will-based views* [14] of trust are both instances of strategic trust as defined by Uslaner, while *participant stance views* [12] and *virtue-based accounts* [13] of trust are both instances of moralistic trust.

The second dimension, indicated in [25] as the *whom* dimension, distinguishes between trust definitions that are particular and those that are general. A *particular* definition of trust identifies the phenomenon of trusting as a one-to-one relation, where trust can only be placed on specific individuals. In particular, the individuals that are considered to be trust bearers are those on whom the trustor has a fair amount of information, such as, e.g., family members, friends or colleagues. On the other hand, a *general* definition of trust identifies the phenomenon of trusting as a one-to-many relation, where trust can be placed also on anonymous individuals or strangers and such that there is no specific task or context of evaluation. In such a case, it might be said that trust is considered as an omnicomprehensive attitude towards a specific group of individuals (often those attitudes are determined by stereotypical categories). This dimension has an obvious relation with the first one: moralistic trust seem to lend well to general trust, while strategic trust is strictly tied to particular trust. However, those links are not absolute, leaving open the possibility for strategic general trust and moralistic particular trust. The former case is typical of views in which trust is seen as a stereotype: specific information about a given group of agents, i.e., the stereotyped group under consideration, is taken into consideration to determine whether the group falls indeed under the category at the base of the stereotype; then, this information is used to compute a trust value on the whole group. The latter case identifies views for which agents are morally inclined to cooperate with (and therefore trust) close relatives and known others and base their decisions to trust only on those moral values and not on specific information about the person they must interact with. As it was the case for the *how* dimension, also in the case of the *whom* dimension it is possible to find support for the relevance of this dimension by looking at major accounts of trust. In particular, the trust literature is divided between accounts that treat trust as an interpersonal phenomenon (which are the dominant paradigms of trust) and what is labelled as "institutional trust", i.e., the trust that agents place on specific institutions. In the former case, there is an obvious relation to particular definitions of trust, while the latter represent obvious instances of general definitions of trust.

The third, and final, dimension, indicated in [25] as the *what* dimension, is not directly presented in Uslaner's paper, but seems to capture a distinctive feature of trust conceptions. According to such dimension, it is possible to distinguish between trust definitions that are simplex and those that are multiplex. A *simplex* definition of trust identifies the phenomenon of trusting as being highly context-specific, where trust is granted according to a very narrow and clearly identifiable task. On the other hand, a *multiplex* definition of trust identifies trust as an extended phenomenon, which might either take into consideration variuos contexts at the same time or consider no contexts at all. In the former case, trust is evaluated on a case-by-case basis and the same two agents might trust each other in specific contexts and refuse to do so in different situations. Given the variety of different scenarios that might happen in the real world, the assumption that trust is context-specific seem to be a suitable one for a good definition of trust. However, it is important to note that there are times in which an agent trusts others blindly or in different (and multiple) situations. For instance, a child trusts his parents blindly[1]. Moreover, even admitting that context-free multiplex phenomenon of trust are impossible (independently from how much you trust someone, that someone might not be able to perform given actions, e.g., flying a plane, and thus he shall not be trusted in such contexts), it is still plausible that mild-versions of multiplex trust exist, where trust is granted with respect to a set of contexts sharing some core features, rather than a single one.

Given the three dimensions introduced, it is possible to allocate trust definitions into eight different categories (in figure 1 each quadrant represents a

[1] Note that some authors might claim that the child isn't actually trusting the parents, since he has no choice other than relying on them.

category). Each category corresponds to a given idea of what trust is. In particular, the categories are the following:

1. **Strategic particular simplex trust**: trust is seen as a specific belief about another person's reliability on a specific issue.
2. **Strategic general simplex trust**: trust is seen as a specific expectation about strangers' reliability on a specific issue.
3. **Strategic particular multiplex trust**: trust is seen as a specific belief about another person's reliability in general.
4. **Strategic general multiplex trust**: trust is seen as a specific expectation about strangers' reliability in general.
5. **Moralistic particular simplex trust**: trust is seen as a general trusting attitude towards specific individuals in specific circumstances.
6. **Moralistic general simplex trust**: trust is seen as a general trusting attitude towards strangers in specific circumstances.
7. **Moralistic particular multiplex trust**: trust is seen as a general trusting attitude towards specific individuals.
8. **Moralistic general multiplex trust**: trust is seen as a general trusting attitude towards strangers.

This conceptual map will help all further discussion on trust, by allowing the indication of a specific class which can be placed into correspondence with the formal evaluations that will be made in subsequent sections.

In the next section, the syntax and semantics of a logical language for trust will be introduced. This language will provide a proper formal framework to model all the distinct notions of trust identified in this section.

3 Modal Logic for Trust

The core idea behind the language is to describe the information possessed by an agent and then transform this knowledge into a trust value about a given proposition. In MLT, propositions substitute direct relationships between agents (or between an agent and an object). The reason is straightforward: a propositional language (rather than a predicative one) makes it easier to think about implementations of the language in computational environment, while, at the same time, retaining an expressivity which is sufficient to describe trust and its relationship to knowledge. The idea is that the relationship between the trustor and the trustee can be expressed through the use of a proposition, which is then assessed by the trustor for trust. Furthermore, while employing a predicative language might allow to express some subtleties related to trust, it also makes it infeasible to obtain positive results for decision-problems, which are, again, desirable results when building a language that is thought as a starting point for practical implementations.

Basically, the language is a modal language augmented with a trust operator, interpreted in a monotonic neighborhood semantics structure[2].

3.1 Syntax

In our language $\mathcal{L}(At)$ (for short $\mathcal{L}$) of logic formulas (which are ranged over by $\phi, \psi, \ldots$), we start with a finite set At of atomic propositions representing basic pieces of information. Given $p \in At$ our language is defined by the following grammar, given in BNF form:

$$\phi := p \mid \neg\phi \mid \phi \wedge \phi \mid K(\phi) \mid T(\phi)$$

All other Boolean connectives are defined in the standard way and we allow for a dual operator for knowledge and for trust (expressing possible knowledge and possible trust).

Formula $K(\phi)$ should be intuitively read as "formula ϕ is known"; we will call such formulas *knowledge formulas.* Formula $T(\phi)$ should be intuitively read as "formula ϕ is trusted"; we will call such formulas *trust formulas.* The degree to which a formula can be trusted goes from 0, complete distrust, to 1, complete trust; the point of transition from distrust to trust will strictly depend on the semantic structure we will now introduce.

3.2 Semantics

The semantics we will provide in this paper is in truth theoretical form and depends on a structure that is a combination of an augmented neighborhood structure for the modal part [24] and an added component to assign weights to formulas for the trust part. The added component is novel in the literature about computational trust and forms the core of the novelties this paper introduces to formalize trust.

We will interpret the above presented language in the following structure:

Definition 1 (Contextual Trust Model) *A* ***contextual trust model*** *is a tuple $M = (S, C, \pi, N, \mathcal{T}, \Theta)$, where*

- *S is a finite set of possible states of the system $s, s', \ldots$.*
- *C is a finite set of primitive evaluation scenarios $c, c', \ldots$.*
- *π is a valuation function, assigning set of states to atomic propositions.*
- *N is an augmented neighborhood function.*
- *$\mathcal{T} = \{\langle \omega_c, \mu_{c,\phi} \rangle \mid c \in C$ and $\phi \in \mathcal{L}\}$ is a trust relevance structure.*
- *$\Theta = \{\theta_c \mid c \in C\}$ is a family of trustworthiness threshold functions.*

[2] See [4, 9] for a general introduction to modal logics and monotonic neighborhood structures. Moreover, see [29] for an approach that interprets the same language in a standard relational structure.

Intuitively, a possible state $s \in S$ represents a way in which the system can be specified; hence, two states differ from one another by what propositions hold in such states. It is assumed that states are *maximally consistent* descriptions of the system. They are maximal insofar as the truth value of each proposition is specified. They are consistent insofar as a proposition and its negation can't both be true in the same state.

Set C is a finite set of primitive scenarios. Intuitively, a scenario is a situation in which trust must be assessed. The main reason to include such a set in the semantical structure comes from the consideration that most conceptions of trust see the phenomenon as a context-dependent phenomenon [8, 20] (other authors also use terms as, e.g., "scope", "purpose", "aim", and so on). Thus, in order to achieve a properly general formal system for trust, it is necessary to include in the system a component dealing with the (possible) contextuality of the phenomenon. For instance, someone might trust his mechanic when it comes to fixing cars, but might not trust him for financial advice. In the previous example, "fixing cars" and "giving financial advice" are to be considered two separate contexts of evaluation. Informally, contexts could be seen as labels assigned to states of the system, where each context is a different label that can be assigned to the same state. Thus, the whole evaluation space of formulas is equivalent to the cartesian product between the set of states S and the set of contexts C.

Function π is a valuation function that assigns to each proposition $p \in At$ a set of states, i.e., $\pi : At \to \wp(S)$; a state is included in the set if, and only if, the proposition holds in the given state.

Function N is an augmented neighborhood function that assigns to each state $s \in S$ a finite set of subsets of S, i.e., $N : S \to \wp(\wp(S))$; the set of subsets obtained by applying N is closed under superset, i.e., for each $X \subseteq S$ and each $s \in S$, if $X \in N(s)$ and $X \subseteq Y \subseteq S$, then $Y \in N(s)$. Moreover, N contains its core, i.e., $\cap N(s) \in N(s)$. Intuitively, function N assigns to each state the sets of states *corresponding to the known* propositions in such state[3]. The neighborhood function is employed to interpret the knowledge operators of the language. Note that using neighborhood functions knowledge is defined directly: thus, the informative content of a proposition is determined (in the specific case of this language by applying function π or, as it will be shown later, an extension of such a function), and the function N assigns to each state of the system a set containing all those contents corresponding to the known propositions. The closure under superset condition expresses the intuitive idea that when something is known, weakened pieces of information derived from the knowledge possessed are also known [4]. The closure under core, on the other hand, indicates that an agent is always aware of the conjunction of the information he possesses.

[3] To make the exposition simpler during the course of the paper, elements of $\wp(S)$ will be indicated with letters from the end of the alphabet capitalized and with eventual superscripts and subscripts, i.e., $X, X_2, Y, X', X_2', Y' \ldots$.

[4] For instance, if a proposition p is known at a state s, i.e., $\pi(p) \in N(s)$, then also $p \vee q$ is known at s, i.e., $\pi(p \vee q) \in N(s)$.

$\mathcal{T}$ is a trust relevance structure, where for each $\phi \in \mathcal{L}$ and each $c \in C$, there is an ordered couple $\langle \omega_c, \mu_{c,\phi} \rangle$. ω_c is a function that assigns to each formula $\phi \in \mathcal{L}$ a consistent[5] set of subsets of S, i.e., $\omega_c : \mathcal{L} \to \wp(\wp(S) - \emptyset)$ (the consistency condition expresses the informal idea that contradictions should never be considered relevant for trust formulas). This consistent set, which we call $\Omega_{c,\phi}$, contains the sets of states corresponding to the formulas relevant for trust in ϕ. $\mu_{c,\phi}$ is a trust weight function, assigning to elements in $\Omega_{c,\phi}$ rational numbers in the range [0, 1] according to their relevance for trust in the formula, i.e. $\mu_{c,\phi} : \Omega_{c,\phi} \to [0, 1] \in \mathbb{Q}$. 0 represents no trust relevance and 1 represents full trust relevance. It is assumed that the weights assigned are subadditive to 1, i.e., $\sum_{X \in \Omega_{c,\phi}} \mu_{c,\phi}(X) \leq 1$, guaranteeing that it is never possible to exceed full trust (i.e. the value 1). Intuitively, the functions $\mu_{c,\phi}$ assign to the trust relevant formulas a specific weight for trust, with respect to a given formula ϕ, which is evaluated for trust and a context of evaluation c. The notion of relevance employed here is an intuitive one: an information related to a formula is relevant for trust, if knowing such information would modify the trust assessment made towards that formula. Obviously, having no trust relevance means that whether or not the information is known, the trust assessment would be the same; on the other hand, full trust relevance means that knowing the information is the only way it is possible to modify the trust assessment.

Finally, Θ is a trustworthiness threshold structure, where, for each context $c \in C$, θ_c assigns to each formula $\phi \in \mathcal{L}$ a rational number between 0 and 1[6], i.e. $\theta_c : \mathcal{L}(At) \to [0, 1] \in \mathbb{Q}$. This rational number indicates the minimum threshold needed to trust the given formula.

Before providing the truth definition for a formula in a model, we must add some further functions; those functions will help us in defining the truth of knowledge and trust formulas.

First note that a neighborhood function N can induce a map m_N, which is a function that associates to each element $X \in \wp(S)$ another element $Y \in \wp(S)$, according to the neighborhood function N, i.e. given $N : S \to \wp(\wp(S))$, there is a map $m_N : \wp(S) \to \wp(S)$. The function m_N is defined formally as follows:

$$m_N(X) = \{s \mid X \in N(s)\} \tag{1}$$

Intuitively, m_N returns, for each set of states corresponding to a formula (i.e. the formulas informative content), a set of states such that a state is in the set if, and only if, the formula is known in the state. The function m_N will help in defining the truth of knowledge formulas.

A second and important derived element of the semantic structure is the family of functions $\Lambda = \{\tau_{c,\phi} \mid c \in C \text{ and } \phi \in \mathcal{L}\}$, which contains functions that assign ideal trust values to formulas in states of the system, and in a given context. Intuitively, a function $\tau_{c,\phi}$ ($\tau_{c,\phi} : S \to [0, 1] \in \mathbb{Q}$) indicates how much

[5] $\mathcal{U}$ is **consistent**, if $\emptyset \notin \mathcal{U}$.

[6] Real numbers could have been employed. However, it is believed that density is sufficient to capture the different grades of trust and continuity is not required. For this reason, the choice to use rational numbers is made.

trust an agent has in the formula ϕ (representing the parameter of $\mu_{c,\phi}$) in the given state and context denoting the argument of $\tau_{c,\phi}$, provided that the agent is aware, in such a state, of all the relevant basic information related to ϕ, i.e., the agent knows all the relevant propositions which are true in that state. Another way to put it is the following: if an agent knows exactly which one is the current state of the system (thus possessing all possible knowledge regarding the system), then $\tau_{c,\phi}$ will specify the amount of trust the agent has towards ϕ. Therefore, $\tau_{c,\phi}$ represents an ideal measurement of trust. Note that, even though ideal, this is a trust measure indicating how much an agent trusts the proposition ϕ in the given state and it still remains a subjective measurement.

Functions $\tau_{c,\phi}$[7] are defined as follows:

$$\tau_{c,\phi}(s) = \sum_{X \in \Omega_{c,\phi} : s \in X} \mu_{c,\phi}(X) \tag{2}$$

It is assumed that if in equation 2 there is no X such that $s \in X$ then $\tau_{c,\phi}(s) = 0$. Moreover, the subadditivity criterion on $\mu_{c,\phi}$ guarantees that $\tau_{c,\phi}$ itself never exceeds 1 (this is to be expected, since trust, even in an ideal setting might never exceed the maximum value of 1, i.e., full trust). Note that it is possible that $\tau_{c,\phi}(s) = 0$ and $\tau_{c,\neg\phi}(s) < 1$, thus the functions do not complement each other. This is perfectly reasonable, given the fact that trust, especially in ideal settings, might not be closed under complementation. In fact, it is perfectly acceptable that an agent does not trust a given proposition at all and, at the same time, he does not fully trust the negation of such proposition.

Given the family of functions $\tau_{c,\phi}$, it is possible to define a trust value for each $X \in \wp(S)$. The functions performing such task will be defined as $\tau^{ext}_{c,\phi}$ and are formally specified as follows:

$$\tau^{ext}_{c,\phi}(X) = min_{s \in X}\{\tau_{c,\phi}(s)\} \tag{3}$$

Intuitively, the function $\tau^{ext}_{c,\phi}$ looks at all states in the set X under analysis and selects the worst-case scenario, i.e., that in which the trust value is the lowest. This choice models the behaviour of a cautious agent, which will only consider the information he possesses to make an evaluation on trust and will not, therefore, make any other assumption on the trustworthiness of the formula under analysis. However, other possibilities for the definition of $\tau^{ext}_{c,\phi}$ are possible, such as taking the maximum (which would model the behaviour of an optimistic agent) or the average value between all the $\tau_{c,\phi}(s)$ (which would model the behaviour of an agent which is neither cautious nor optimistic).

Specifically, such definition would be formalized as follows.

For the maximum (optimistic agent):

$$\tau^{ext}_{c,\phi}(X) = max_{s \in X}\{\tau_{c,\phi}(s)\} \tag{4}$$

For the average (neutral agent):

[7] Again, one for each $\phi \in \mathcal{L}$.

$$\tau_{c,\phi}^{ext}(X) = \frac{\sum_{s \in X}\{\tau_{c,\phi}(s)\}}{|X|} \tag{5}$$

Where $|X|$ stands for the cardinality of X.

The various $\tau_{c,\phi}^{ext}$ equations return the ideal trust value of the set under consideration given a specific attitude of the trusting agent. The equations just given identify the trust value of a formula when the states (worlds) compatible with an agent's knowledge are selected.

It is interesting to observe that if the formula is applied to a singleton set containing only a single state s (i.e., $X = \{s\}$), then the value of the function $\tau_{c,\phi}^{ext}(X)$ is equal to the value of $\tau_{c,\phi}(s)$. This proves that $\tau_{c,\phi}^{ext}$ is indeed a proper extension of $\tau_{c,\phi}$.

To improve the readability of the truth theoretical definition for the formulas, a definition of truth set is given for each formula of the language.

Definition 2 (Extension of the Valuation Function) *Given a contextual trust model $M = (S, C, \pi, N, \mathcal{T}, \Theta)$, then the truth set of a formula, denoted π_M^{ext} (M will be omitted when the model is clear in the discussion), is defined recursively as follows:*

- $\pi_M^{ext}(p) = \pi(p)$ *for all* $p \in At$;
- $\pi_M^{ext}(\neg\phi) = S - \pi_M^{ext}(\phi)$;
- $\pi_M^{ext}(\phi \wedge \psi) = \pi_M^{ext}(\phi) \cap \pi_M^{ext}(\psi)$;
- $\pi_M^{ext}(K(\phi)) = m_N(\pi_M^{ext}(\phi))$;
- $\pi_M^{ext}(T(\phi)) = \{s \mid \tau_{c,\phi}^{ext}(\bigcap_{X \in N(s)} X) \geq \theta_c(\phi)\}$.

Two things that characterize the truth sets of trust formulas are: $\bigcap_{X \in N(s)} X$, which can also be indicated with $\bigcap N(s)$, is the core of $N(s)$ and indicates the minimal set of states which are compatible with all the knowledge of the agent; to compute the π^{ext} of $T(\phi)$, it must be checked whether in a given state the trust value of the core of N in such state is greater than or equal to the trustworthiness threshold for the formula.

Now that we introduced all the elements of our semantical structure, we can provide the truth definition of a formula ϕ at a contextual pointed model (M, s, c):

Definition 3 *Given a contextual trust model $M = (S, C, \pi, N, \mathcal{T}, \Theta)$, a state $s \in S$ and a context $c \in C$, then a formula ϕ is satisfied at a contextual pointed model (M, s, c) if:*

$(M, s, c) \models p$ iff $s \in \pi(p), \forall p \in At$;
$(M, s, c) \models \phi$ iff $s \in \pi^{ext}(\phi)$.

Given the above satisfiability conditions, it is possible to define some slightly more complicated satisfiability conditions. Those will help in defining the validity classes for MLT.

Definition 4 *Given a contextual trust model $M = (S, C, \pi, N, \mathcal{T}, \Theta)$, a state $s \in S$ and a set of contexts $A \subseteq C$, then the following holds:*
$(M, s, A) \models \phi$ *iff* $\forall c \in A, (M, s, c) \models \phi$.

All the above definitions allow to identify four different validity concepts.

Definition 5 *Given a contextual trust model $M = (S, C, \pi, N, \mathcal{T}, \Theta)$, a formula ϕ is context-valid with respect to a set of contexts $A \subseteq C$ if:*

$$\forall s \in S : (M, s, A) \models \phi \tag{6}$$

A formula ϕ is state-valid with respect to a state $s \in S$ if:

$$\forall c \in C : (M, s, c) \models \phi \tag{7}$$

A formula ϕ is model-valid if:

$$\forall s \in S \; \forall c \in C : (M, s, c) \models \phi \tag{8}$$

Finally, a formula ϕ is valid ($\models \phi$) if it is model-valid for every model M.

Those validity concepts will be analysed, one at a time, in the next section.

4 Validity Classes for MLT

When assessing trust formulas according to the validity principles introduced at the end of the last section, nice considerations about trust might be derived. Those considerations will also be made with respect to the conceptual map introduced in section two. It will be shown that the logical language introduced in this paper is expressive enough to talk about all varieties of trust indicated by the conceptual map. Before proceeding to the discussion, it is important to notice that in the model introduced in the previous section to interpret the language $\mathcal{L}$, the contexts of evaluation are employed to identify what are the issues for which trust must be assessed for trust, while the states of the system define how the system under analysis is structured (which is not directly impactful on trust) and, moreover, what is known (which greatly influences trust). Therefore, a context-valid formula, w.r.t. a context $c \in C$, might be seen as a formula that is always considered trustworthy, in that specific context $c \in C$, independently from what is known. Furthermore, a state-valid formula, w.r.t. a state $s \in S$, might be seen as a formula that is always considered trustworthy, given a specific set of known facts (i.e., the facts known in s), independently from what is the issue for which the formula must be assessed for trust. Finally, a model-valid formula is a formula that is always considered trustworthy, independently from what is known and what is the issue. With those small clarifications in hand, it is now possible to compare the semantical expressivity of the language MLT with the dimensions of trust introduced in section 2.

4.1 Context-validity

If a trust formula is context-valid with respect to a set A of contexts, then the notion of trust analysed is one for which, in the given set of context A, what might be known by the trustor is irrelevant for the attribution of trust. Thus, whatever the state of the system is, in that set of contexts trust will be granted. This kind of trust is typical of situations in which there is little choice other than trusting and no matter what is the level of knowledge, trust is always the best decision. A possible example could be a situation where the cost of not trusting and therefore not collaborating with (or not relying on) another agent/object can be so high that even if the other agent will defect the collaboration (or the object won't serve the purpose for which is was trusted), the loss is still less than or equal to the cost of not trusting. Take as an example a worn rope which an agent must choose whether to use or not to escape his house during a fire[8]. Assuming that the cost for the agent of not using the rope is death, no matter what he knows about the rope, he will trust it and use it as a possible escaping tool. This is because, even if the rope breaks (defects the trusting relationship), the worse that can happen to the agent is that he breaks his leg falling, while if he refuses to use the rope, he might face death.

Note that context-validity allows a modeller to move along the *how* dimension of trust. A formula that is context-valid w.r.t. a class of contexts can represent well moralistic versions of trust. Recall that moralistic trust is indeed based on the ethical and moral values of the trustor and, thus, specific knowledge about the trustee or, in the case of MLT, about the formula to be trusted seldom enter the picture in this typology of trust. When a trust formula is context-valid with respect to a class of contexts (or a single context), the only important factor is *that* specific class of contexts in which the formula is evaluated. Those contexts determine the exact situations in which the moral evaluations of the trustor condition him to trust the proposition under analysis. Note that trust formulas that are not context-valid represent (at least partially) strategic conceptions of trust. This is due to the fact that non-context-valid formulas distinguish between states of the system in order to assess trust and, therefore, the way the system (or world) is represented and what is known in each state matters. This can only be the case if some specific information about the proposition under evaluation is relevant for trust and, thus, knowing such information can change the trust assessment concerning the proposition. As said in section two, this kind of interaction between knowledge and trust is typical of strategic conceptions of trust, as was claimed above. Therefore, purely strategic versions of trust can be modelled using trust formulas for which there is no context making them context-valid, while it is possible to gradually move towards moralistic versions of trust by looking at trust formulas that are context valid with respect to bigger and bigger sets of contexts.

[8] In this case, there is only one element in the set A of contexts, i.e., escaping a burning house.

4.2 State-validity

If a trust formula is state-valid with respect to a state $s \in S$, then the notion of trust analysed is one for which, in the given state s, the context of evaluation is irrelevant for the attribution of trust. This means that the knowledge possessed is sufficient to have trust in the formula independently from the scenario in which trust must be assessed. This might be the case when an agent evaluates some general factors as relevant for trust independently from the contexts or where contexts have no impact on trust at all (e.g., trusting that Charlie has blond hair). For example, he might believe that, independently from the situation, a Buddhist monk would never fail to collaborate or maintain his word, therefore, knowing that someone is a Buddhist monk is sufficient to trust him, no matter the context.

Note that state-validity allows a modeller to move along the *what* dimension of trust. A formula that is state-valid w.r.t. a given state can represent well multiplex versions of trust, when the contextual model that is built contains various contexts. However, by looking at sub-contextual models for MLT that contain only a subclass of all the possible contexts or by simply looking at formulas that are not state-valid, it is possible to move towards the simplex vertex of the what-dimension. In particular, the fewer the contexts taken into consideration in the contextual model or the fewer contexts for which a formula is satisfied in a state, the closer the notion of trust taken into consideration is to the simplex notion of trust. In fact, all non state-valid formulas cover the whole length of the what dimension, while trust formulas that are state-valid in contextual model represent only multiple versions of trust.

4.3 Whom-Dimension: a Matter of Modelling

Note that there seems to be no class of validities that can capture the *whom* dimension of trust, i.e., no references have been made to the distinction between particular and general conceptions of trust. This is due to the fact that this dimension is not captured by validity principles but, instead, by the choice of trusting formulas that are evaluated. Recall that in MLT propositions substitute the relationship between agents or between an agent and an object. Thus, in case the modeller wants to capture a *particular* notion of trust, he will employ formalization that highlight only one-to-one relations, e.g., Alice will help me. On the other hand, if the modeller wants to focus on *general* notions of trust, he will employ formalizations that express a relation between the trustor and a bigger group of agents (or an institution), e.g., Microsoft will sell me a non-defective product. The language here proposed leaves complete freedom to the modeller to evaluate all the formulas he considers important. Such formulas might express propositions about single agents or multitudes of those. Therefore, the whom dimension enters the language a step before the other dimensions and is characterized by the appropriate choice of propositions to evaluate. Obviously, this dimension of trust can be integrated with the others by evaluating the specifc trust proposition in the model(s) and determining the other two dimensions according to the previously presented validity principles.

This concludes the comparison between the dimension of trust as introduced in section 2 and the way MLT identifies, through its characteristics and its validity classes, different conceptions of trust.

5 Conclusion and future works

It has been shown in the paper that different conceptions of trust are possible. Those conceptions have been categorized according to a conceptual map, which might aid in understanding the important features of all the different conceptions of trust. Then, a novel formal language to reason about trust has been introduced. Moreover, it has been shown that such a formal language is capable of representing all the different conceptions of trust by just employing the formal tools present in it. The comparison proved to be successful insofar as the language is expressive enough to talk about various conceptions of trust.

However, the language still requires a syntactic representation of the validity classes, in order to determine which rules of inference and which axioms are necessary to obtain the various validity principles. Having such a representation might help in the future to understand which are the rational mechanisms that produce trust in social and economical environment. This is a great improvement for computer science, since knowing how trust is fostered in social communities might help in reproducing the same mechanisms in digital communities, thus fostering digital versions of trust. Moreover, it might be interesting to understand if the core ideas of the formal language here introduced can be employed to improve the quality of already existing computational trust models, providing them with the tools that allow the representation of other conceptions of trust over and above the ones considered for the specific applications those models are applied to.

References

1. B. Barber, “The Logic and Limits of Trust”, Rutgers University Press, 1983.
2. P. Bateson, “The Biological Evolution of Cooperation and Trust”, in: D. Gambetta (ed.), Trust: Making and Breaking Cooperative Relations, Blackwell, pp. 31–48, 1988.
3. J. van Benthem, D. Fernández-Duque, E. Pacuit, “Evidence Logic: a New Look at Neighborhood Structures”, Advances in Modal Logic 9, pp. 97-118, 2012.
4. B.L. Chellas, “Modal Logic: an Introduction”, Cambridge University Press, 1980.
5. J. Coleman, “Foundations of Social Theory”, Harvard University Press, 1990.
6. P. Dasgupta, “Trust as a Commodity”, in: D. Gambetta (ed.), Trust: Making and Breaking Cooperative Relations, Blackwell, pp. 49–72, 1988.
7. E. Fehr, “On the Economics and Biology of Trust”, Journal of the European Economic Association 7, pp. 235–266, 2009.
8. D. Gambetta, (Ed.), “Trust: Making and Breaking Cooperative Relations”, Blackwell, 1988.

9. H.H. Hansen, “Monotonic Modal Logic”, Master’s Thesis.
10. R. Hardin, “Trust and Trustworthiness”, Russell Sage Foundation, 2002.
11. R. Hardin, “The Street-Level Epistemology of Trust”, Politics and Society 21, pp. 505–529, 1993.
12. R. Holton, “Deciding to Trust, Coming to Believe”, Australasian Journal of Philosophy 72(1), pp. 63–76, 1994.
13. K. Jones, “Trustworthiness”, Ethics 123(1), pp. 61–85, 2012.
14. K. Jones, “Second-Hand Moral Knowledge”, Journal of Philosophy 96(2), pp. 55–78, 1999.
15. A. Jøsang, “Subjective Logic”, Springer, 2016.
16. A. Jøsang, “Trust and Reputation Systems”, in: A. Aldini, R. Gorrieri (eds.), Foundations of Security Analysis and Design IV, pp. 209–245, 2007.
17. A. Jøsang, R. Ismail, C. Boyd, “A Survey of Trust and Reputation Systems for Online Service Provision”, Decision Support Systems 43(2), pp. 618–644, 2007.
18. I. Kant, “Groundwork of the Mataphysic of Morals”, 1785.
19. M. Levi, “A State of Trust”, in: V. Braithwaite, M. Levi, K.S. Cook, R. Hardin (eds.) Trust and Governance, Russell Sage Foundation, pp. 77–101, 1998.
20. N. Luhmann, “Trust and Power”, John Wiley and Sons Inc, 1979.
21. J. Mansbridge, “Altruistic Trust”, in: M.E. Warren (ed.), Democracy and Trust, Cambridge University Press, pp. 290–309, 1999.
22. O. O’Neill, “Autonomy and Trust in Bioethics”, Cambridge University Press, 2002.
23. E. Ostrom, W. James, (Eds.), “Trust and Reciprocity”, Russell Sage Foundation Series on Trust, Vol. VI, 2005.
24. E. Pacuit, “Neighborhood Semantics for Modal Logic”, Springer, 2017.
25. B.G. Robbins, “What is Trust? A Multidisciplinary Review, Critique, and Synthesis”, Sociology Compass 10(10), pp. 972–986, 2016.
26. B.G. Robbins, “On The Origins of Trust”, Ph.D. Thesis, University of Washington, 2014.
27. T. Schelling, “The Strategy of Conflict”, Harvard University Press, 1960.
28. M. Tagliaferri, “A Logical Language for Computational Trust”, Ph.D. Thesis, University of Urbino, 2019.
29. M. Tagliaferri, A. Aldini, “From Knowledge to Trust: a Logical Framework for Pre-Trust Computations”, Procs. of the 12th IFIP International Conference on Trust Management (IFIPTM’18), IFIP AICT 528, pp. 107–123, Springer, 2018.
30. M. Tagliaferri, A. Aldini, “A Trust Logic for Pre-Trust Computations”, Procs. of the 21th International Conference on Information Fusion (Fusion’18), pp. 2010–2016, IEEE, 2018.
31. R.L. Trivers, “The Evolution of Reciprocal Altruism”, The Quarterly Review of Biology 46(1), pp. 35–57, 1971.
32. R.L. Trivers, “Natural Selection and Social Theory: Selected Papers of Robert Trivers”, Oxford University Press, 2002.
33. E.M. Uslaner, “Who Do You Trust?”, in: E. Shockley, T.M.S. Neal, L.M. PytlikZillig, B.H. Bornstein (eds.), Interdisciplinary Perspectives on Trust: Towards Theoretical and Methodological Integration, Springer, pp. 71–83, 2016.
34. E.M. Uslaner, “Varieties of Trust”, European Political Science 2, pp. 43–49, 2003.
35. O. Williamson, “Calculativeness, Trust, and Economic Organization”, Journal of Law and Economics 36(2), pp. 453–486, 1993.