PhD-FSTC-2013-24
The Faculty of Sciences, Technology and Communication

DISSERTATION

Defense held on 21 October 2013 in Luxembourg

to obtain the degree of

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG

EN INFORMATIQUE

by

Tim Johannes Christina Muller
Born on 9 April 1985 in 's-Hertogenbosch (the Netherlands)

A Formal Analysis of Trust Operations

Trust Aggregation, Trust Chaining and Logical Trust Operations

Supervisor:

Prof. Dr. Sjouke Mauw (University of Luxembourg)

Defence committee:

Prof. Dr. Audun Jøsang (University of Oslo, Norway)
Prof. Dr. Yang Liu (Nanyang Technical University, Singapore)
Prof. Dr. Sjouke Mauw (University of Luxembourg)
Prof. Dr. Ulrich Sorger (University of Luxembourg)
Prof. Dr. Leon van der Torre (University of Luxembourg)

The author was employed at the University of Luxembourg.

Summary

Trust is a concept used in everyday life, both in off-line interactions and increasingly often in on-line interactions. On-line trust appears in e-commerce, as most transactions leave a party vulnerable at some point (the buyer when there is an advance payment, and the seller if there is no advance payment). On-line trust appears in the cloud, as clients need to trust providers not to destroy, lose or snoop into the data. In public key infrastructures, websurfers need to trust certifiers to be honest and not to be compromised, especially since even a (Dutch) governmental certifier "DigiNotar" turned out not to be trustworthy. Furthermore, on-line trust appears when users provide private information to third parties, who may or may not be trusted to respect the privacy of their users.

In computer security, the goal is typically to remove the necessity of trust in on-line communication. Often, cryptographic commitments, signatures or encryptions can be used to guarantee a satisfactory outcome of an interaction. In some cases, however, there are no known techniques to remove the need for trust altogether, and sometimes the need for trust is shifted from one party to another party; a party which is often called a trusted third party. This thesis focusses on dealing with trust, in the perspective of security. That means that we focus on analysing and formulating the most precise trust assessments users can make. In lieu of hard guarantees, as in classical computer security, we want to enable users to be able to make accurate assessments, and to enable users to deduce the probability that their assessment is mistaken.

Rather than analysing probabilities on a case-by-case basis, we want to generalise the procedure that leads to accurate trust assessments. We refer to such a generalised procedure as a correctness trust model, and to a trust assessments as trust opinions. In a correctness trust model, basic real-world data is translated into trust opinions. Trust opinions can be combined in different ways to reflect more complex real-world data. These ways of combining trust opinions – called trust operations – are central to this thesis.

The trust operations that we study are trust aggregation, trust chaining, trust disjunction, trust conjunction and trust negation. Each of these operations combines two trust opinions into a more sophisticated trust opinion. (Except for trust negation, which transforms a single trust opinion.) Trust aggregation allows us to merge collections of data. Trust chaining allows us to construct a trust opinion from a recommendation. Trust conjunction, trust disjunction and trust negation, together called the logical trust operations, allow us to construct trust opinions about groups of users in various combinations. These five operations together allow us to construct all opinions from just two data points: successful interactions and failed interactions. In the thesis, we study the properties of these five oper-

ations using two different methodologies. We study the operations axiomatically, and we study the operations in provability theory.

In the axiomatic method, we first study the mathematical structure of an existing model called Subjective Logic. Subjective Logic contains five operations with the same objectives as our five operations. In the exploration of the mathematical structure of Subjective Logic, we formulate axioms, statements which should be self-evident. After having formulated Subjective Logic in the language of axioms, we remove or alter those axioms which are not self-evident. The remaining axioms can be seen as a rule-book for the operations that we study.

In the probabilistic method, we encode the assumptions about the relationships between data points into probability theory. We also provide a probabilistic semantics for each of the operations. On the basis of these assumptions and semantics, we can mathematically deduce computations which perform the trust operations. These computations are undoubtedly true under the assumptions. We analyse the computations in several ways. First, we compare them to computations found in the literature which have similar assumptions. We find that some common aspects of existing computations that are consistently implemented wrongly, and some aspects that are commonly implemented correctly. Second, we analyse mathematical properties of the computations. Notably, this contains a formal analysis of the possible behaviour of recommenders. Finally, we combine the computations into a comprehensive model. This model is finally compared to the rule-book set up in the axiomatic method.

Acknowledgements

By regulation, a thesis is written by just the PhD student. Although the text in the thesis comes from one person, the thesis itself can only be conceived with the help and support of others. There are many people to whom I owe gratitude, be it for their input or for support.

Without Sjouke '*I believe it works in practice, but does it work in theory?*' Mauw this thesis would not exist. His ability to cut through the nonsense, and to the core of a problem is unparalleled. I am grateful to Sjouke for offering me a research position with a lot of freedom to explore ideas. Furthermore, I want to thank Sjouke for his insights into writing without generating nonsense and his scientific pointers.

Thanks to Patrick Schweitzer, whose mathematical background and fresh insights played an important role in the research based on probability theory. I enjoyed his well thought-through opinions, that are always odd non-the-less. Especially when these thoughts are voiced at Christmas parties, a regular game of pool or on an icy, misty and freezing ski slope.

I would like to thank the Uli Sorger and Leon van der Torre for participating in my CET committee, and for the feedback they provided in their role. The feedback on our first CET meeting was for me to not try to solve all problems in one general approach, which turned out to be especially valuable advice. My thanks also go to the remainder of my defence committee, Audun Jøsang and Yang Liu, for taking the time to provide their feedback and travelling to Luxembourg to attend the defence. It is an honour to have them in my defence committee.

Satoss was, and is, a friendly and diverse group. Diverse in personality and in academic background. This made our weekly research seminars always a lot more interesting. I enjoyed the events, the birthday cakes and – to a lesser extent – our lunches together. The people that made Satoss great while I was there are: Baptiste Alcalde, Xihui Chen, Ton van Deursen, Naipeng Dong, Wojtek Jamroga, Hugo Jonker, Barbara Kordy, Piotr Kordy, Simon Kramer, Sjouke Mauw, Matthijs Melissen, Andrzej Mizera, Jun Pang, Georgios Pitsilis, Saša Radomirović, Rolando Trujillo Rasa, Patrick Schweitzer, Chenyi Zhang and Yang Zhang.

Furthermore, I owe people at the TU/e, and in particular Jos Baeten, Bas Luttik, Paul van Tilburg and the other former members of the former Formal Methods group, gratitude for sparking my interest in formal methods, and for attending me on an interesting and challenging PhD position here in Luxembourg.

There are many friendly people I have met in Luxembourg that made my stay here enjoyable. In particular, I want to thank the friends I have made here: Djamila Aouada, Giuseppe Bonavolonta, Silvano Colombo Tosatto, Glenn Le Coz, Charles

Demange, Thorben Kätzel, David Khudaverdyan, Jessica Kubern, Adam Nawrot, Antoine Prignon, Gavin Robinson, Patrick Schweitzer. I furthermore want to thank my friends from the Netherlands that have visited me here: Harm Baarens, Bas Bergervoet, Eric Elsackers, Joost Hamelink, Harmen van Heist, John van Herk, Diederik van Houten, Marieke Meeuwissen, Harm 'Sander' Philipse, Tim Righart, Marco Verstege, Bram 'Bert' Vonk. Further thanks to Diederik for input in the early stages of the thesis. Finally, I would like the people that I have enjoyed skiing trips with: Joey Claessen, Roel Coset, Claudia Donkersloot, Denis Gerritsen, Ivo van der Linden, Bob Kubista, Wouter Verheijen, Bas van Zelst and Tamaris Zwickler.

I owe gratitude to my family. In particular to my parents, Angela Muller and Paul Muller, and my brother, Ruud Muller, for their support, their visits, and for keeping a free bedroom for me to visit every once in a blue moon. And I appreciate my family members who took the time travel from far (or from even further) to see me defend this thesis: Nicole Attwood, Dorothé Muller, Paul, Vronie and Simone De Nijs, Bas and Mieke Ten Bosch and especially Paul Veugelers, who travelled from Canada to Luxembourg to visit my defense.

Finally, a big thanks to my girlfriend Naipeng Dong, for her love, time, sense of humour, trips to China and support.

Contents

List of Figures

1

Introduction

What people mean with the concept of trust differs from person to person, and from research field to research field. When one asks an ethicist for a definition of trust, one might expect a response such as: "Trust (...) is letting other persons (natural or artificial, such as firms, nations, etc.) take care of something the trustor cares about, (...)" [Bai86]. An economist might say: "Trust is the willingness to permit the decisions of others to influence your welfare" [Sob02]. A sociological definition is: "(Trust is the) undertaking of a risky course of action on the confident expectation that all persons involved in the action will act competently and dutifully." [LW85]. A game theoretical view on the issue is formulated as: "(Trust) is the mutual confidence that one's vulnerability will not be exploited in an exchange." [BH95]. The exact interpretation of trust depends not only on taste, but for a large part on the viewpoint and the context. Our viewpoint is related to computer security.

The notion of trust over the internet is becoming increasingly relevant. People submit personal information to social media, trusting these sites to handle their data prudently. There are people buying (and selling) goods over the internet. They need to trust that their goods will be delivered correctly (or that they will receive the correct payment). Websites may have a certification from a certain certifier, who may in turn possess a certificate from another certifier, essentially forming a chain of certifiers. *Users* may or may not trust certain certificates, websites or certifiers. Amazon provides a list of products to people, which Amazon expects those people to enjoy. A user may decide to trust the *recommendation* by Amazon. Trust over the internet is, therefore, an interesting subject of study.

Trust over the internet usually concerns a certain transaction involving two parties, where one party has no control over the outcome of the transaction. That party simply needs to trust the other party, if that transaction happens. Such a transaction is called an *asymmetric interaction*, due to the asymmetry between the two parties. The party that has no control over the outcome is the *subject*, and the other is the *target*. The subject has a desired outcome from the asymmetric interaction, if the outcome matches the expectation, then the interaction was a *success*, otherwise it was a *failure*. There are many reasons why an interaction may fail, including maliciousness and incompetence of the target. We will not distinguish between motives. The subject will estimate the likelihood of success and failure, we call such an estimate the *trust opinion*.

Assume that a user wishes to purchase goods from a second-hand seller from a site like eBay, where the seller requires (partial) payment up front. Such a purchase is an example of an asymmetric interaction, since the buyer needs to pay in advance without controlling the seller. Hence, the buyer is called the subject and the seller is

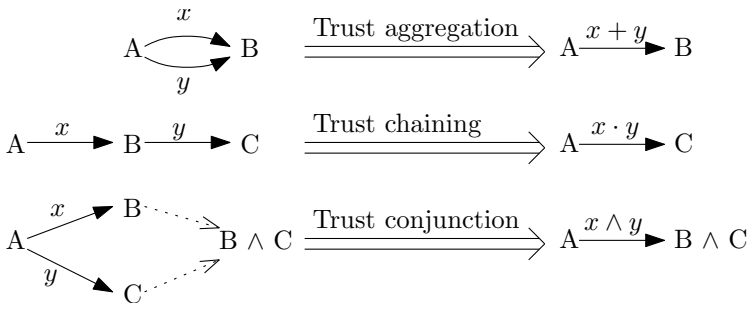

Figure 1.1: A depiction of trust aggregation, trust chaining and trust conjunction (as an instance of a logical trust operation).

called the target. The buyer expects the goods advertised by the seller delivered on time and in advertised condition. If the goods do arrive as such, then the purchase was successful, if, on the other hand, the goods were damaged, late or not delivered at all, the purchase was a failure. The buyer wants to avoid failures, and uses his trust in the seller as a guide to deciding whether to initiate the purchase.

The fact that a decision based on a trust opinion tends to be binary (either you interact, or you do not), may seem to suggest that trust is black-and-white. In reality, however, trust is far more nuanced. First off, trust opinions reflect a degree of trust. Users may want to buy a cheap bracelet from a shady merchant, but not purchase a car from a similarly shady salesman, for example. And secondly, trust opinions have a degree of certainty or confidence. For example, a new customer of a seller, when confronted with a non-delivery, will drastically lower his trust in the seller, in the belief that the seller may be scamming. On the other hand, a regular customer may be sufficiently confident in the seller not to let a single non-delivery affect his opinion too much. The different nuances, assumptions and intended application areas constitute a paradigm. We will refer to this paradigm as the *Beta paradigm* in this thesis, and explain it in Section 1.1.1.

Ways to combine trust opinions are called (trust) operations. We focus on several trust operations: *trust aggregation*, *trust chaining*, and *logical trust operations*, depicted in Figure 1.1. Trust aggregation is the act of taking several trust opinions about a single target, and combining them into a single trust opinion. Under the Beta paradigm, a mathematically correct and unique computation that performs trust aggregation has been found [MM02, JI02]. Hence, we do not analyse trust aggregation with an intent to solve it, but for the interplay with other operations, as well as similarities to other operations. The second operation, trust chaining, is the act of taking a trust opinion on an intermediate user and a recommendation (about a target) made by that user, and combining them into a single trust opinion (about that target). Trust chaining has also received much attention, but there is no consensus on an approach, even within our paradigm. In this thesis, we derive trust chaining rigourously, in a way similar to the way trust aggregation is derived. Contrary to trust aggregation, trust chaining does not have a unique computation, but a family of computations. Finally, the three logical trust operations that we study are *trust conjunction*, *trust disjunction* and *trust negation*. Trust conjunction (or disjunction) is the act of taking two trust opinions about different targets, and turning it into a trust opinion that represents that both (either) user(s) suc-

ceed(s). Logical trust operations already have a computational definition in other paradigms; in this thesis we derive their computations in the Beta paradigm.

The choice of a representation of trust opinions plus the choice of computations for operations is called a *trust model*. It is possible to roughly distinguish trust models in two groups (see Figure 1.2). Models that attempt to capture trust opinions that people hold in certain contexts – *cognitive trust models*, and models that adhere to notions of correctness – *correctness trust models*. Both types of trust models share characteristics. In both, trust in a user will increase with successes and distrust will increase with failure in interactions with that user. Similarly, certainty or confidence increases with more information, in both types of models. The rationale for these effects, however, differs. In cognitive models, the effects are simply mirroring people's tendency to adjust their opinions of that user in such a manner; the effect is empirically understood. Similarly, more interactions make people feel more confident, which the cognitive trust models merely reflects. Correctness models, however, increase trust with successes because the likelihood that the user is trustworthy has increased (by *Bayes' theorem*). And mathematically, more interactions means more data points, which in turn implies more certainty. The paradigm used in this thesis concerns correctness models.

There are two different classes of correctness models that we discuss in the paper, differing in methodology. First, we present the *axiomatic approach*. There, we define models in which correctness is defined by a small set of *axioms* (self-evident statements). There may be several models that exist under a set of axioms. In each of the models, the operations are defined as a collection of axioms. Second, we present the *probabilistic approach*. There, we define random variables and their relations as basic principles. We then derive the operations from the basic principles.

1.1 Problem Statement

The description provided above, regarding the contents of the thesis, can be concisely formulated as the following research question: *How can we correctly combine trust opinions of users on a system where users interact sparsely and with an explicit goal?* In this section, we make this question more precise. We define the types of situations and contexts that we use in our paradigm, i.e. what we mean by "a system where users interact sparsely" and "explicit goals", in the section on the scope. The ways in which we want to combine trust opinions are trust aggregation, trust chaining and logical trust operations. These are explained in the section on objectives. The phrase "correctly combine trust opinions" is explained in the section about methodology. There are two methodologies, that are not mutually exclusive, in this thesis.

It is important to note the level of abstraction in which the thesis is set. In Figure 1.2, we show the relationship between *trust systems* – explained below – and trust models. The dotted line marks the area of interest of this thesis. Note that that area partially contains game theory and applications of trust models, because we merely touch upon these subjects.

A reason that the study of trust has gained much traction, is the rise of trust

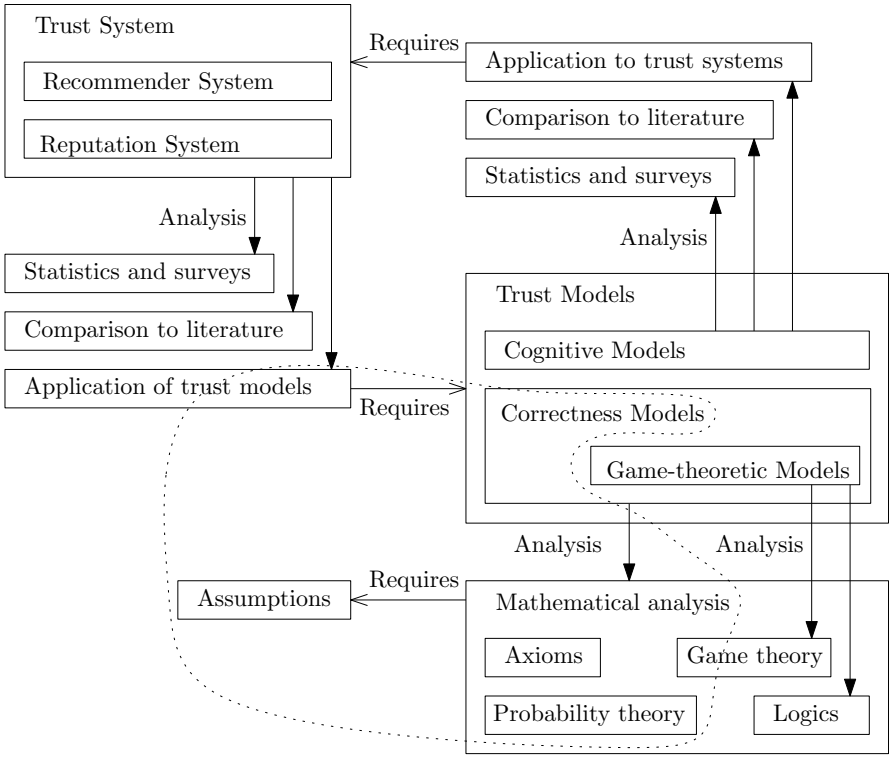

Figure 1.2: An overview of the relationships between trust systems and models.

systems on the internet. A trust system is a system that assists in or automates the construction of trust opinions. Examples of trust systems are movie rating systems, user feedback rating systems, or even Public Key Infrastructures (PKI's). Some trust systems are considered to be more effective or accurate than others. To understand why, we need a higher level of abstraction: analysis of trust systems. There are several ways of analysing trust systems, experimentally or by using trust models. An experimental analysis of a trust system could consist of surveying users of an existing and running system, or by feeding a certain (annotated) data set into a prototype of a trust system, and comparing results. The downside is that these methods may not provide insight on possible improvements, nor can it be properly determined how changing circumstances will affect its operations. Trust models can help providing insights, and they can model changing circumstances. Again, we note that one trust model may be more effective than another, which requires analysis of trust models. For cognitive models experiments, surveys and intuition are appropriate choices, due to the empirical nature of cognitive models. For correctness models, we need logics and mathematics, due to the formal nature of correctness models. We focus on the trust model analysis layer, but also look at the trust model layer. See also Figure 1.2 for a graphical depiction of the relationships mentioned in this paragraph.

In the section on the scope, we look at the setting and context that we work in. In particular, the assumptions of the model are discussed in detail. In the section on objectives, we look at the type of trust operations we analyse in trust models. In the section on methodology, we identify the mathematical tools that we apply in the analysis of trust models.

1.1.1 Scope

Our general perspective is that of computer security. The perspective strongly influences the scope of any discussion on trust. We are interested in constructing trust opinions from observation data and extracting (probabilistic) information from them. Unlike perspectives from social psychology and artificial intelligence, we are not primarily interested in motives, reciprocity, becoming trusted or the value of trust. As a consequence, our choice of perspective largely eliminates the need for a game-theoretic view in favour of a probabilistic view at trust. The view can be considered probabilistic, since our view focusses on the expected probability of success, the certainty tied to that expectation and how to update probabilities when new data is available. The three main restrictions that we introduce in this section are justified from a security perspective. However, unless all users are assumed to be rational, the issue of establishing a trust opinion based on observations is at least a subproblem of the alternative views on trust. Therefore, the work in the thesis – of which we are currently defining its scope – has value even in a perspective where the scope-restrictions cannot be justified.

There are many different situations and interactions to which we tend to ascribe trust. Given that different perspectives on trust occasionally contradict each other, is impossible to create a single trust model that captures all notions of trust. We restrict our paradigm to the appearance of trust in e-commerce and e-services, giving rise to three particular restrictions: 1) That trust is completely interaction-oriented, rather than based on personality or character (and subjects are thus completely interchangeable in the target's view[1]), and that these interactions are sufficiently similar to be grouped into a single category. 2) That interactions have a clear goal, and failure or success to achieve the goal can be objectively determined. 3) That the available information is too restricted to discover a user's full internal state, motives and incentives.

Before expanding upon these three restrictions, we observe a threefold motivation for these restrictions. The first is on a theoretical level: there is a rigourous correctness model of trust aggregation in the Beta paradigm. It is interesting to view a model – called the *Beta model* – with trust chaining and logical trust operations as merely extensions of that model (within the same paradigm). The Beta model is explained in Section 2.4.1. The second is on a pragmatic level: there currently is no other rigourous formal model of trust chaining or the logical trust operations. Liberally generalising the paradigm complicates both the formal results and their presentation. We consider it more sensible to first study the new operations within the paradigm, and then generalise the results. A rigourous model (the Beta model) and generalisations thereof already exist for models restricted to only trust aggregation, we discuss them in Section 2.4.3. An argument against trying to be too general with regards to trust is made in [MC01]. The third is on a practical level: Trust in security – trust in clouds, trust on online market places and trust in PKI's and Webs of Trust (WoT's) – fit the paradigm remarkably well. In short, we believe this paradigm allows us to rigourously formulate ideas which are interesting on a theoretical and practical level, without obfuscating the ideas

[1] This is relevant when subjects communicate with each other about interactions with a certain target.

with additional problems.

The notion (1) that our paradigm restricts to interaction-oriented trust allows us to tie a concrete meaning to trust opinions: Trust opinions reflect an estimate that the next interaction is successful. Since we assume that interactions are similar, we can form a trust opinion on a target, which applies to any next interaction we may have with that user. Another restriction of interaction-oriented trust (rather than interpersonal trust) is that a target's likelihood to succeed does not depend on the subject, since the target is not interested in the subjects themselves.

The restriction (2) that interactions are binary, i.e. results can objectively be classified as successes or failures, together with the fact that the target's behaviour is independent of the subject, entail that trust opinions can be shared between all subjects (the implication is argued in Section 2.3). Hence, an honest recommendation provides completely reliable information, since the observations by the *recommender* are identical to the observations the subject would have made in his place (since outcomes are objective) and targets behave the same towards the recommender as to the subject. That implies that subjects that receive recommendations need only be concerned with the honesty of these recommendations, not with their relevance (due to differences in taste or attitude of target). Systems under this assumption are often called *reputation systems*, whereas systems where subjective taste is the key factor are called *recommender systems*[2]. Of course, this distinction is not black-and-white; a recommendation about a hotel may be bad for subjective reasons (e.g. kitsch interior) and objective reasons (e.g. unannounced extra charges). We discuss recommender systems, provide existing examples, and their relation to reputation systems in Section 2.3.1

The inability (3) to deduce anything about the full internal states, motives and incentives of targets, means that subjects can only really know one thing about targets, namely how often they succeeded and failed in interactions. This reduces a target to a single, but unknown, quantity which represents how likely successes and failures are, called the *integrity* of that user. He may, in reality, have two states of mind, with different integrities. However, under the restriction that we cannot have knowledge regarding his state of mind, there is no point in representing both integrities. Similarly, a user may be more inclined to fail in certain types of interactions, due to differences in incentives. By not incorporating such incentives and internal states we can only reason about the typical interaction. In other words, operating under this assumption may force subjects not to use certain (statistical) information. We can classify this extra information into two types: information that transcends particular users, and information that does not. A typical example of the former case, is that the price of a product may influence the likelihood that it will not be delivered. A typical example of the latter case, is a user who has fallen ill and fails to deliver for that reason. Influences as in the former case are measurable. Our results can be easily updated to deal with these kinds of effects, when simply provided (due to measurements). Influences as in the latter case cannot be measured on a system-wide scale. However, for systems over the internet, the number of interactions between two users is so small, that

[2] Despite the fact that "recommender" is part of the term "recommender system" and not in "reputation system", reputation systems may have recommenders. In the thesis, the word recommender refers any user making trust-related claims.

it is infeasible to distinguish information about user-specific incentives or their internal states from statistical noise. Intuitively speaking, after n interactions, we merely have n pieces of binary data (for some small n), which is not nearly enough to draw conclusions regarding personality and character (other than the very broad and general). We look at systems that do look at the internals of users in Section 2.4.3.

The *Beta paradigm* is the paradigm laid out in this section. In other words, in the beta paradigm, we want to be able to make formally correct statements about the integrity of targets. To allow us to make such rigourous statements, the paradigm makes restrictions along the lines of (1), (2) and (3). The *beta distribution* is a central tenet of the Beta paradigm.

1.1.2 Objectives

We study trust in a trust-over-the-internet paradigm (outlined in Section 1.1.1, and explained in Section 2.4). The basic building block of trust is the *trust opinion*. The trust opinion reflects the estimate of the integrity of a target by the subject, and is based on a number of interactions with that target. Since we are interested in a correctness model (rather than a cognitive model), we demand that these estimates correspond with the actual integrity (rather than correspond to estimates of actual people). Recall that the actual integrity is the relevant information in the perspective of computer security, hence our choice for correctness models. Moreover, it is possible to combine trust opinions into new trust opinions. In this thesis we look at the following ways of combing trust opinions:

- *Trust aggregation* (also known as *fusion*, *consensus* or Dempster's rule of combination). Given two valid trust opinions on a target (based on distinct data), the aggregate trust opinion reflects the estimate based on those two trust opinions.
- *Trust chaining* (also known as transitive trust, *dilution*, *discounting* or trust propagation). Given a trust opinion on a recommender, and a recommendation by the recommender, the chained trust opinion reflects the estimate based on these two pieces of data.
- *Logical trust operations* (trust conjunction, trust disjunction and trust negation). Given a trust opinion on target A and a trust opinion on target B, the trust conjunction (disjunction) reflects the estimate that both (either) users succeed. Trust negation reinterprets successes as failures, and vice versa.

Due to the fact that we are interested in correctness models, we can dismiss definitions of these operators as incorrect when they are inconsistent with the basic principles of the Beta paradigm. We can also characterise those definitions that are correct. In fact, when considering only trust aggregation in our paradigm, there is a uniquely correct definition of this operator (modulo isomorphism). In other words, in that setting every correct definition of trust aggregation is the same (that definition was formulated in [JI02, MM02])

For trust chaining and the logical trust operations, such characterisations do not yet exist. Nor was it known in advance whether there exist (uniquely) correct models

with these operators. Furthermore, given the existing model of trust aggregation, we cannot know in advance whether its definition remains correct with the addition of extra operations. In other words, there is little known about correctness models on these trust operations.

The goal of this paper is to provide insight into and to formally characterise these operations and their interplay, and ultimately to define these operations. We do so in two ways, first axiomatically, then probabilistically.

1.1.3 Methodology

Assume we are interested in constructing a simple trust model based on intuition. We would reason how the result of certain trust operations correlate with their input, and choose the definition of the operation to reflect this. Say that the amount of belief or trust contained in an aggregate trust opinion positively correlates with the belief in the constituent trust opinions, we could simply define the resulting belief as a sum of the belief in its components. However, this raises the question of why not using a product or maximum. Intuition or experiments may or may not provide satisfying answers. As an alternative to picking a seemingly arbitrary specific operation, we can make a restriction to those operators where belief of the output correlates positively with belief of the inputs. Such a restriction may be called an *axiom*.

The *axiomatic approach* requires us to define a set of axioms. These axioms are statements that are undoubtedly true. A set of axioms does not necessarily define a unique model. There may not be any model for axiomatisations. This would imply that there cannot be a proper definition of the operations, and that at least one of the things believed to be undoubtedly true was not. This does not happen for any axiomatisations we consider in this thesis. There may be several models that satisfy the axioms, but that are essentially the same, just different ways of describing the same thing. We call such models *isomorphic*, and consider them to be identical. An axiomatisations of which all models are isomorphic is called a complete axiomatisations. Complete axiomatisations are useful, because they show that a model for such axioms is the only correct model (provided all the axioms are indeed true). Finally, there may be several models that satisfy the axioms, but that are fundamentally different. Even if this is the case, axiomatisations is useful. First, because we gain the insight that there may be alternative models which appear equally correct. Secondly, because we can still identify properties of all models under an axiomatisations. These properties automatically hold for all correct models.

We apply the axiomatic approach in two ways. We start by taking an existing trust model (*Subjective Logic*), and build an incrementally more powerful axiomatisations of that trust model. These candidate axioms are analysed, and we study whether or not they are undoubtedly true. Axioms that are not undoubtedly true may have alternative axioms, or generalisations, that we also analyse. Furthermore, we study a core group of axioms that remain after elimination of dubious axioms.

Another alternative to simply thinking up intuitive trust models, is to attempt to

derive a (family of) models from a small number of basic principles. This is a *probabilistic approach*, meaning that these basic principles are formulated as relations between random variables. We distinguish the probabilistic approach from the axiomatic approach (despite the existence of basic principles that could function as axioms) for several reasons, one of which is the fact that objects (trust opinions, subjects, targets, recommendations, integrity, etc) have immediate semantics, allowing a better interpretation as a (family of) model(s). To note some of those direct semantics, trust opinions are taken to be probability distributions, each target has a random variable denoting their integrity and the interactions between every pair of subjects and targets are captured in random variables. Given these translations to random variables and the relations between random variables, trust aggregation, trust chaining and the logical trust operations can be formulated as equations. The equations belonging to the operations can be seen as a direct translation of the semantics of the operations. The main objective becomes to find a general solution for these equations. Other interesting things to study are the properties of the equations. Every theorem or proposition regarding the equations has a direct semantics regarding the operations. Therefore, the probabilistic approach allows us to make general statements regarding operations (and their appearance in existing models) in the form of theorems.

There are two popular philosophical interpretations of probability: the frequentist and the Bayesian interpretation [Sti86]. Under the probabilistic assumptions that we formulate, the interpretation is irrelevant. However, the assumptions are not god-given. The assumptions that define the models are justified under the Bayesian interpretation. The reason for this is that the integrity effectively models an unknown (if you could read the mind of the target, you would most likely see that the outcome has already been determined, like a face-down card waiting to be turned). The integrity parameter can be seen as a hypothesis about future behaviour, which is not 1 or 0, but any value in between. We also apply Bayes' theorem to update a prior distribution into a posterior distribution, which in turn can be used as a prior distribution for yet another update; another characteristic of Bayesian methods. In general, therefore, we do probability under the Bayesian interpretation, although we do justify our assumptions under the frequentist interpretation in Section 6.2.

The axiomatic approach and the probabilistic approach are not mutually exclusive, and many conclusions are shared. In fact, we can show that a model derived in the probabilistic approach is a model of an axiomatisation that we provided. However, we discuss their overlap only where relevant, and mainly treat the two approaches in isolation.

1.2 Contributions

We can classify our contributions according to the two methods we have employed. First the results obtained axiomatically, then the results of the probabilistic method.

In Part I – Chapters 4 and 5, we apply the axiomatic approach to trust. First, we axiomatise a fragment of Subjective Logic, then we analyse those axioms, and reduce these axioms to those that qualify as self-evident. In the axiomatisation of

parts of Subjective Logic, we have the following contributions in particular:

- We define several sound axiomatisations of subtly different fractions of Subjective Logic. The axioms are studied, and problematic expressions are reviewed.
- We prove that all non-*dogmatic opinions* (with rational numbers as elements) can be constructed using fusion (trust aggregation) and dilution (trust chaining) and two constants.
- We introduce tuple averaging – averaging over tuples of rational numbers – as a generalisation of both arithmetic means and opinion averages, and provide a complete axiomatisation of tuple averaging.
- Using the new operation of opinion averaging, we provide a complete and finite axiomatisation of non-dogmatic opinions in the fraction of Subjective Logic that concerns with fusion, dilution and AND, OR and inverse.
- We discuss alternate axiomatisations, with respect to dogmatic opinions, alternative basic experiments and alternate interpretations for trust chaining.

We furthermore select the self-evident axioms that we have identified above. We discuss why certain axioms are selected and others are removed. Important items noted are:

- Axioms of Subjective Logic allow an unreliable recommender to strongly alter parts of the subject's opinion, which is generally undesirable.
- We argue that left commutativity of dilution holds, while associativity does not.
- We provide an axiomatisation of the expected value and weight of a trust opinion.

In part II – Chapters 6, 7, 8, 9 and 10 – we apply the probabilistic approach to trust. First, we provide a clear and systematic formalisation of the assumptions in the paradigm, also called the Beta model. Second, we formulate the logical trust operations in the setting of the Beta model, and derive the equations for the logical trust operations. Third, we formulate trust chaining in the setting of the Beta model, and derive the family of equations for trust chaining. Fourth, we link the choice for a particular equation for trust chaining to information theory and game theory. Finally, we use these techniques to formulate a single trust model (and a methodology to create similar trust models). In particular, we note the following achievements:

- We provide the only correct definition of trust conjunction and trust disjunction (of independent opinions) and trust negation, up to isomorphism, in the Beta paradigm.
- We prove that existing models in the Beta paradigm that attempt to capture logical operations either fall outside of the Beta paradigm, or incorrectly implement the logical trust operations.

- We provide the correct parameterised definition of trust chaining, where the parameters are called the *lying strategy* and the *entanglement*, w.r.t. the Beta paradigm.
- We prove that the trust operations do not interfere with each other.
- We prove that *endogenous filtering* – an effective tool in trust chaining in other settings – does not work in the Beta paradigm; only *exogenous filtering* works. In endogenous filtering, the weight of a recommendation is determined by the likelihood that its contents are true, and in exogenous filtering, by the likelihood that its issuer is honest.
- We prove that existing models in the Beta paradigm that attempt to capture trust chaining either fall outside of the Beta paradigm, or incorrectly implement the logical trust operations.
- We show the relationship between recommendations and information theory and game theory.
- We provide a correctness trust model that captures all operations (trust aggregation, trust chaining and the logical trust operations), and prove that it satisfies the axioms derived in Part I.

1.3 Organisation

The organisation of this paper is as follows: There are three parts: Part I details the axiomatic approach, Part II details the probabilistic approach and Part III concludes the thesis.

Part I contains three chapters. Chapter 3 provides the preliminary knowledge for Part I. It introduces the notion of trust networks, a fraction of the existing trust model Subjective Logic, and the formal notions surrounding axiomatisations. The fraction of Subjective Logic is completely and finitely axiomatised in an iterative fashion, in Chapter 4. Each axiomatisation, and its implications, are analysed. In Chapter 5 we identify those candidate axioms that are not self-evident, and reject them. We also provide additional axioms that we deem self-evident but not consistent with Subjective Logic.

Part II contains five chapters. The first chapter, Chapter 6, serves both as source of preliminary knowledge and to provide a formal definition of the Beta model; an example of the probabilistic approach. The following two chapters both build on the foundation of the Beta model. The first of these two chapters, Chapter 7, adds the logical trust operations to the Beta model. The other chapter, Chapter 8, add the notion of trust chaining to the Beta model to yield a family of models. The family of models differs in the amount of information carried in recommendations, which we study in Chapter 9. Finally, in Chapter 10, we merge the results discussed in Part II into a single trust model.

Finally, in Part III, we conclude the thesis and discuss future work.

2

Background

Trust is an everyday concept. Similar to most everyday concepts, there are many distinct ways of interpreting trust. Some interpretations may be mutually exclusive. For example, is the use of Trusted Third Parties (TTP's) application of trust. Some say it is, since the subject has no control over the behaviour of the TTP (e.g. [BFL96]). Others say it is not, since depending on the TTP is not optional (e.g. in [LIB+07] "trusted" is contrasted with "trustworthy"). Therefore, there cannot be a completely general notion of trust. In Section 1.1.1, we have precisely described our interpretation of trust, dubbed the *Beta paradigm*.

In this chapter, we have three goals: The first is to put the choices from Section 1.1.1 (Scope) into context, relative to the current state of research. The second is to discuss the implications of alternate choices (e.g. is it possible to obtain similar results by changing a particular choice). The third is to define the terminology concretely, within the restrictions of the Beta paradigm.

Sections 2.1, 2.2, 2.3 and 2.4.3 address the first two goals. These sections are not prerequisites for understanding our results, but rather assist in placing results into context. Sections 2.4.1 and 2.4.2 address the third goal. Most basic terminology used throughout the thesis is formally defined here. The definitions are not technical in nature. Since technical definitions vary between the axiomatic and probabilistic approaches, technical definitions are provided separately for the part they apply to.

2.1 Asymmetric Interactions

By trust concerning asymmetric interactions, we consider trust along the lines of:

"Trust is a particular level of the subjective probability with which a user assesses that another user or group of users will perform a particular action, both before he can monitor such action and in a context in which it affects his own action."

(Gambetta [Gam88], condensed for readability.)

The quote asserts that trust involves interactions, since there is an action that the *target* performs that affects the *subject*. And the quote asserts that those interactions have some level of asymmetry, since it happens before the subject can monitor the action of the target (which implies that the subject cannot control the action). The exact asymmetry of the interaction is not captured in the quote.

In the most asymmetric case, potential subjects and potential targets are mutually exclusive. In that case, an asymmetric interaction between Alice as subject and Bob as target can never happen with Bob as subject and Alice as target. The

only game-theoretical reason for Bob to attempt to be perceived as trustworthy, is because of social capital [Put93]. If he is trusted by many, more people will want to interact with Bob, which (presumably) is beneficial to Bob.

Alternatively, we can imagine that both Alice and Bob may fulfill the roles as subjects and targets at different times. So each interaction is asymmetric, but their roles are symmetric. Now, Bob has an additional reason to attempt to be perceived as trustworthy, namely reciprocity [BDM95]. If Bob is trusted as a target by many, people are more likely to act beneficial when Bob is a subject.

Game-theoretical trust models, such as Liu et al., assign or calculate values for social capital and reciprocity [LZL12]. One problem with applying game-theory (directly) to our problem is that we do not know a priori whether reciprocity is relevant. Another, more general, problem is that the value of social capital and reciprocity strongly depends on specific costs that are hard to quantify, like the cost to create an account on a trust system.

The game-theoretical trust models are often contrasted with *cognitive trust model* (which not always require asymmetric interactions, hence they are explained in more detail in Section 2.1.1). We argue that such a division would be a false dichotomy, as better antitheses of the cognitive trust models are *correctness trust models* (see Figure 1.2). In fact, it makes sense to see game-theoretical trust models as a special case of correctness trust models, since the former are interested in finding sensible *trust opinions* which represent actual (subjective) probabilities, much like the latter. The essential difference is that not all correctness trust models apply game-theory in finding trust opinions.

Many correctness trust models derive notions of trust opinions from observations [SFE08, ZLTV10]. We previously defined a notion of trust where observations are an integral part of the definition [Mul11]:

"An observation is any contingent fact, that is witnessed to be true. A trust assessment is a boolean expectation based on own observations and possibly observations of others. Trust is a positive trust assessment, and distrust a negative."

These observations could be about interactions similar to the interaction that an assessment is being made upon, essentially applying induction. This approach is popular in several computational models [Mar94] – known as evidence based trust – and in Eigentrust [KSGM03], as well as probabilistic models (such as TRAVOS [PTJL05]). However, the observations can also be completely different, such as certifications [BFL96], certain relationships [GS03] or credentials [EFL+99]. All these views can be characterised as having their philosophy rooted in Dempster-Shafer theory [Dem67, Sha76].

Asymmetric interactions can be of the sort "good or bad", or they could be more nuanced. To offer some examples of classifications that are more nuanced than just a binary classification: First, different types of bad behaviour may be classified according to intention (e.g. malicious versus incompetent), which adds an extra (cognitive) dimension to trust, as in [FC01]. Second, different types of bad (or good) behaviour may be classified according to result (e.g. late delivery versus delivery of broken goods), which allows subjects to differentiate between types of bad outcomes, as in [JH07]. Finally, a continuous range of outcomes with an order, and a supremum (e.g. response in n seconds versus m seconds), as in [ARH00]. And

of course combinations of those three nuances may exist, as in [SS01].

We remark that these additional nuances are not in our paradigm, but that they are compatible to it. Our setting (where the *beta distribution* is effective), can be generalised to a setting with extra nuances on the behaviour (using Dirichlet distributions), as is done formally in [JH07]. In the case that categories are causally linked (e.g. a successful interaction requires good intent and competence), our techniques on *logical trust operations* can also be applied. This follows from a generalisation of logical trust models where categories of outcomes are allowed, such as [MS13b].

2.1.1 Alternate Types of Trust

Not all notions of trust are strictly based on asymmetric interactions, or even on interactions, particularly, many notions are based on character. The general sentiment of trust not regarding interactions is, quoted liberally after McKnight and Chervany: "If one is predictable, benevolent, competent and honest, then one is worthy of trust indeed." [MC96] Such a notion of trust is far more suitable for situations where information is paramount, and it makes sense to reason about motives and internal states. That means, in a social setting, rather than an online setting. As we are interested in trust over the internet, notions as from McKnight and Chervany are less suitable for our purpose.

Another reason we opted for trust based on interactions, is because our objective is to study (and develop) correctness trust models. Whereas for trust not based on asymmetric interactions, it may not be possible to define a correct notion of trust chaining. That is the point argued by Christianson et al. [CH97], where the authors show that social notions of trust and cognitive trust models are (necessarily) intertwined when it comes to the transitivity of trust (i.e. whether trust chaining can be defined).

There are many interesting relations between such character-based trust notions, as explored by Robert Demolombe in [Dem04]. There, he uses modal logics to find relationships between trust notions such as sincerity, cooperativity, credibility and vigilance. The techniques used in that paper rely heavily on modal logic, and cannot be immediately quantified. By having a less rigourous relation between partial notions of trust, Castelfranchi and Falcone have quantified partial trust notions [CF98]. In particular, the four notions they treat are called competence, disposition, dependence and fulfilment.

Rather than looking just at interaction-based history, or just at alternatives, it is possible to look at trust at a level general enough to encompass both. Moyano et al. analyse trust in a top-down fashion, where concepts related to trust are viewed as components [MFGL12]. Rather than looking at how these components work (e.g. how trust is established, how it is used), they study the relationships between these components. What we call trust models based on interactions, they call evaluation models. These evaluation models comprise only a part, albeit a critical part, of their notion of trust models.

2.2 Aspects of Trust Opinions

A trust decision regarding asymmetric interactions is, in a sense, binary. Either the subject places enough trust in the target to interact, or he does not. That the decision is binary does not mean that our *trust opinions* need only be boolean. To reinforce this notion, a distinction between so-called decision trust and reliability trust is made in [JKD05]. They propose the term decision trust as "(..) willingness to depend on something or somebody in a given situation with a feeling of relative security, even though negative consequences are possible." Which means that the argument about trust being boolean may apply to decision trust (although arguably, one can be more or less inclined to trust, implying a gradual scale rather than boolean). An example of a trust model purely for decision trust is given in [QH07]. The other type of trust, reliability trust, they defined using the aforementioned quote from Gambetta. Our notion of trust opinions relates to their notion of reliability trust. Deriving decision trust from reliability trust is the objective of another field, namely risk management [LGTL85, KMRS12]. On a high level, decision trusts consists of reliability trust, (relative) risk a (relative) value [LSS10].

The notion of trust opinions must be sufficiently powerful to fulfill the role of reliability trust. In its role it should contain enough data, not just to reflect likelihood of success, but also uncertainty ([SYHL05]) and potentially more. The notion of trust opinions must also be powerful enough to operate as a *carrier set* for the trust operations (i.e. trust aggregation, trust chaining and the logical trust operations). By that, we mean that the result of these operations should be expressible as a trust opinion. We look at some existing representations of trust opinions, and study whether they are powerful and flexible enough to fulfill these two objectives.

2.2.1 Binary Trust Opinions

The obvious representations for binary trust opinions, are first order logic and modal logics. Logical representations are limited in the sense that they cannot provide quantified opinions. However, as a positive trade-off, logical representations tend to be capable of being both general and formal. Hence, we discuss these models here.

Many logical trust models have the ideas of Castelfranchi and Falcone [CF98, FC01] as their basis. A model where the ideas of Castelfranchi and Falcone are formalised and axiomatised can be found in [HLHV10], where reputation is furthermore defined similarly to trust. The predicates in such models usually refer to beliefs and other internal states, and their relation to actual entities is not always provided or analysed. Kramer et al. not only provide a logical trust model, like Castelfranchi and Falcone, but relate this to entities such as TTP's and cryptographic primitives [KGO10].

Finally, Katz et al. look at using default logic to study trust [KG06]. In default logic, one can express, for example, that birds can typically fly. Also, by default, animals with flippers cannot fly. Hence, we can conclude both that penguins can typically fly and that they typically cannot. To resolve this issue, one can prioritise

defaults, making it prioritised default logic. Katz et al. propose to use a trust metric to obtain the required priorities, in a model called TidalTrust ([Gol05]). Interestingly, TidalTrust uses natural numbers, which have more possible values than a boolean. Hence, even for a model designed for binary trust predicates, internally, trust opinions need a richer representation.

2.2.2 Degrees of Freedom in Trust Opinions

Closest to binary trust opinions, are trust opinions expressed in three-valued logic, or in fuzzy logic [Zad65]. Rather than having "yes" or "no" on trust-predicates, they allow "unknown" or "somewhat". The advantage of such an approach is that it maps directly to cognitive notions. Direct trust is represented in this way in [ARH00]. Fuzzy logic is not very suitable as a representation for trust opinions with respect to the trust operations. Since all opinions are discrete, nuances may be lost, which due to particular nesting of operations have large effects on the resulting opinion.

A more informative scale is achieved by using a single continuous parameter. Such a parameter may range in $[0, 1]$, e.g. UniTEC [KBR05], $[-1, 1]$ e.g. [Mar94], or even $\mathbb{R}$ e.g. EigenTrust [KSGM03]. In the first two, 0 and 1 correspond to distrust and 1 to trust. In all three, values halfway the range, such as 0 or $1/2$ represent neutrality. With just one parameter, it is difficult to distinguish whether a target with trust 0 is unknown or whether it is known that the target can neither be trusted nor distrusted. For example, 1 good experience and no bad ones, may lead to a trust opinion of $1/2$, and 93 good experiences and 46 bad ones may also lead to $1/2$. If we want to aggregate $(0, 1)$, i.e. no good and a single bad experience, with $(93, 46)$, it means we trust aggregate $(-1/2)$ with $1/2$ which naturally becomes 0. However, $(93, 47)$ – the pair of total successes and failures – should not be represented by 0. Another issue with one dimensional representations of trust opinions is the Ellsberg paradox, termed in [Ell61], which shows that uncertainty by itself may be an interesting measure to report.

A standard representation of trust opinions uses two dimensions, one to represent the ratio of successes versus failures, and one to represent the number of interactions (which functions as a measure of certainty). Such a representation maps trivially to beta distributions (see Section 2.4.1). A representation with two degrees of freedom corresponding to beta distributions is used often (e.g. in [Jøs97, TPJL06, Rie07, SFE08]). The aforementioned issue for one-dimensional metrics with trust aggregation can be solved using the second dimension; the measure of uncertainty. Whether these two degrees of freedom are sufficient for the other operators (trust chaining, trust conjunction and trust disjunction) was an open question. We prove that two degrees of freedom is not enough for all three operators (Theorems 7.10 and 8.11).

2.3 Recommendations

The most informative type of information are first-hand experiences. That is, observing the outcomes of interactions in which the subject was involved is more

informative than observing claims of outcomes (called *recommendations*) of interactions where others were involved. The two main reasons for this effect, are the fact that people may make false claims, and the fact that others may perceive the same interactions differently. The former applies to all situations, the latter tends to hold only in situations where outcomes are subjective. For that reason, we subdivide this section into a part about subjective recommendations and a part about objective recommendations. Subjective recommendations are relevant in *recommender systems*, whereas objective recommendations are relevant in *reputation systems*.

In the analysis of recommendations, we need to be careful with our terminology. We distinguish trust in a recommendation (an opinion on the likelihood that the recommendation is honest), trust expressed by a recommendation (the opinion that the *recommender* claims to have) and trust from a recommendation (the trust opinion that is the result of the trust chain of the previous two). An increase in trust in a recommendation does not necessarily lead to an increase in trust from a recommendation (rather a decrease of uncertainty [JMP06, AM09]). The uncertainty in the trust expressed by a recommendation is at most as large as the uncertainty from the recommendation, since there is uncertainty added by the fact that there is a non-zero probability that the recommendation was dishonest [AM09].

2.3.1 Subjective Recommendations

When recommendations regard taste or other subjective qualifications, the important concern is whether the recommendation is applicable. To use movie recommender systems (e.g. CinemaScreen [SA06]) as an example, the recommendation from a fan of romantic comedy may be irrelevant to a horror fan. Other types of recommender systems are Cobot [SR11], YouTube [DLL+10] or GroupLens [RIS+94]. The concept of matching similar tastes to determine relevance of recommendations is known as collaborative filtering ([BHK98, SKKR01]).

There are several techniques that can be applied to do collaborative filtering. The Pearson product-moment correlation coefficient [Pea96] can be used to measure the distance between two *users* (as in [SKKR01]). This distance can be used to determine the weight of a recommendation of one user to the other. Interestingly, weights can have negative value. Alternatively, the k-nearest neighbours algorithm [CH67], can be applied (as in [SKKR01]). There, the k people with most similar tastes are selected to provide recommendations, others are typically ignored.

Another interesting method, is the application of Kalman filtering [Kal60] (as in [WLS12]). Kalman filtering is designed for robustness in noisy channels. In this case, noisy channels would be formed by malicious users, that fake some of their recommendations. Rather than having a single method for filtering for relevant recommendations and against malicious, one can add an extra filtering method. A possible way to do this, is to filter outlying or strange recommendations using intrusion detection [FZAB11]. Another possible way to do this, is to adapt methodology used for objective recommendations.

2.3.2 Objective Recommendations

The goal of reputation systems is similar to that of recommender systems, namely to assist in (or automate) trust-decision making. The difference is the domain of the recommendations; in reputation systems, recommendations are regarding fact rather (or more so) than taste. In particular, that means that if a recommendation is honest, then it is immediately applicable.

Rather than filtering for relevance, we merely need to filter for honesty. In the literature (e.g. [TPJL06, JIB07]), two general types of filtering are discussed: *endogenous filtering* and *exogenous filtering*. Endogenous filtering compares a recommendation to other recommendations or to the subject's own experience, and filters out those who are too far away and/or weighs closer recommendations more heavily. Exogenous filtering looks at the recommender, rather than the recommendation, and assigns a weight according to the reliability of the recommender.

Endogenous filtering is applied in [BLB04]. However, we prove that, in the context of the Beta paradigm, endogenous filtering is superfluous at best, and exogenous filtering suffices. We formally prove this in Theorem 8.9 in Section 8.3 and informally and intuitively explain this phenomenon in Section 10.1.4.

Exogenous filtering is applied in [Jøs97, TPJL06]. There, exogenous filtering is defined as a computation. In our probabilistic approach, we are capable of deriving the formulas involved from the semantics of *trust chaining* from the principles of the Beta paradigm.

2.4 The Beta Paradigm

The Beta paradigm is based on the *beta distribution*, which in turn is deeply connected with Bayesian probability. As established before, we typically use the Bayesian interpretation of probability. Therefore, this section uses Bayesian methods until Section 2.4.4, where we discuss an alternative to the Bayesian methodology.

The Beta paradigm was introduced in Section 1.1.1 (Scope). Essentially, the Beta paradigm is the collection of assumptions, trust systems and examples where the *Beta model* applies to. Hence, we discuss the Beta model in this section. Furthermore, a core fraction of *Subjective Logic* falls under the Beta paradigm. For this reason, and the fact that we are using Subjective Logic extensively in Part I, we discuss a fraction of Subjective Logic in this section. Finally, we discuss other models that have a foundation in the Beta paradigm, but do not completely fall under the Beta paradigm. These models are, however, sufficiently close to the Beta paradigm, that our results (found under the Beta paradigm) can be inserted into their models without necessitating new insights.

The beta distribution is central to the Beta paradigm. The beta distribution is formally defined in Definition 6.7. For now it suffices to note that the beta distribution, based on s successes and f failures, denoted $\vartheta_{s,f}(x) = \frac{x^s \cdot (1-x)^f}{\int_0^1 y^s \cdot (1-y)^f \ \mathrm{d}y} \propto x^s \cdot (1-x)^f$. The $\propto$ symbol means "proportional to". If two functions are proportional to each other, they represent the same distribution.

2.4.1 The Beta Model

The Beta model is central to the Beta paradigm. Since the Beta paradigm is, in turn, central to our results, the Beta model is formally introduced in Chapter 6. In this section we informally introduce ideas and techniques of the Beta model that help the reader put abstract notions into the right perspective.

By definition, a target with high integrity is more likely to succeed than a target with low integrity, which by Bayesian logic means that after observing a success, the likelihood of the former has increased relative to the latter. In order to make this line of reasoning rigourous, and to be able to formulate it more precisely, we need to define the exact assumptions. We formulate the assumptions of the Beta paradigm more precisely, but refer the reader to Section 1.1.1 for motivation or explanation, and the preceding sections in this chapter for alternatives. Recall the three main restrictions in the paradigm, trust is interaction-oriented, interactions can objectively be labelled *successes* or *failures*, and subjects cannot deduce the internal states or motives of the target. We base the more precise and concrete assumptions on these three restrictions from Section 1.1.1

Every target has some behaviour or strategy, even though this strategy may be unknown to all other users, and perhaps even to the target himself. These strategies may range from completely deterministic (e.g. only fail every 3rd interaction) to completely probabilistic (e.g. always fail with probability 1/3), and everything in between. However, subjects cannot distinguish between these strategies. For both strategies, the only information available is that 1/3 of the interactions fail. The expected fraction of successes is called the *integrity*, which happens to be 1/3 for both strategies.

The restriction to interaction-oriented trust contained the assumption that interactions were similar. In other words, the integrity of a target is the probability of success for past interactions and for present and future interactions. Consider an algebraic analysis of the integrity parameter, given a certain number of observations. The first step is an application of Bayes' theorem, the second step assumes that all interactions happened independently. Let R be a random variable whose outcomes are integrity parameters, i.e. values in $[0, 1]$. Let $\overrightarrow{O} = O_1, \ldots, O_n$ be random variables denoting observations. Let f_E be the probability density function of E. If there are n observations:

$$f_R(x|\overrightarrow{O}) \propto P(\overrightarrow{O}|x) \cdot f_R(x) = f_R(x) \cdot \prod_{1 \leq i \leq n} P(O_i|x)$$

Now, by definition, $P(O_i = \text{S}|R = x) = x$ (and $P(O_i = \text{F}|R = x) = 1-x$), since the integrity parameter is the probability of success. So if s of the n interactions were successes, and the remaining $f = n-s$ are failures, we get (modulo a multiplicative factor):

$$x^s \cdot (1 - x)^f \cdot f_R(x)$$

Which means that (except for the factor $f_R(x)$) the beta distribution expresses the probability distribution over the integrity of the target. This is formally proven in Theorem 6.5. Hence, the beta distribution is suitable for representing trust opinions, as discovered by Mui and Mohtashemi [MM02] and Jøsang and Ismail [JI02]. They assume $f_R(x) = 1$, which means they assume the prior distribution to be

uniform, which is a good choice for two reasons. First, it is the prior distribution with maximal entropy (that makes it the best choice according to the principle of maximum entropy [Jay57]). Second, it allows for the simplest definition of trust aggregation: multiplication.

With a uniformly distributed prior, $f_R(x)$, it is almost immediate that *trust aggregation* is multiplication. The semantics of trust aggregation, is that the trust aggregation of two opinions yields an opinion that reflects the observations of the two original opinions. In other words, we define the trust aggregation of two trust opinions $f_R(x|O)$ and $f_R(x|O')$ simply as $f_R(x|O, O')$. We can derive that:

$$\begin{aligned} & f_R(x|O,O') && \propto P(O,O'|x)\cdot f_R(x)\cdot 1 \\ =& P(O|x)\cdot P(O'|x)\cdot f_R(x)\cdot f_R(x) && \propto f_R(x|O)\cdot f_R(x|O') \end{aligned}$$

The interesting thing to note is that we have not defined trust aggregation by defining how to compute it, but by providing a semantics and deriving its computation. The result is also derived formally, in Lemma 6.6.

The idea of representing trust opinions as beta distributions provides us with the simplest notion of a correctness trust model; the Beta model. The Beta model uses beta distributions to represent trust opinions, and trust aggregation is defined as multiplication. The authors of [JI02] named the Beta model the Beta reputation system, however, we see the Beta model as a model (which can indeed be applied to reputation systems), see Figure 1.2. We follow the authors of [ESN10] in referring to the model as the Beta model.

It is possible to select alternative representations for trust opinions, that are *isomorphic* with respect to trust aggregation. Rather than taking the actual beta distribution, one can take the corresponding successes and failures. More interesting is to have one parameter for the fraction of successes (success rate), and one parameter for the total number of interactions (weight). Alternatively, trust opinions can be represented with three parameters that have two degrees of freedom. The last representation is adopted by *Subjective Logic*, which is introduced in Section 2.4.2 and discussed in detail in Section 3.2.1. Their three parameters are belief, disbelief and uncertainty, where the ratio between belief and disbelief translates roughly to success rate and uncertainty to the inverse of weight.

In this thesis, we extend the Beta model by adding the remaining operators; trust chaining and the logical trust operations. The assumptions remain in the spirit of the Beta paradigm, notably that results of interactions are independent of each other, independent of the subject and independent of context. As with trust aggregation, we define these operators by their semantics, and derive their computation. This is in contrast to other trust models discussed in Section 2.4.3.

2.4.2 Subjective Logic

Subjective Logic [Jøs97] is a trust model with a wide range of operators. Amongst those operators are *consensus*, *discounting*, *multiplication*, *comultiplication* and *complement*; which correspond to trust aggregation, trust chaining, trust conjunction, trust disjunction and trust negation, respectively. Trust opinions in Subjective Logic are known as opinions or *belief triples*, and represented as a triple

(b, d, u), where $b + d + u = 1$. The three components represent belief, disbelief and uncertainty.

As Subjective Logic has grown more powerful over time, expressiveness and generality have increased to more complex notions. Currently, opinions need not be triples, but tuples, depending on the number of outcomes of interactions. Further, opinions now include a base rate, which allow for different priors than the uniform prior. Throughout the thesis, we study a fraction of Subjective Logic without these features, unless stated differently. However, every general conclusion regarding Subjective Logic regards the entire calculus, unless stated differently.

Subjective Logic's definition of consensus is based on beta distributions [Jøs02]. In fact, consensus over opinions is isomorphic to multiplication over beta distributions (i.e. trust aggregation). This property makes at least a fraction of Subjective Logic fall under the Beta paradigm.

The other operators are formalised differently, using something akin to Kleene logic [Kle50]. If we reinterpret an opinion (b, d, u) as "with probability b, P is true, with probability d, P is false, and with probability u, P is unknown", then we have mapped every triple to a probability distribution. Multiplication, comultiplication and complement map (as defined in [Jøs01][1]) to conjunction, disjunction and negation in Kleene logic. Discounting maps to an operator equivalent to $(P \Rightarrow Q \wedge \neg P \Rightarrow \text{unknown})$ [JP05]. The intuition for discounting is that, if the recommender tells the truth (i.e. P holds), then Q accurately describes the target, and if the recommender lies (i.e. $\neg P$ holds), then the target is unknown.

However, we axiomatically show in Chapter 5, that these two different views (beta distributions and three-valued logics) are not compatible. And we show probabilistically in Chapter 7, that the view as Kleene logic is incompatible with the Beta paradigm. There, we provide alternatives to the operators that are compatible with the Beta paradigm.

2.4.3 Other Models in the Beta Paradigm

In this section, we discuss models based on the Beta model, which are an extension or generalisation thereof. By extensions, we understand that extra functionality or information is added. For example, additional operations, or more data in trust opinions. By generalisations, we understand models that require fewer assumptions or that apply to more scenarios.

The Beta model was developed independently in [MM02, JI02]. The latter also provided an immediate generalisation, wherein recent interactions are taken as more relevant than less recent ones, using a so-called decay-factor.

A common generalisation of the Beta model, is to drop the notion that the prior distribution must be the uniform distribution. In Subjective Logic, a prior distribution can be selected based on a parameter called the base rate [JOO10]. Cer-

[1] Later, multiplication and comultiplication were defined differently in [JM05], and the notions from [Jøs01] were renamed as simple multiplication and simple comultiplication. The general ideas in this thesis are independent of the choice of either operator. Since the simple versions are closer to the theory they originate from, we opt to use these. The modern version tends to be more suitable for analysis of trust systems, but this falls outside of our scope.

tainTrust [Rie07] and CertainLogic [RHMV11] also have a single parameter, called base trust, to select a prior.

Another generalisation is achieved by dropping the restriction that we cannot derive internal states. ElSalamouny et al. propose Hidden Markov Models (HMM's) – defined in [BP66] – to represent internal states of targets [ESN10]. HMM's are Markov models with unknown parameters. That means that a target is assumed to be in a particular state with unknown integrity, but can switch to another state with another unknown integrity. If behaviour is observed to be bursty, e.g. periods of many successful interactions are interleaved with periods of many failures, this translates to a HMM with two states, one with a high integrity the other with a low integrity. The HMM approach uses Bayesian techniques similar to those in the Beta model, and the other restrictions from the Beta paradigm.

In [Sta10], trust aggregation uses beta distributions as in the Beta model. In the trust system they are interested in, the subject can determine the number of observed interactions, by sending fake requests. However, sending fake requests is costly. Hence, they use the probability distribution, not to optimise the amount of information received, but to minimise the amount of fake requests whilst maintaining sufficiently informative. This work falls out of the scope of the Beta paradigm, due to the differences in applications and goals.

Many models have an operator for trust chaining. Examples of models with trust aggregation based on the Beta model, and an operator for trust chaining are Subjective Logic [Jøs97], TRAVOS [TPJL06], CertainTrust [Rie07] and the trust model by Buchegger and Le Boudec [BLB04]. Subjective Logic, TRAVOS and CertainTrust apply exogenous filtering of recommendations, whereas Buchegger and Le Boudec apply endogenous filtering. As stated before, we prove that exogenous filtering should be applied (alone), in Theorem 8.9. Subjective Logic, TRAVOS and CertainTrust all have a definition where the resulting trust opinion is (isomorphic to) a beta distribution. In Subjective Logic and CertainTrust, the definition is provided as a computation with a justification (in probability theory), rather than from the principles of the Beta paradigm. In TRAVOS, the mechanism is based on a derivation of probabilistic relations. We show in Theorem 8.11, that none of these definitions are compatible with the Beta paradigm.

The logical trust opinions are also often represented with an operator. Examples of models with logical trust operations based on the Beta model, and logical trust operations are Subjective Logic [Jøs97] and CertainLogic [RHMV11]. As mentioned before, Subjective Logic has two sets of definitions, an older one (later renamed as simple) and a more recent one (named normal). The latter requires reasoning over the base rate. The definitions from CertainLogic are identical to the operations from Subjective Logic involving the base rate. In both formalisms, the definitions are provided as a computation with a justification, rather than derived from the principles of the Beta paradigm.

2.4.4 Dempster-Shafer Theory

The Beta model's objective is *trust aggregation* based on observation data. The correctness of the Beta model follows from the Bayesian interpretation of probability, via Bayes' theorem. However, there are two problems: To apply Bayesian inference, we need a prior (a distribution $f_R(x)$), which we typically do not have. And a more philosophical issue is that Bayesian methods are not necessarily accepted by frequentists (see e.g. [Fie06]). Frequentists may, for example, reject our strategy of treating the unknown integrity parameter as a probabilistic entity, on philosophical grounds. More generally, frequentists differ from Bayesianists by not allowing unknowns to be assigned probability (unless the outcomes can be measured repeatedly).

Formally reasoning about such unknowns poses a problem for frequentists. To illustrate, say you have an unfair coin. Due to the laws of physics, even the most unbalanced coin cannot land on one side more than 70% of the time. That means that the probability of heads (H) is $0.3 < P(H) < 0.7$. A Bayesianist would let U be a random variable that stands for the unfairness of the coin, such that if the coin has $U = 0.6$, $P(H) = 0.6$. A frequentist rejects this methodology. To formulate and reason about such a scenario poses a problem to him.

Frequentists may deal with this problem by generalising probability theory, to include not just probability, but also uncertainty. A famous way of doing this, is Dempster-Shafer theory [Dem67, Sha76]. In Dempster-Shafer theory, rather than having probability of X, i.e. $P(X)$, there is mass of X, belief in X and plausibility of X, denoted $m(X)$, $Bel(X)$ and $Pl(X)$. Belief, $Bel(X)$ is the minimum probability that an element of X holds. For example, $Bel(\{H\}) = 0.3$, as an unfair coin may give heads as little as 30% of the time. But also $Bel(\{H, T\}) = 1$, as a coin must always give heads or tails. Plausibility is the dual, both in the sense that $Pl(X) = 1 - Bel(\ X)$ and in the sense that its the maximal probability; e.g. $P(H) = 0.7$. For atomic events, mass and belief are the same, e.g. $Bel(\{H\}) = m(\{H\}) = 0.3$, and $Bel(\{T\}) = m(\{T\}) = 0.3$. However, for non-atomic events, mass is the difference between the belief and the sum of the constituent beliefs; in our example and $m(\{H, T\})$ is the remainder, 0.4. The quantity 0.4 can be seen as uncertainty regarding the fairness of the coin. Therefore, Dempster-Shafer theory has an effective method of denoting uncertainty. Each of mass, belief and plausibility is sufficient to derive the two others, so typically we only need a mass function $m : \mathcal{P}X \to [0, 1]$. Observe, for example, that $m(\{H\}) = 0.3$, $m(\{T\}) = 0.3$, $m(\{H, T\})$ is sufficient to describe the unfair coin flip, and the associated belief and plausibility. There is an operation in Dempster-Shafer theory which performs essentially the same operation as trust aggregation, namely Dempster's rule of combination. According to Dempster's rule of combination, the combined mass m of m_1 and m_2 is defined, for any event $A \neq \emptyset$:

$$m(A) = \frac{1}{\sum_{B \cap C \neq \emptyset} m_1(B) \cdot m_2(C)} \cdot \sum_{B \cap C = A} m_1(B) \cdot m_2(C).$$

Dempster's rule of combination collapses to probability theory, when only atomic events have non-zero probability. In that case, Dempster's rule of combination collapses to $m(A) = \frac{1}{\sum_B m_1(B) \cdot m_2(B)} \cdot m_1(A) \cdot m_2(A)$, which is remarkably similar to

our definition of trust aggregation (as exemplified in Section 2.4.1 and defined in Definition 6.9).

Dempster-Shafer theory is, at least superficially, similar to Subjective Logic, in the sense that a trust opinion (b, d, u) in Subjective Logic can be interpreted as a belief in Dempster-Shafer theory, as shown by Haenni [Hae06]. For all operations that we consider, except trust aggregation, Haenni provides a translation from trust opinions in Subjective Logics to beliefs in Dempster-Shafer theory. In the translation, Haenni translates (b, d, u) to a belief with masses $m(\{\text{S}\}) = b$, $m(\{\text{F}\}) = d$ and $m(\{\text{S}, \text{F}\}) = u$. Interestingly, Dempster's rule of combination is similar to, but distinct from, consensus in Subjective Logic. We believe the similarity is caused by the fact that Dempster-Shafer theory and Subjective Logic share similar intuitions, and the difference is caused by the fact that Subjective Logic is designed to adhere to Bayesian probability in case of consensus.

Dempster-Shafer theory is an important alternative to Bayesian methods for our application. There are formalisms with subtly different rules of combination [SF02]. However, Dempster-Shafer theory, its variations, and other generalisations of probability theory, do not have the rigourous formal foundation that probability theory itself has. Nevertheless, notions and intuitions formulated in Dempster-Shafer theory (and other rules of combination) can provide insights or explanations for our probabilistic notions, especially seeing that both are dealing with uncertainty and unknowns in probability theory.

Part I

The Axiomatic Method

3

Models and Axiomatisations

In Part I, we study the various trust operations: *trust aggregation*, *trust chaining*, and the *logical trust operations*. The trust operations can be applied in various combinations. A (nested) application of trust operations can be compared with another (nested) application of trust operations. To compare two expressions, it is possible to apply the definition of the operations, and reduce both sides to a single *trust opinion* which can be compared. Alternatively, it is possible to define a set of rules to transform one into the other. To illustrate the difference between these two alternatives, take an example in arithmetics, such as the question whether $3 \cdot 2 + 3$ is equal to $3 \cdot 3$. The first method is to compute that $3 \cdot 2 + 3 = 9$ and $3 \cdot 3 = 9$, thus they are equal. But it is not necessary to compute the answer, as the rule $x = x \cdot 1$ can be applied to get $3 \cdot 2 + 3 = 3 \cdot 2 + 3 \cdot 1$, and distributivity $(x \cdot y + x \cdot z = x \cdot (y + z))$ can be applied to get $3 \cdot 2 + 3 \cdot 1 = 3 \cdot 3$. The first method uses a model, and the second method uses an axiomatisation. Models provide a computation for an expression. An axiomatisation provides a set of equalities called *axioms*. A formal definition of the concepts surrounding axiomatisations is found in Section 3.2.3.

Axiomatisations have several advantages. An important advantage is that axioms can be studied in isolation from specific scenarios. If an operation is associative and commutative, then you know that order is not important for the operation, without needing further knowledge on the operation. Another advantage is that upon finding a counterintuitive or undesirable property, it is easier to identify and drop (or alter) the responsible axioms than to update the computation in the model correctly. To reap the benefits of the *axiomatic approach*, when provided with a model, it is possible to axiomatise the model. A complete axiomatisation is particularly useful, as all properties of the model are captured by the axiomatisation.

In Part I, we introduce many models and axiomatisations, we created a general overview in Figure 3.1. Figure 3.1 contains an overview of all the signatures[1] (rectangles), models (ellipses) and axiomatisations (hexagons) used in this thesis. Note the typographical differences between the names of models and axiomatisations, which is consistent with the typographic in the entire Part I, to assist in identifying their class. The fat hexagons are axioma schemes, and those with regular thickness represent finite axiomatisations. The relationships are provided as arrows. Signatures may have regular arrows to strictly more restricted signatures, accompanied by a description of the restriction. Models and axiomatisations may have solid arrows to weaker models or axiomatisations. If an axiomatisation is sound with respect to a model, there is an open arrow from the axiomatisation to the model. If

[1] Signatures determine the set of expressions we reason with. In particular a signature defines the set of constants and functions.

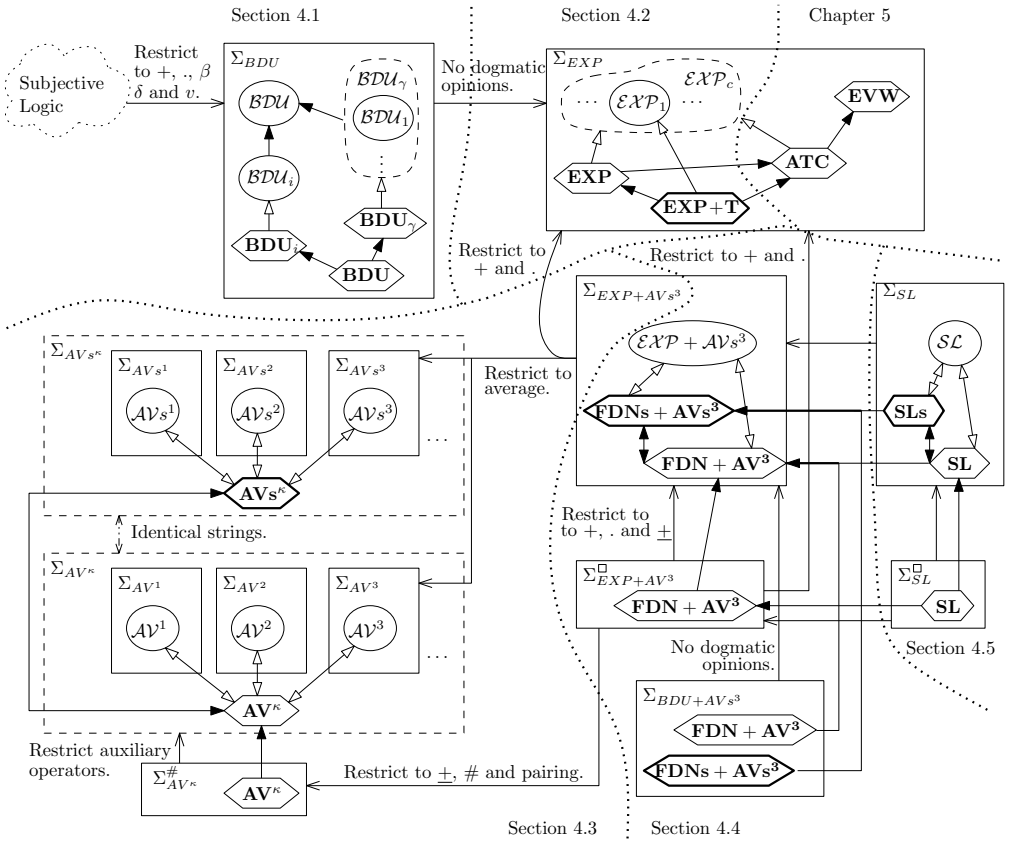

Figure 3.1: Overview of all signatures, models and axiomatisations in Part I.

that axiomatisation is furthermore complete, the open arrow will be bidirectional. Dashed rectangles or ellipses represent families of signatures or models. Arrows point towards the family, if they apply to all members. Figure 3.1 is furthermore partitioned, by the dotted lines, into 5 parts, each corresponding to a section in this thesis. The figure can be used as a quick reference for signatures, models and theories; either for an understanding of their relation to others or for finding their location in the thesis.

3.1 Representation of Operations

In this section, we introduce three different notations to express nested trust operations. The graphical representation only serves as a visual assistance, based on Figure 1.1. The other two representations are used for the axiomatisations and for the models.

Trust aggregation, trust chaining, trust conjunction, trust disjunction and trust negation can be seen as abstract operations over trust opinions. In order to axiomatise these operations, we introduce a *symbolic* representation of the operations. These symbols can form expressions, and the axioms are equations on these expressions. For example, *fusion*, denoted $\_ + \_$, is the symbolic representation of trust aggregation. The notion that order is irrelevant for trust aggregation can be expressed symbolically with the axiom $x + y = y + x$.

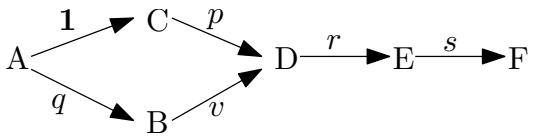

Figure 3.2: A simple trust network.

The symbol for arbitrary trust opinions is a variable (e.g. x), whereas specific trust opinions can be represented by constants (e.g. $\mathbf{1}$, as from Definition 3.1). The symbol for trust aggregation is fusion, denoted $\_ + \_$, and the symbol for trust chaining is *dilution*, denoted $\_ \cdot \_$. An example of a trust expression is $((\mathbf{1} \cdot p) + (q \cdot v)) \cdot r \cdot s$.

A complicated expression may be easier to understand graphically, hence we provide a graphical notation in Series Parallel Graphs (SPG's) [Duf65]. To define SPG's, we need to define the graph K_1 and the parallel and serial compositions of graphs that have a source and a sink. The graph K_1 consists of two vertices, called the source and the sink, and one directed edge from source to sink. The parallel composition G of two graphs G_1 and G_2 unifies the sources of G_1 and G_2 into the source of G, and unifies the sinks of G_1 and G_2 into the sink of G. The serial composition G of two graphs G_1 and G_2 unifies the sink of G_1 with the source of G_2. The source and sink of G are the source of G_1 and the sink of G_2. A SPG is a graph that can be recursively constructed from K_1 using serial composition and parallel composition. An example of a series parallel graph can be found in Figure 3.2.

The link between SPG's and expressions is simple: The variables and constants (i.e. trust opinions) are represented by K_1. Fusion (i.e. trust aggregation) is represented by parallel composition. Dilution (i.e. trust chaining) is represented by serial composition. The source of the network is the *subject*, and the *target* of the network is the sink. Hence, the expression $((\mathbf{1} \cdot p) + (q \cdot v)) \cdot r \cdot s$ represents the same trust network as Figure 3.2.

We can view a *trust model* as the semantics of such a network. That means that we provide a meaning to the network. The particular model we look at is *Subjective Logic*. There, direct trust opinions are represented as *belief triples* (or opinions), denoted (b, d, u). The trust operations take a pair of belief triples as argument, and return another belief triple. The implementation of trust aggregation is called *consensus*, denoted $\_ \oplus \_$. The implementation of trust chaining is called *discounting*, denoted $\_ \otimes \_$. The semantics (in Subjective Logic) of the network depicted in Figure 3.2 is $(((1/2, 0, 1/2) \otimes (b_p, d_p, u_p)) \oplus ((b_q, d_q, u_q) \otimes (0, 0, 1))) \otimes (b_r, d_r, u_r) \otimes (b_s, d_s, u_s)$. If two networks yield the same triple, they are semantically equivalent. A sound and complete axiomatisation of this fraction of Subjective Logic, therefore, equates two networks if and only if the two networks are semantically equivalent.

We have not yet defined the symbolic representation for the logical trust operations. The symbol for *trust conjunction* is *AND*, denoted $\_ \wedge \_$, the symbol for *trust disjunction* is *OR*, denoted $\_ \vee \_$ and the symbol for *trust negation* is *inverse*, denoted $\overline{x}$. An example of a trust expression is $\overline{p} \wedge (q \cdot (r + s))$.

The logical trust operations are naturally represented as binary trees. The leafs

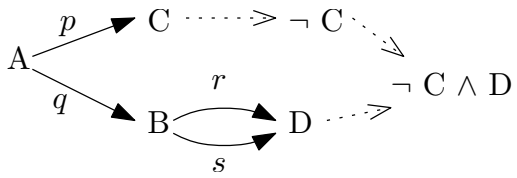

Figure 3.3: A trust network with logical trust operations.

of the binary trees contain *users*. A node can be the conjunction or disjunction of two subtrees, or a node can be the negation of a single subtree. The root of the tree contains the entire propositional formula. Notice that this is shown for the proposition $\neg C \wedge D$ in Figure 3.3, where the dotted arrows form the binary tree. The subject has a trust opinion of all the constituent users of the target, meaning that there are graphs with the subject as source and each leaf as sink. For $\neg C \wedge D$, this means that the subject has a trust opinion about C and a trust opinion about D, as also depicted in Figure 3.3, using solid lines. Hence, the entire graph has the subject as a source, and the target as sink, as depicted in Figure 3.3, where A is the source and $\neg C \wedge D$ is the target. Figure 3.3 depicts the trust network $\overline{p} \wedge (q \cdot (r + s))$. Note that the opinion of A about C is p, and the opinion of A about D is $q \cdot (r + s)$, therefore, the opinion of A about $\neg C \wedge D$ is $\overline{p} \wedge (q \cdot (r + s))$.

Of course, Subjective Logic has an implementation of the logical trust operations. Trust conjunction is implemented by *multiplication*, denoted $\_ \mathbin{⩓} \_$. Trust disjunction is implemented by *comultiplication*, denoted $\_ \mathbin{⩔} \_$. Trust negation is implemented by *complement*, denoted $\ominus(\_)$.

3.2 Preliminaries

The goal of Part I of this thesis is to axiomatise trust via trust aggregation, trust chaining and the logical trust operations. Before we do so, we introduce Subjective Logic, a tool to reason with and visualise Subjective Logic and the notion of axioms in this section. First we introduce Subjective Logic and its notion of a trust opinion, belief triples, and we discuss its relation to trust. We will formally define the operations in Subjective Logic corresponding to trust aggregation, trust chaining and the logical trust operations, and discuss their properties. There are several subtly different fractions of Subjective Logic that we consider; these are different *models*. We furthermore formally introduce concepts such as axiomatisations, derivations and theories. We informally discuss the relation between models and axiomatisations.

3.2.1 Subjective Logic

Subjective Logic is a formalism to denote trust opinions, and to do calculus with opinions [Jøs97]. We call the objects representing trust opinions *belief triples*, although the name opinion is preferred, to distinguish them from the more general notion of trust opinions that we discuss throughout the thesis. Furthermore, we may refer to a trust opinion represented by a belief triple as a belief. Belief triples

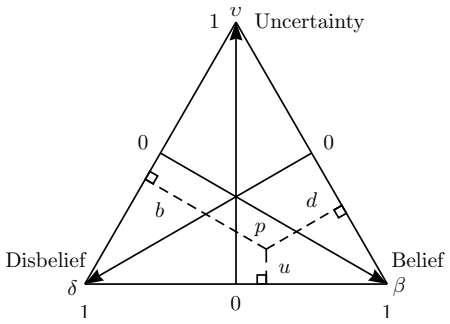

Figure 3.4: The Subjective Logic triangle.

represent only trust and distrust, and do so with particular uncertainty. In Subjective Logic, trust, distrust and uncertainty form a triple (b, d, u), with $b+d+u = 1$. Each such triple can be represented as a point in a triangle with corners β, δ and υ (see Figure 3.4). Coordinate b of a point $p = (b, d, u)$ determines the (perpendicular) distance between p and side $\delta\upsilon$. Likewise, d determines the distance between p and side $\beta\upsilon$, and u the distance between p and $\beta\delta$. Some examples: point β has coordinates (1, 0, 0) and represents full belief (or full trust), the middle point between β and δ is $(1/2, 1/2, 0)$ and represents the fully certain belief that there is as much belief (trust) as disbelief (distrust) in the trustee. In this thesis we will restrict ourselves to the fragment of Subjective Logic in which the coordinates consist of rational numbers. This simplifying assumption enables axiomatic reasoning.

Jøsang proposed two operators in [Jøs97], called consensus and discounting, which are intended to model aggregation and trust chaining, respectively. When a subjects hold two independent beliefs about a target simultaneously, then consensus allows these two belief triples to be merged into a new belief triple, which has less uncertainty. Consensus is particulary useful for aggregating beliefs:

Example 3.1. Assume that subject A had some interactions with user B at some point, and formed belief (b, d, u) on the basis of these interactions. Later A had more interactions with B, which lead him to the belief (b', d', u'). He can then apply consensus on (b, d, u) and (b', d', u'), to get $(b, d, u) \oplus (b', d', u')$, which represents A's belief about B on the basis of all interactions.

The definition of consensus in Subjective Logic is:

$$(b, d, u) \oplus (b', d', u') = \left(\frac{bu' + b'u}{u + u' - uu'}, \frac{du' + d'u}{u + u' - uu'}, \frac{uu'}{u + u' - uu'}\right) \tag{3.1}$$

The idea behind this definition is that the belief component of the aggregated belief triple corresponds to the sum of the belief components of the individual belief triples, weighted by the uncertainty of the other belief triple: $bu'+b'u$. Similarly, we obtain for the disbelief component $du'+d'u$. The reduced uncertainty becomes the multiple of the individual uncertainties uu'. Finally, these values are normalised as to retain the invariant $b+d+u = 1$. This gives denominator $bu'+b'u+du'+d'u+uu'$, which equals $u + u' - uu'$.

The calculus in Subjective Logic is meaningful only when the definitions of the operators are. There are several ways to analyse the validity and relevance of the operators, and studying their defining axioms is a common way to do so.

The next Subjective Logic operation is discounting. Discounting is the operation intended to model trust chaining.

Example 3.2. Assume that user B said that his belief about C is (b', d', u'). Further assume that A's belief about B's capabilities to refer to other users is (b, d, u). Then A's derived belief about C is $(b, d, u) \otimes (b', d', u')$. If A strongly trusts user B, he will form an belief triple similar to (b', d', u') about C. If A strongly distrusts user B, he will almost completely disregard B's claimed belief triple, leaving us with a belief with a lot of uncertainty about C. Naturally, consensus can be applied on beliefs resulting from *recommendations*, so A can combine several different users' beliefs about C, possibly with some prior personal belief. In other words, A can have belief $z \oplus (x \otimes y)$, taking the consensus of a discounted belief $(x \otimes y)$ and his own belief (z).

The definition of discounting in Subjective Logic is:

$$(b, d, u) \otimes (b', d', u') = (bb', bd', bu' + d + u) \tag{3.2}$$

The idea behind this operator is that the belief b in the referring user determines the chained belief bb' and the chained disbelief bd'. The chained uncertainty $bu' + d + u$ follows from the invariant $b + d + u = 1$.

The third Subjective Logic operator is multiplication. Multiplication is designed to model trust conjunction.

Example 3.3. Let A be a customer, B be a seller and C be a delivery service. The subject, A, has belief (b, d, u) about B, and belief (b, d, u) about C. The packet will only arrive intact and on time, if both B succeeds (i.e. sends the packet intact and on time) and C succeeds (i.e. does not break the package or delay delivery). Of course, A is interested in the delivery itself, and thus neither B alone or C alone is the target, rather their conjunction is. The belief of A about the conjunction of B and C is $(b, d, u) \oslash (b', d', u')$. The resulting beliefs should be less trusted than either of the original beliefs, since if either B or C fails, the whole interaction fails.

The definition of multiplication is:

$$(b, d, u) \oslash (b', d', u') = (bb', d + d' - dd', bu' + b'u + uu') \tag{3.3}$$

Comultiplication is designed to model trust disjunction and has a similar rationale:

$$(b, d, u) \otimes\!\!\!\!\bigcirc\, (b', d', u') = (b + b' - bb', dd', du' + d'u + uu') \tag{3.4}$$

The last Subjective Logic operator that we will use in our study is the complement operator. The complement operator models trust negation. This operator swaps belief and disbelief:

$$\ominus\,((b, d, u)) = (d, b, u) \tag{3.5}$$

If the belief triples relate to the trust opinions of users, and consensus and discounting correctly model aggregation and trust chaining, then Subjective Logic is exactly what is needed to model trust networks. In other words, if Subjective Logic

has the rigour of correctness trust models, then Subjective Logic is the answer to our research question.

In trust networks, users are able to chain trust, and aggregate information, as modelled by discounting and consensus, respectively. The purposes of the axiomatisations are therefore twofold: If consensus and discounting correctly model aggregation and trust chaining, then its axioms are the axioms of trust. Otherwise, there should be axioms that are not self-evident, or even false, when applied to aggregation and chaining. As will turn out, Subjective Logic seems too strong, as there are truths that seem not self-evident, but not contradictory to self-evident truths. It is, however, not impossible that more self-evident truths are found that contradict Subjective Logic. An analysis of this can be found in Section 5.2.

The reason that we suspect that consensus is a good candidate for aggregation, is that there is a strong link between belief triples over consensus and beta distributions [AS64], as also shown in [JJ98]. Beta distributions form the core of the beta paradigm, in which we operate.

We have two collections of *experiments* about a single user. The first collection has s successes, and f failures, and the second collection has s' and f', respectively. We may aggregate these collections to $s + s'$ successes and $f + f'$ failures. This allows us to map belief triples to beta distributions. Let us map (b, d, u) to (s, f) as follows:

$$\pi((b, d, u)) = (\frac{b}{u}, \frac{d}{u})$$

then

$$\pi^{-1}((s, f)) = (\frac{s}{s+f+1}, \frac{f}{s+f+1}, \frac{1}{s+f+1})$$

Consequently, π is an isomorphism between beta distributions under pairwise addition, and belief triples with consensus:

$$\begin{aligned}
&\pi(\pi^{-1}((s, f)) + \pi^{-1}((s', f'))) \\
=&\pi((\frac{s}{s+f+1}, \frac{f}{s+f+1}, \frac{1}{s+f+1}) + (\frac{s'}{s'+f'+1}, \frac{f'}{s'+f'+1}, \frac{1}{s'+f'+1})) \\
=&\pi(\frac{s+s'}{s+s'+f+f'+1}, \frac{f+f'}{s+s'+f+f'+1}, \frac{1}{s+s'+f+f'+1}) \\
=&(s+s', f+f')
\end{aligned}$$

The *isomorphic* mapping π has previously been formulated in [Jøs97]. Recall that the expected value of a beta distribution based on s successes and f failures (i.e. $\alpha = s + 1$ and $\beta = f + 1$) has an expected value of $\frac{s+1}{s+f+1}$. This means that we can tie an expected value to a belief triple (b, d, u), via π, namely $\frac{b+u}{1+u}$.

As consensus (of belief triples) is isomorphic to pairwise addition (of beta distributions), it has the same algebraic structure as pairwise addition. There are, however, more isomorphic mappings. We present a class of mappings π_c, for $c \in \mathbb{R}^+$ as follows:

$$\pi_c((b, d, u)) = (c\frac{b}{u}, c\frac{d}{u})$$

$$\pi_c^{-1}((s, f)) = (\frac{s}{s+f+c}, \frac{f}{s+f+c}, \frac{c}{s+f+c})$$

There is no reason why one value of c should be inherently better than another. Jøsang concentrated on the mapping π_1 in [JJ98]. Since 1 is the identity element of multiplication, and thus requires fewer symbols, we have no good reason to stray from that choice. An interesting fact about the mappings is that, due to transitivity of isomorphisms, $\tau_c = \pi_c^{-1} \circ \pi$ is also an isomorphism over consensus. In other words,

$$\tau_c((b,d,u)) = (\frac{b}{b+d+cu}, \frac{d}{b+d+cu}, \frac{cu}{b+d+cu})$$

is an automorphism on belief triples under consensus. The underlying intuition is that, with regard to consensus, relative certainty is important, not absolute certainty. That τ_c is an automorphism over complement is immediate.

There is no such automorphism over discounting. We can immediately verify that. We have that $(b,d,u) \otimes (b',d',u') = (bb', bd', bu' + d + u)$, but:

$$\begin{aligned}
&\tau_c^{-1}(\tau_c((b,d,u)) \otimes \tau_c((b',d',u'))) \\
=&\tau_c^{-1}((\frac{b}{nf}, \frac{d}{nf}, \frac{cu}{nf}) \otimes (\frac{b'}{nf'}, \frac{d'}{nf'}, \frac{cu'}{nf'})) \\
&\quad \text{where } nf = b+d+cu, \text{ and } nf' = b'+d'+cu' \\
=&\tau_c^{-1}((\frac{b}{nf}\frac{b'}{nf'}, \frac{b}{nf}\frac{d'}{nf'}, \frac{b}{nf}\frac{cu'}{nf'} + \frac{d}{nf} + \frac{cu}{nf})) \\
&\quad \text{where } nf = b+d+cu, \text{ and } nf' = b'+d'+cu' \\
=&(\frac{bb'}{p}, \frac{bd'}{p}, \frac{bu' + \frac{db'}{c} + \frac{dd'}{c} + du' + ub' + ud' + ucu'}{bb' + db' + bu' + \frac{db'}{c} + \frac{dd'}{c} + du' + du' + ub' + ud' + ucu'}) \\
&\quad \text{where } p = bb' + bd' + bu' + \frac{d(b'+d'+cu')}{c} + \frac{u(b'+d'+cu')}{c} \\
\neq&(bb', bd', bu' + d + u), \text{ for } c \neq 1
\end{aligned}$$

Hence, τ_c is not an automorphism of belief triples over discounting. Similar equations show that τ_c is not an automorphism over multiplication and comultiplication. Therefore, different choices of π_c yield different models. In other words, we can map experiments to belief triples in several ways, yielding different results.

As mentioned before, we will pick $c = 1$, thus mapping a single *success*, $s = 1$ and $f = 0$, to $(1/2, 0, 1/2)$ and a single *failure*, $s = 0$ and $f = 1$, to $(0, 1/2, 1/2)$.

Definition 3.1 (Experiment)**.** A beta distribution with one success and no failures, that is $\alpha = 2$ and $\beta = 1$, is referred to as the *successful experiment*, denoted **1**. A beta distribution with no successes and one failure, that is $\alpha = 1$ and $\beta = 2$, is referred to as the *failed experiment*, denoted **0**. If we map beta distributions to belief triples with π, then **1** denotes $(1/2, 0, 1/2)$ and **0** denotes $(0, 1/2, 1/2)$.

By changing c in the mapping, one changes the interpretations of **0** and **1** in Subjective Logic. Hence picking c or picking **0** and **1** is essentially the same. In this section, we will not yet be able to reason axiomatically about the effects of picking **0** and **1**, but we will do so in section 4.2. There we will see that the axiomatic insight in the experiments is clearer than the above technical modelling approach, since the latter tends to get cluttered with nested divisions.

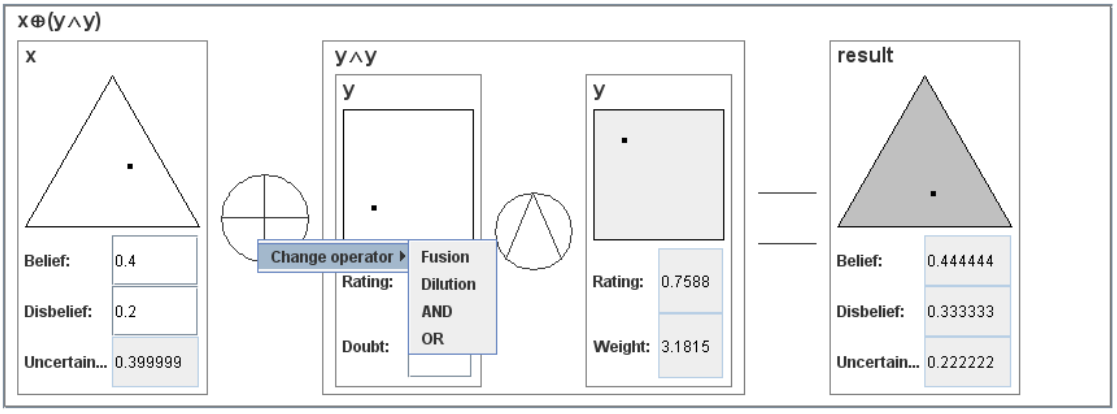

Figure 3.5: A screen capture of a trust network.

3.2.2 SLVisualiser

Subjective Logic is an intuitive model for computing a belief triple that reflects a trust network. In this thesis, we are not primarily interested in computing trust networks, but in understanding trust networks and their relations. For example, given an expression, how does the result change if we change one belief triple in the network. We study the theoretical implications of trust networks in Subjective Logic in Chapter 4. However, to get a more practical view on trust networks in Subjective Logic, it may also be helpful to study these trust networks graphically. The *SLVisualiser* tool[2] allows a user to construct and alter trust networks, and study resulting belief triples in different graphical representations.

The SLVisualiser is a tool with a graphical user interface. Its defining strength is that formulae can be created, altered, computed and viewed via a graphical user interface, as depicted in Figure 3.5. Users can create arbitrary formulae, potentially with duplicated belief triples (e.g. $(x \oplus y) \otimes x$). Each element has a graphical and numerical representation, and users can input values via both representations. The advantage of numerical input is that belief triples can be input with high precision. The advantage of graphical input is that users can continuously modify belief triples, and immediately see the result of the trust network with that belief triple. The user interface not only shows the beliefs, but furthermore provides the symbolical notation of the trust network (see top left corner of each belief triple in Figure 3.5, e.g. $x + (y \wedge y)$).

The SLVisualiser not only allows representation of belief triples as Subjective Logic triangles, but allows three alternative representations. The first alternative represents belief triples as pair of rating and doubt, where doubt corresponds to uncertainty and rating to the ratio between belief and disbelief. The second alternative representation also depends on rating, but rather than doubt, it uses weight. A weight of 0 corresponds to a doubt (or uncertainty) of 1, and in the limit weight ∞ corresponds to doubt of 0. In general $\frac{1}{1+weight} = doubt$. Since weight has no upper bound, we represent weight in a hyperbolic projection. In the third alternative representation, a belief triple can be represented as a beta distribution. We have

[2]The tool is available at http://satoss.uni.lu/members/tim/SLVisualizer.php.

seen the relation between belief triples and beta distributions in Section 3.2.1. All four representations are present in Figure 3.5, in order, from left to right.

In order to reason about equality of formula, we allow belief triples to be reused. Not only can belief triples be reused within one trust network, but also across trust networks. That means that it is possible to construct a trust network $x \oplus y$ and a trust network $y \oplus x$, and to visualise the fact that a change in x has the same effect in either trust network. It is even possible to nest such equations. Put differently, the tool allows the user to obtain intermediate results of computations.

3.2.3 Axiomatisation

Before we define models and axioms formally, we need terms over which the models and axioms range. The terms are defined by a signature, usually denoted by Σ. A signature Σ is a set of constant and function symbols, with their arities. We assume the signature(s) are given, and define models and theories on top of signatures. A model defines the semantics of the terms:

Definition 3.2 (Model)**.** A *model* $\mathcal{M}$ consists of a set of elements X, and a set of constants in X and functions on X. If the constants and functions of the model yield a signature Σ_X, we call Σ_X the signature $\mathcal{M}$. We refer to X as the *carrier set* of $\mathcal{M}$. If we apply all the functions in a term y in the specified order, then it yields an element in X. If two terms y and z in a model yield the same element in X, we write $\mathcal{M} \models y = z$.

Note that every finite term in a model yields exactly one element (called the solution) in X, as it consists only of constants and function applications. Possible models of Subjective Logic are defined by the sets of belief triples, together with consensus and discounting. As we will see later, there are different relevant carrier sets of belief triples possible.

An axiomatisation of a model is a set of equations that should be true in the model. An axiom of a model must, therefore, either be proven as true in the model, or be used to alter or reject a model.

Definition 3.3 (Axiomatisation)**.** An *axiomatisation* T (alternatively, a set of axioms or an axiom scheme) is a set of (conditional) equations over a given signature Σ.

It is possible to use the axioms to prove equalities that are not axioms themselves. This is captured by the notion of a *theory*:

Definition 3.4 (Theory)**.** Let T be an axiomatisation over a signature Σ. Let s, t and u be (possibly open) terms in Σ. We denote that an equation $s = t$ is derivable from axiomatisation T, as $T \vdash s = t$.

ax If $s = t$ is an equation in T, then $T \vdash s = t$.

cond If $\bigwedge_i \varphi_i \Rightarrow s = t$ is a conditional equation in T and for all i, $T \vdash \varphi_i$, then $T \vdash s = t$.

sub If $T \vdash s = t$ then, for any valid substitution σ, $T \vdash s[\sigma] = t[\sigma]$.

ref $T \vdash t = t$.

sym If $T \vdash s = t$ then $T \vdash t = s$.

trans If $T \vdash s = t$ and $T \vdash t = u$ then $T \vdash s = u$.

func Let f be an n-ary function, if $T \vdash s = t_i$, then
$T \vdash f(t_1, \ldots, t_{i-1}, s, t_{i+1}, \ldots, t_n) = f(t_1, \ldots, t_n)$.

We formulate axioms over the operations fusion and dilution. The axioms of fusion and dilution should match our intuition of trust aggregation and trust chaining. Like consensus and discounting form a model of aggregation and chaining, fusion and dilution are the theory of aggregation and chaining.

In this thesis, two approaches are used to reason about trust aggregation and trust chaining. The first approach is to take a model, formulate its axioms, and study the axioms. Obviously, if the model correctly implements aggregation and chaining, the axioms are automatically the theory of aggregation and chaining. Seeing whether the model is correct is difficult, and the axioms are a great tool to study the properties of the model. We can study the axioms, and check whether they are self-evident, true, false, counterintuitive, too weak, too strong, etc. This approach can be found in Chapter 4. Another approach is to formulate properties of trust aggregation and trust chaining by immediately formulating axioms of fusion and dilution. We follow this approach in Chapter 5.

4

Axiomatisation of Subjective Logic

Subjective Logic can be seen as a semantics for trust networks, examples of which are graphically depicted in Figures 3.2 and 3.3. Another way of saying that, is to say that Subjective Logic is a trust model. In this part of the thesis, we are interested in axiomatisations (or theories), rather than models. The goal of this chapter is to provide a complete axiomatisation of Subjective Logic. That means that the axiomatisation equates exactly those trust networks that have the same semantics (i.e. that result in the same trust opinion).

The operations that we analyse first, are *trust aggregation*, *trust chaining* and trust negation. The symbols or *operators* that we denote these operations with are *fusion*, *dilution* and *inverse*. The first axiomatisation of these three operators is provided in Section 4.1. That axiomatisation is based on *dogmatic opinions*, opinions without uncertainty. There, we identify issues arising from the choice of basing the axioms on the dogmatic opinions. Then, in Section 4.2, we alter the axioms from Section 4.1 to reason about non-dogmatic opinions. There, we discuss problems that arise when trying to formulate a complete axiomatisation of *Subjective Logic* with non-*dogmatic beliefs*, using only fusion, dilution and inversion. To circumvent these problems, we introduce a new operation – opinion mean – and a new operator – opinion average – in Section 4.3. We also introduce a generalisation of these concepts: tuple mean and tuple average, respectively. The tuple average (and opinion average) completely axiomatises the tuple mean (and opinion mean). In Section 4.4 we link the fusion, dilution and inversion to tuple averaging. Using that link, we can provide a complete axiomatisation of the three operators; fusion dilution and inversion. In that section, we also study properties of the axiomatisation. Finally, in Section 4.5, we link *AND* and *OR* – the operators for trust conjunction and trust disjunction. to tuple averaging. That allows us to provide a complete axiomatisation of these two operators.

4.1 Dilution and Fusion of Boundary Opinions

In this section we give a finite axiomatisation of fusion, dilution and inversion, based on the three extremal trust values $\beta = (1, 0, 0)$, $\delta = (0, 1, 0)$ and $\upsilon = (0, 0, 1)$. These three constants correspond to the three corners of the trust triangle in Figure 3.4 and denote full trust, full distrust and full uncertainty, respectively.

Definition 4.1 (Signature Σ_{BDU})**.** We define the signature Σ_{BDU} as:

$$\varphi ::= \beta \mid \delta \mid \upsilon \mid \varphi + \varphi \mid \varphi \cdot \varphi \mid \overline{\varphi}$$

Dogmatic opinions.

A complicating factor when performing calculations in Subjective Logic is that the fusion of two *dogmatic beliefs* is troublesome, since $(b, d, 0) \oplus (b', d', 0)$ gives $(\frac{0}{0}, \frac{0}{0}, \frac{0}{0})$. In order to deal with consensus of dogmatic opinion in Subjective Logic, we consider two amendments of the consensus operation. The first concerns the extension with a *limit construction* and the second concerns the introduction of *inconsistencies*. The limit construction is due to Jøsang et al. [JMP06, JDV03]. They provide the following definition for the fusion of dogmatic belief triples:

$$(b, d, 0) \oplus (b', d', 0) = \left(\frac{\gamma b + b'}{\gamma + 1} , \frac{\gamma d + d'}{\gamma + 1} , 0 \right).$$

According to Jøsang et al., γ is defined by $\gamma = \lim_{u,u' \to 0} \left(\frac{u'}{u}\right)$. It expresses the relative dogmatism between the expressions $(b, d, 0)$ and $(b', d', 0)$ or, rather, between the *users* expressing these dogmatic beliefs. The higher the value of γ, the higher the relative weight of belief triple $(b, d, 0)$ in a fusion with $(b', d', 0)$. The "default value" of γ is 1, meaning that in general dogmatic values are averaged. We shall denote this extension of Subjective Logic by $\mathcal{BDU}_\gamma$.

The second interpretation of the fusion of dogmatic beliefs is based on inconsistencies, as introduced by Alcalde and Mauw [AM09]. We extend Subjective Logic with the special element i, which stands for inconsistency. This element is the result of fusing two contradictory dogmatic beliefs, such as $\beta + \delta$. We set $(b, d, 0) \oplus (b', d', 0) = i$ and assume that inconsistencies proliferate through Subjective Logic expressions, i.e., $x + i = i + x = x \cdot i = i \cdot x = i$, for every expression x. Consequently, we have $\beta + \beta = i$. Further, we set $\overline{i} = i$. We denote this extension by $\mathcal{BDU}_i$.

Axiomatisation.

The basic axioms for fusion and dilution are given in Figure 4.1, ranging over terms in Σ_{BDU}. Axioms (**B1**) and (**B2**) express that the fusion operator is commutative and associative. This means that the fusion of opinions does not depend on the order in which the trust opinion are aggregated. Axioms (**B3**) and (**B4**) capture associativity and left-commutativity of the dilution operator. They express that in a trust chain the order in which the referral trust opinions (graphically: edges not ending in the source) are combined is irrelevant. Further, the last element in a trust chain, which expresses functional trust, cannot be mixed with the referral trust opinions. Axioms (**B5**) and (**B6**) define that full uncertainty behaves like a zero element. Adding a fully uncertain opinion to an opinion x does not give any extra information. The symmetric version of axiom (**B6**), $x \cdot \upsilon = \upsilon$, is also valid (Proposition 4.1). The next two axioms define the properties of full trust. Axiom (**B7**) states that full trust fused with itself remains full trust because this is the element with maximal trust. Axiom (**B8**) expresses that full trust is a left-unit for dilution. This follows from the assumption that if we fully trust another user, we adopt his opinion without any hesitation. The definition of full distrust follows a similar reasoning. Full distrust fused with itself remains full distrust (axiom (**B9**)).

(**B1**) $x + y = y + x$	(**B5**) $x + \upsilon = x$	(**B7**) $\beta + \beta = \beta$
(**B2**) $x + (y + z) = (x + y) + z$	(**B6**) $\upsilon \cdot x = \upsilon$	(**B8**) $\beta \cdot x = x$
(**B3**) $x \cdot (y \cdot z) = (x \cdot y) \cdot z$		(**B9**) $\delta + \delta = \delta$
(**B4**) $x \cdot (y \cdot z) = y \cdot (x \cdot z)$		(**B10**) $\delta \cdot x = \upsilon$
(**I1**) $\overline{\overline{x}} = x$	(**I3**) $\overline{\upsilon} = \upsilon$	(**I5**) $\overline{x \cdot y} = x \cdot \overline{y}$
(**I2**) $\overline{x + y} = \overline{x} + \overline{y}$	(**I4**) $\overline{\delta} = \beta$	

Figure 4.1: Basic axioms of Fusion and Dilution and basic axioms of the Inverse operator (**BDU**)

Finally, if we consider the opinion of someone whom we distrust, it will give us no information at all (axiom (**B10**)).

There are several properties that we will not consider because they are not valid in Subjective Logic. An example is that dilution is not fully commutative. This can be seen intuitively by a simple example. Assume that user A fully trusts user B's opinion on user C and assume that B fully distrusts C. Then A should also fully distrust C. However, if we swap the values, i.e., A has full distrust in B, who fully trusts C, then A should not necessarily (dis)trust C, so $\beta \cdot \delta \neq \delta \cdot \beta$.

From the above example we can also derive that β is not a right-unit. Further, the symmetric version of axiom (**B10**), $x \cdot \delta = \upsilon$, does not hold either, since $\beta \cdot \delta = \delta$.

In order to capture more properties of the dilution and fusion operators we give an axiomatisation of inversion in Figure 4.1. Axiom (**I1**) expresses that double inversion is the identity. Axiom (**I2**) states that inversion distributes over fusion. This means that trust and distrust are treated similarly when fusing trust opinions. Axiom (**I3**) states that if one has full uncertainty (thus neither trust nor distrust), inversion has no effect. This also stresses that υ is a zero element. Axiom (**I4**) expresses the duality of trust and distrust. Finally, we see that negation satisfies a particular semi-distributivity property over dilution (**I5**). This property expresses that the ratio between the trust and distrust component of a dilution only depends on the ratio of trust and distrust in the final element of the chain, which corresponds to functional trust. The set of axioms from Figure 4.1 is denoted by **BDU**.

The statement $x \cdot \upsilon = \upsilon$ denotes that if anyone proclaims to have complete uncertainty, the resulting opinion is also complete uncertainty. Either that person truthfully stated to have complete uncertainty, or that person was lying and he has an unknown opinion, in either case no information is gained. However, $x \cdot \upsilon = \upsilon$ need not be added to the axioms, as it is a proposition.

Proposition 4.1. $\mathbf{BDU} \vdash x \cdot \upsilon = \upsilon$

Proof. $\mathbf{BDU} \vdash x \cdot \upsilon = x \cdot (\delta \cdot \delta) = \delta \cdot (x \cdot \delta) = \upsilon$ □

Soundness.

Next, we study which axioms are sound in the completed variants of Subjective Logic. First, we observe that Subjective Logic with limit construction does not satisfy associativity of the fusion operator (**B2**). If we take $\gamma = 1$, we can use associativity to derive $(1/2, 1/2, 0) = (1, 0, 0) \oplus (0, 1, 0) = (1, 0, 0) \oplus ((0, 1, 0) \oplus (0, 1, 0)) =$

$((1,0,0) \oplus (0,1,0)) \oplus (0,1,0) = (1/2, 1/2, 0) \oplus (0,1,0) = (1/4, 3/4, 0)$. The remaining axioms are satisfied by this model.

Theorem 4.2. *Subjective Logic with limit construction $\mathcal{BDU}_\gamma$ is a model of* **BDU** *minus* (**B2**).

Proof. Axiom (**B1**) follows by observing the symmetry in the definition of the fusion operator (e.g. $\frac{bu'+b'u}{u+u'-uu'} = \frac{b'u+bu'}{u'+u-u'u}$). Axioms (**B3**) and (**B4**) are proven by calculating that $((b,d,u)\otimes(b',d',u'))\otimes(b'',d'',u'')$, $(b,d,u)\otimes((b',d',u')\otimes(b'',d'',u'))$ and $(b',d',u') \otimes ((b,d,u) \otimes (b'',d'',u'))$ are all equal to $(bb'b'', bb'd'', 1 - bb'b'' - bb'd'')$. The remaining axioms trivially follow from the definitions of the constants and operators involved. For instance, (**I5**) follows from $\overline{(b,d,u) \otimes (b',d',u')} = \overline{(bb', bd', d + u + bu')} = (bd', bb', d + u + bu') = (b,d,u) \otimes (d',b',u') = (b,d,u) \otimes \overline{(b',d',u')}$. □

Next, we look at the validity of the axioms for Subjective Logic extended with the constant i, which corresponds to inconsistency. Contrary to the previous model, this model satisfies commutativity of the fusion operator. However, it does not satisfy axioms (**B6**), (**B7**), (**B9**) and (**B10**). One can derive, for instance, that $\beta + \beta = (1,0,0) + (1,0,0) = i \neq \beta$.

Theorem 4.3. *Subjective Logic with inconsistency $\mathcal{BDU}_i$ is a model of* **BDU** *minus* (**B6**), (**B7**), (**B9**) *and* (**B10**).

Proof. Axiom (**B1**) follows from the symmetry in the definition of the fusion operator. Axiom (**B2**) follows from a simple case distinction. If at least one of the summands equals i or at least two of the summands are dogmatic opinions, then it follows from the persistence of inconsistencies and the fact that dogmatic opinions prevail in a fusion context. In the other case it follows from a simple calculation. For instance, it is easy to calculate that the first components of $((b,d,u) + (b',d',u')) + (b'',d'',u'')$ and $(b,d,u) + ((b',d',u') + (b'',d'',u''))$ are both equal to $\frac{bu'u''+b'uu''+b''uu'}{uu'+uu''+u'u''-2uu'u''}$. Likewise for the other components. The proofs of the other axioms follow from straightforward calculations and the propagation of inconsistencies. □

The axiomatisation provided above is a revision of the axiomatisation from [AM09]. In particular we have omitted the ordering axioms and added the axioms (**B4**) and (**I5**). That paper also considers a number of possible models of Subjective Logic, especially models where inconsistencies do not fully propagate.

Completeness

Even though the presented axioms capture many of the properties of Subjective Logic, they do not form a complete axiomatisation. Calculations with non-extremal values are not derivable, e.g., $(1/2, 0, 1/2) \cdot (0, 1/2, 1/2) = (0, 1/4, 3/4)$. It is clear that, in order to complete this axiomatisation, non-extremal values have to be added. In the following sections we will develop an axiomatisation based on the constants $(0, 1/2, 1/2)$ and $(1/2, 0, 1/2)$. We will show that these two values, plus an *averaging operator* will suffice for a complete axiomatisation.

4.2 Dilution and Fusion of Opinions from Experiments

As established in the previous section, the fusion operator is difficult to axiomatise due to the existence of *dogmatic opinions*. Two dogmatic *belief triples*, say $(b, d, 0)$ and $(b', d', 0)$, are fused to $(\frac{0}{0}, \frac{0}{0}, \frac{0}{0})$. In other words, there is a divide by zero, when applying consensus to dogmatic beliefs. Two solutions were offered. We can denote the result of a division by zero by a special inconsistency symbol, in which case $\mathbf{BDU} \vdash \beta + \beta = \beta$, but $\mathcal{BDU}_i \not\models (1,0,0) \oplus (1,0,0) = (1,0,0)$. On the other hand, the limit model is not associative, as $\mathbf{BDU} \vdash (x + x) + y = x + (x + y)$ but $\mathcal{BDU}_\gamma \not\models ((b,d,u) \oplus (b,d,u)) \oplus (b',d',u') = (b,d,u) \oplus ((b,d,u) \oplus (b',d',u'))$. If we restrict consensus to non-dogmatic beliefs, it turns out that very elegant axiomatisations exists. As proven in [Dan03], the non-dogmatic opinions under consensus form an ordered commutative monoid, with identity element $(0, 0, 1)$ being the only idempotent element. The basic elements will be the *experiments* **1** and **0**, rather than the dogmatic β and δ. The reason that fusion in Subjective Logic can be characterised elegantly, is its link to beta distributions, and thus pairwise addition.

We formally defined belief triples in $\mathcal{BDU}_i$ and $\mathcal{BDU}_\gamma$ as a triple of rational numbers (b, d, u), where $b + d + u = 1$. We concluded that the cases where $u = 0$ are problematic, so in this section we restrict ourselves to cases where $u > 0$. We replace β and δ with atoms correspond to triples with $u > 0$, and provide an axiomatisation of the resulting calculus.

Definition 4.2 (Model $\mathcal{EXP}$)**.** The model $\mathcal{EXP}$ consists of the fragment of Subjective Logic that contains only non-dogmatic beliefs, and consensus, discounting and complement. Non-dogmatic beliefs are triples (b, d, u), where $b + d + u = 1$, $0 \leq b$, $0 \leq d$ and $0 < u$.

Recall, from Definition 3.1, that **0** is $(0, 1/2, 1/2)$, and **1** is $(1/2, 0, 1/2)$ in the model, with the default mapping π.

We use fusion, dilution and inversion to model consensus, discounting and complement. Furthermore, variables stand for non-dogmatic opinions which model non-dogmatic beliefs.

We define the signature Σ_{EXP} as:

Definition 4.3 (Signature Σ_{EXP})**.**

$$\varphi ::= \upsilon \mid \mathbf{0} \mid \mathbf{1} \mid \varphi + \varphi \mid \varphi \cdot \varphi \mid \overline{\varphi}$$

Belief triples in $\mathcal{EXP}$ are exactly the closed terms over Σ_{EXP}.

Lemma 4.4. *Every non-dogmatic Subjective Logic belief triple (belief triple in $\mathcal{EXP}$) corresponds to a term in Σ_{EXP}.*

Proof. Every triple of non-negative rational numbers (that includes every belief triple) can be represented as a triple of non-negative integers, divided by a normalisation factor, which is a positive integer: $(\frac{b_n}{\mathrm{NF}}, \frac{d_n}{\mathrm{NF}}, 1 - \frac{b_n}{\mathrm{NF}} - \frac{d_n}{\mathrm{NF}})$. We consider

$(\mathbf{B1_{0,1}})\ x + y = y + x$	$(\mathbf{B5_{0,1}})\ x + \upsilon = x$	
$(\mathbf{B2_{0,1}})\ x + (y + z) = (x + y) + z$	$(\mathbf{B6_{0,1}})\ \upsilon \cdot x = \upsilon$	
$(\mathbf{B3_{0,1}})\ x \cdot (y \cdot z) = (x \cdot y) \cdot z$		
$(\mathbf{B4_{0,1}})\ x \cdot (y \cdot z) = y \cdot (x \cdot z)$		$(\mathbf{B10_{0,1}})\ \mathbf{0} \cdot x = \upsilon$
$(\mathbf{I1_{0,1}})\ \overline{\overline{x}} = x$	$(\mathbf{I3_{0,1}})\ \overline{\upsilon} = \upsilon$	$(\mathbf{I5_{0,1}})\quad \overline{x \cdot y} = x \cdot \overline{y}$
$(\mathbf{I2_{0,1}})\ \overline{x + y} = \overline{x} + \overline{y}$	$(\mathbf{I4_{0,1}})\ \overline{\mathbf{0}} = \mathbf{1}$	

Figure 4.2: Axioms of Fusion, Dilution and Inversion (**EXP**)

only non-dogmatic belief triples, hence $1 - \frac{b_n}{\text{NF}} - \frac{d_n}{\text{NF}} > 0$, and as b_n, d_n, NF are non-negative integers, $\text{NF} \geq b_n + d_n + 1$. Hence, it suffices to provide at least one term for every $(\frac{b_n}{\text{NF}}, \frac{d_n}{\text{NF}}, 1 - \frac{b_n}{\text{NF}} - \frac{d_n}{\text{NF}})$:

$$\begin{cases} \sum_{b_n} \mathbf{1} + \sum_{d_n} \mathbf{0} & \text{if } \text{NF} = b_n + d_n + 1 \\ (\sum_{b_n + d_n + 1} \mathbf{1} + \sum_{\text{NF} - b_n - d_n - 2} \mathbf{0}) \cdot (\sum_{b_n} \mathbf{1} + \sum_{d_n} \mathbf{0}) & \text{if } \text{NF} \geq b_n + d_n + 2 \end{cases}$$

It can readily be observed that for $n \geq 0, m \geq 0$:

$$\sum_{n} \mathbf{1} + \sum_{m} \mathbf{0} = (\frac{n}{n+m+1}, \frac{m}{n+m+1}, \frac{1}{n+m+1})$$

and hence, in the case that $\text{NF} = b_m + d_n + 1$ we trivially get $(\frac{b_n}{\text{NF}}, \frac{d_n}{\text{NF}}, 1 - \frac{b_n}{\text{NF}} - \frac{d_n}{\text{NF}})$. In the case that $\text{NF} \geq b_m + d_n + 2$, consider the following equation:

$$\begin{aligned} &(\sum_{b_n + d_n + 1} \mathbf{1} + \sum_{\text{NF} - b_n - d_n - 2} \mathbf{0}) \cdot (\sum_{b_n} \mathbf{1} + \sum_{d_n} \mathbf{0}) \\ =&(\frac{b_n + d_n + 1}{\text{NF}}, \frac{\text{NF} - b_n - d_n - 2}{\text{NF}}, \frac{1}{\text{NF}}) \cdot (\frac{b_n}{b_n + d_n + 1}, \frac{d_n}{b_n + d_n + 1}, \frac{1}{b_n + d_n + 1}) \\ =&(\frac{b_n + d_n + 1}{\text{NF}} \frac{b_n}{b_n + d_n + 1}, \frac{b_n + d_n + 1}{\text{NF}} \frac{d_n}{b_n + d_n + 1}, \frac{1}{\text{NF}} + \frac{\text{NF} - b_n - d_n - 2}{\text{NF}} + \frac{1}{\text{NF}}) \\ =&(\frac{b_n}{\text{NF}}, \frac{d_n}{\text{NF}}, 1 - \frac{b_n}{\text{NF}} - \frac{d_n}{\text{NF}}) \end{aligned}$$

□

In **EXP**, $x \cdot \upsilon = \upsilon$ holds, similar to Proposition 4.1.

Proposition 4.5. $\mathbf{EXP} \vdash x \cdot \upsilon = \upsilon$

Proof. $\mathbf{EXP} \vdash x \cdot \upsilon = x \cdot (\mathbf{0} \cdot \mathbf{0}) = \mathbf{0} \cdot (x \cdot \mathbf{0}) = \upsilon$ □

Soundness

We refer to the axioms of $\mathcal{EXP}$ as **EXP**, which can be found in Figure 4.2. All axioms in **EXP** are axioms from **BDU**, or minor variations thereof. Therefore, the soundness of these axioms nearly follows from the soundness of **BDU**.

Lemma 4.6. *If* $\mathbf{EXP} \vdash x = y$ *then* $\mathcal{EXP} \models x = y$.

Proof. Soundness of axiom $(\mathbf{B10_{0,1}})$ follows from $(0, 1/2, 1/2) \otimes (b, d, u) = (0, 0, 1)$ and of axiom $(\mathbf{I4_{0,1}})$ from $\overline{(0, 1/2, 1/2)} = (1/2, 0, 1/2)$. The proof of the other axioms is similar to the proof of Theorem 4.2. □

Completeness

As mentioned in Section 3.2, there are several choices for $\mathbf{0}$ and $\mathbf{1}$, that yield different models parameterised by c. We parameterise $\mathcal{EXP}$ for different choices of $c > 0$, and get $\mathcal{EXP}(c)$. The axioms in **EXP** are sound with respect to all $\mathcal{EXP}(c)$, which can be seen by a simple adaptation of Lemma 4.6. The only axioms that deal with the basic experiments $\mathbf{0}$ and $\mathbf{1}$ are $(\mathbf{B10_{0,1}})$ and $(\mathbf{I4_{0,1}})$. In all models $\mathcal{EXP}(c)$, the trust component of $\mathbf{0}$ is 0, thus $\mathbf{0} \cdot x$ always equals υ. The choice of c also does not influence $\overline{\mathbf{0}} = \mathbf{1}$, as the trust component of $\mathbf{1}$ has the same factor as the distrust component of $\mathbf{0}$. Hence, the soundness of **EXP** with respect to $\mathcal{EXP}(c)$ does not depend on c.

Since we have showed that $\mathcal{EXP}(c)$ and $\mathcal{EXP}(d)$ are not isomorphisms for $c \neq d$, there must be true statements in $\mathcal{EXP}$, that are not true in $\mathcal{EXP}(c)$ for $c \neq 1$. Furthermore, **EXP** contains no truths that are falsified by picking different c. Hence, there is a class of truths in $\mathcal{EXP}$ that cannot be derived in **EXP**. As will be illustrated by Examples 4.1 and 4.2, distributive statements are an important example of truths that cannot be proven in **EXP**.

Additional axioms

As **EXP** is not complete, we can look for additional (temporary) axioms. These axioms range from natural to (at least seemingly) artificial.

Example 4.1. One can easily verify that $\mathcal{EXP} \models (\mathbf{0} + \mathbf{0}) \cdot x = \mathbf{0} \cdot x$, while this is not derivable from **EXP**. One might be inclined to simply add this equality as an axiom $(\mathbf{T1a_{0,1}})$ $(\mathbf{0} + \mathbf{0}) \cdot x = \mathbf{0} \cdot x$. However, there will still be valid equalities that are not derivable, such as $\mathcal{EXP} \models (\mathbf{0} + \mathbf{0} + \mathbf{0}) \cdot x = \mathbf{0} \cdot x$. Therefore, one may want to formulate a more general axiom scheme:

$$(\mathbf{T1_{0,1}}): \qquad (\sum_{n} \mathbf{0}) \cdot x = \upsilon$$

Note that $(\mathbf{T1_{0,1}})$ is closely related to $(\mathbf{B10_{0,1}})$ $(\mathbf{0} \cdot x = \upsilon)$ in two ways. First, $(\mathbf{B10_{0,1}})$ is an instance of $(\mathbf{T1_{0,1}})$. Second, if we accept $(\mathbf{B10_{0,1}})$, then we should accept $(\mathbf{T1_{0,1}})$ as well, because a greater number of failed experiments (without more successes) should not make us less skeptic to *recommendations*. If we have one failure and no successes, we already are completely skeptic, as expressed by $(\mathbf{B10_{0,1}})$. Additionally, since the identity element of fusion in $\mathcal{EXP}$ is υ, we can argue that the empty summation equals υ, and then $\upsilon \cdot x = \upsilon$ $(\mathbf{B6_{0,1}})$ is also an instance of $\mathbf{T1_{0,1}}$.

Axiom scheme $(\mathbf{T1_{0,1}})$ from Example 4.1 is valid for all choices of c, hence the extension of **EXP** with this axiom scheme cannot be complete yet. Besides $(\mathbf{T1_{0,1}})$, we will explore some candidate axioms that do depend on the choice of c in the following example.

Example 4.2. Whereas one expects $(\mathbf{T1_{0,1}})$ to hold before making any calculations in the model, this example will show possible axioms and axiom schemes that are not immediate. We will not prove the validity of the proposed axioms in the models, as they can straightforwardly be verified. Moreover they

are not the central topic of the example. Consider the following truth ($\mathbf{T2a_{0,1}}$): $\mathcal{EXP} \models (\mathbf{1} + (\mathbf{1} + \mathbf{0})) \cdot x = \mathbf{1} \cdot x$. This, contrary to ($\mathbf{T1_{0,1}}$), is only true for $c = 1$. Recall (Definition 3.1) that if $c = 1$ then $\mathbf{1} = (1/2, 0, 1/2)$ and $\mathbf{0}$ is its inverse. This property can be generalised to the following scheme:

$$(\mathbf{T2_{0,1}}): \qquad (\sum_{(n+1)\cdot k} \mathbf{1} + \sum_{n} \mathbf{0}) \cdot x = \sum_{k} \mathbf{1} \cdot x$$

Contrary to most (if not all) aforementioned axioms, we lack straightforward intuition for this axiom scheme, and postulate it only because it is sound with respect to $\mathcal{EXP}$. Without any formal analysis, we can already see that even this scheme, although quite expressive, is not yet powerful enough to prove all equalities, as the right-hand side of the dilution remains invariant under all axioms of **EXP** with one dilution operator, and ($\mathbf{T1_{0,1}}$). We can, however, easily verify that $(\mathbf{1} + \mathbf{1}) \cdot \mathbf{1} = \mathbf{1} \cdot (\mathbf{1} + \mathbf{1})$ ($\mathbf{T3a_{0,1}}$) holds, which is an equality where the right-hand sides of the dilutions differ. This gives rise to yet another axiom scheme to prove this equality:

$$(\mathbf{T3_{0,1}}): \qquad (\sum_{n} \mathbf{1} \cdot \sum_{m} \mathbf{1}) = (\sum_{m} \mathbf{1} \cdot \sum_{n} \mathbf{1})$$

which is a generalisation of $\mathbf{T3a_{0,1}}$. We cannot expect $\mathbf{T3_{0,1}}$ to complete the axiomatisation, as it can only be applied to subformulas containing only sums of positive experiments.

The naive approach to complete the axiomatisation of $\mathcal{EXP}$ - finding unprovable statements that are true in the model, and adding a class of similar true statements - does not seem to work. In order to elegantly and completely axiomatise $\mathcal{EXP}$, it suffices to introduce one auxiliary operator.

4.3 Averaging of Tuples

In the previous section, we presented **EXP**, an incomplete axiomatisation of $\mathcal{EXP}$. We showed that distributive laws are lacking in **EXP**, and gave examples of axiom schemes, ($\mathbf{T1_{0,1}}$), ($\mathbf{T2_{0,1}}$) and ($\mathbf{T3_{0,1}}$), that could be added to **EXP**. As we will see later, they are instances of a more general rule; a rule that contains an auxiliary operator, namely the opinion mean, denoted by $\underline{\oplus}$. In this section we will study this operator as the basis for a complete axiomatisation.

Remark 4.1. To reduce notational complexity we introduce a short hand notation for a comma separated list of identical objects. We shorthand $\underbrace{x, \ldots, x}_{n}$ to $\big(x\big)_n$.

Definition 4.4 (Opinion mean)**.** *The opinion mean* is an unranked function on $\mathcal{EXP}$-beliefs, defined as:

$$\underline{\oplus}((b_1, d_1, u_1), \ldots, (b_n, d_n, u_n)) = (\frac{\sum_{1 \leq i \leq n} b_i}{n}, \frac{\sum_{1 \leq i \leq n} d_i}{n}, \frac{\sum_{1 \leq i \leq n} u_i}{n})$$

We extend the signature Σ_{EXP} to $\Sigma_{EXP+AVs^3}$, and for $1 \leq n$ its syntax is defined by the following scheme:

$$\varphi ::= \upsilon \mid \mathbf{0} \mid \mathbf{1} \mid \varphi + \varphi \mid \varphi \cdot \varphi \mid \pm(\varphi, \ldots, \varphi) \mid \overline{\varphi}$$

Tuple mean

We generalise the opinion mean to *tuple mean* as to study its properties independently of the context of Subjective Logic. We formulate the theory of the opinion mean ($\mathcal{AV}^3$) and tuple means ($\mathcal{AV}^\kappa$), using opinion averaging and tuple averaging, respectively. The complete axiomatisation of tuple averaging is an independent result, due to its close ties to quasi-arithmetic means. We develop a complete axiomatisation of tuple averaging, and set it up in such a way that the results stand independently. Next, in Section 4.4, we use opinion averaging to completely axiomatise $\mathcal{EXP}{+}\mathcal{AVs}^3$. Both sections provide two axiomatisations: first a complete axiom scheme, followed by a finite complete axiomatisation. The axioms in the axiom scheme provide a better intuitive handle than the finite counterpart, and for that reason are likely relevant to the reader.

Tuple averaging assumes a set A^κ of $\kappa \geq 0$ elements, say $a_1, \ldots, a_\kappa$, where $a_i = (\big(0\big)_{i-1}, 1, \big(0\big)_{\kappa-i})$. We refer to such a_i as the atoms in A^κ. Further, we consider κ-tuples of rational numbers $(\chi^1, \ldots, \chi^\kappa)$, with the restriction that $0 \leq \chi^i \leq 1$ and $\sum_{1\leq i\leq\kappa} \chi^i = 1$. One can consider A^κ as a basis of unit vectors spanning up the set of κ-tuples of rational numbers through the application of the tuple mean.

Definition 4.5 (Tuple mean)**.** The *tuple mean* is an unranked function on κ-tuples of rational numbers defined by:

$$\underline{\oplus}((\chi_1^1, \ldots, \chi_1^\kappa), \ldots, (\chi_n^1, \ldots, \chi_n^\kappa)) = (\frac{\sum_{1\leq i\leq n} \chi_i^1}{n}, \ldots, \frac{\sum_{1\leq i\leq n} \chi_i^\kappa}{n})$$

We define the model $\mathcal{AV}^\kappa$ as the set of κ-tuples of rational numbers under the tuple mean. Let $a \in A^\kappa$, then we define the signature Σ_{AVs^κ} terms as:

$$\varphi ::= a | \pm(\varphi, \ldots, \varphi)$$

Note that we use variables from the roman alphabet to reason over tuples, e.g. $x_i = (\chi_i^1, \ldots, \chi_i^\kappa)$, and Greek variables to reason within tuples.

Proposition 4.7. *Every κ-tuple of rational numbers, is expressible as a tuple mean of atoms in A^κ.*

Proof. Let $x = (\chi^1, \ldots, \chi^\kappa)$ be such a tuple. Rewrite the elements of the tuple to a form where they have equal denominators, so every χ^i in the tuple equals $\frac{\mu^i}{\nu}$. Then, we construct a tuple mean where each $a_i \in A^\kappa$ will appear μ^i times. Recall that $\sum_i \chi^i = 1$, therefore $\sum_i \mu^i = \nu$. By Definition 4.5, the i-th component equals $\frac{\sum_{1\leq k\leq\mu^i} 1}{\nu} = \frac{\mu^i}{\nu} = \chi^i$. □

There exists a unique normal form for terms in Σ_{AVs^κ}:

Corollary 4.8. *There exists a unique normal form for terms in Σ_{AVs^κ}, evaluated in $\mathcal{AV}^\kappa$. For atoms $a_1, \ldots, a_n \in A^\kappa$ and $\gcd k_1, \ldots, k_n = 1$, $\pm(\big(a_1\big)_{k_1}, \ldots, \big(a_n\big)_{k_n})$ is a unique normal form.*

Proof. Every tuple $(\frac{n_1}{m_1}, \ldots, \frac{n_\kappa}{m_\kappa})$ has a unique representation, namely one where every $m_i = m$, and the greatest common divider of all n_i is 1. If we apply the

$(\mathbf{A1}_\infty)\ \pm(x_1,\ldots,x_n) = \pm(x_{\pi(1)},\ldots,x_{\pi(n)})$	for any permutation π
$(\mathbf{A5}_\infty)\ \pm(x_1,\ldots,x_n) = \pm(\big(x_1\big)_k,\ \ldots,\ \big(x_n\big)_k)$	
$(\mathbf{A6}_\infty)\ \pm(\big(\pm(x_1,\ldots,x_n)\big)_n, y_1,\ldots,y_m) = \pm(x_1,\ldots,x_n,\ y_1,\ \ldots,\ y_m)$	
$(\mathbf{A7}_\infty)\ \pm(x) = x$	for any $x \in A$

Figure 4.3: Axiom scheme for tuple average ($\mathbf{AV}s^\kappa$)

construction in the proof of Proposition 4.7 to this unique representation of the tuple, we obtain a tuple mean which forms the required unique normal up to symmetry. □

We refer to these expressions as being in the unique normal form.

Definition 4.6 (The unique normal form)**.** For atoms $a_1,\ldots,a_n \in A^\kappa$, we call

$$\pm(\big(a_1\big)_{k_1},\ldots,\big(a_n\big)_{k_n})$$

(where gcd $\overrightarrow{k} = 1$) *the unique normal form.*

Our notion of tuple means is closely related to that of quasi-arithmetic means [Kol30]. Quasi-arithmetic means can be characterised as an infinite collection of continuous, strictly increasing, *symmetric* real functions:

$$M_1(x_1), M_2(x_1,x_2),\ldots,M_n(x_1,\ldots,x_n),\ldots,$$

such that

$$(\textit{reflexivity})\ M_n(x,\ldots,x) = x$$

and

$$(\textit{compositionality})\ \text{if } M_k(x_1,\ldots,x_k) = x \text{ then}$$
$$M_n(x_1,\ldots,x_k,x_{k+1},\ldots,x_n) = M_n(\big(x\big)_k, x_{k+1},\ldots,x_n).$$

The arithmetic mean is the simplest example of a quasi-arithmetic mean. Other examples are the geometric mean, the harmonic mean and the power mean. Contrary to quasi-arithmetic means, the tuple means operator does not range over real values, but the *symmetry*, *reflexivity* and *compositionality* properties of quasi-arithmetic means are also properties of tuple means. Due to Proposition 4.7, we can completely, finitely axiomatise tuple means, as tuple averages. Most notions of means do not have a complete finite axiomatisation.

Complete axiom scheme

We present an axiom scheme ($\mathbf{AV}s^\kappa$) for tuple averaging ($\mathcal{AV}^\kappa$) in Figure 4.3. The axiomatisation assumes the signature Σ_{AVs^κ}.

We see that $\mathbf{A1}_\infty$ describes *symmetry*, that $\mathbf{A5}_\infty$ together with $\mathbf{A7}_\infty$ is sufficient to prove *reflexivity*, and that $\mathbf{A6}_\infty$ is a reformulation of *compositionality*. To determine whether $\mathcal{AV}^\kappa$ is (strictly) increasing is non-trivial. We can define a total order (the lexicographical order of the tuple), that is strictly increasing. Instead

of one ordering, we can define κ orderings, one for each element in the tuples. It can easily been seen that each ordering is monotonic with respect to tuple averaging. As a consequence, for each element, the mean lies between the maximum and the minimum. More formally, let $(\chi^1, \ldots, \chi^\kappa) = \underline{\oplus}((\chi^1_1, \ldots, \chi^\kappa_1), \ldots, (\chi^1_n, \ldots, \chi^\kappa_n))$, then for all i, $\min_j(\chi^i_j) \leq \chi^i \leq \max_j(\chi^i_j)$. Furthermore, slightly changing one variable in the mean will only slightly change the mean. Which is the intuition behind continuous functions. The problem, however, with formally defining continuity is that some elements in the mean go up, while others go down. We can interpret the definition of continuity in such a way that tuple means are continuous. In that case, the only difference is that $\mathcal{AV}^\kappa$ ranges over tuples of rational numbers, and not over single real numbers. It seems that $\mathcal{AV}^\kappa$, therefore adheres to some generalised characterisation of quasi-arithmetic means. The axiom scheme $\mathbf{AV}s^\kappa$ is, however, stronger than *symmetry*, *reflexivity* and *compositionality*, as it implies all three, but $\mathbf{A5}_\infty$ (with $\mathbf{A7}_\infty$) is strictly stronger than *reflexivity*.

We need to prove that $\mathbf{AV}s^\kappa$ is a complete axiomatisation of $\mathcal{AV}^\kappa$. First, we prove soundness:

Lemma 4.9. *If* $\mathbf{AV}s^\kappa \vdash x = y$ *then* $\mathcal{AV}^\kappa \models x = y$.

Proof. It is easy to see that the four axioms are true in the model. □

Then we prove completeness:

Lemma 4.10. *If* $\mathcal{AV}^\kappa \models x = y$ *then* $\mathbf{AV}s^\kappa \vdash x = y$.

Proof. Equality is transitive, therefore, if $\mathcal{AV}^\kappa \models x = y$, then x and y have equal unique normal forms (Corollary 4.8). Without loss of generality, we can therefore assume y to be the unique normal form of x. Now we shall prove, by contradiction, that for all x, such that y is its unique normal form, $\mathbf{AV}s^\kappa \vdash x = y$.

Assume there exists x such that we cannot derive equality with its unique normal form y, then (for a suitable notion of size) there is a smallest x with $\mathbf{AV}s^\kappa \nvdash x = y$. We define the size of a term $x = \pm(x_1, \ldots, x_s)$ based on its nesting depth, and the amount of subterms with maximal nesting depth: Let d be the maximum nesting depth of averaging operators of a term. Let u be the number of (syntactically) unique x_i with depth $d-1$ in a term. We say that (d, u) is the size of x. We say that $x > x'$, when x and x' have size (d, u) and (d', u'), respectively, and either $d > d'$ or $d = d' \wedge u > u'$. We shall use $x \equiv y$ for syntactic equivalence.

Let us distinguish possible smallest terms x, such that $\mathbf{AV}s^\kappa \nvdash x = y$:

- If $x \in A^\kappa$, then $\mathbf{AV}s^\kappa \vdash \pm(x) = x$, and $\pm(x)$ is the unique normal form (Definition 4.6).
- If $x \equiv \pm(\left(a_1\right)_{k_1}, \ldots, \left(a_n\right)_{k_n})$, $\gcd \overrightarrow{k} = 1$, then x is the unique normal form.
- If $x \equiv \pm(\left(a_1\right)_{k_1}, \ldots, \left(a_n\right)_{k_n})$, $\gcd \overrightarrow{k} > g$, then we can apply $\mathbf{A5}_\infty$ with parameter g, to get $\mathbf{AV}s^\kappa \vdash x = \pm(\left(a_1\right)_{\frac{k_1}{g}}, \ldots, \left(a_n\right)_{\frac{k_n}{g}})$, which is the unique normal form.

- If $x \equiv \pm(x_1, \ldots, x_m)$ where all $x_i \in A^\kappa$, then, using $\mathbf{A1}_\infty$, it is reduced to one of the above two cases.

- If $x \equiv \pm(x_1, \ldots, x_n)$ where at least one $x_i \notin A^\kappa$, then there is an x_i with depth $d-1$. The subterm $x_i \equiv (z_1, \ldots, z_m)$ can appear several times, say k times. Then, by $\mathbf{A1}_\infty$, there are $y_1, \ldots, y_{n-k}$, such that $\mathbf{AV}s^\kappa \vdash x = \pm(\big((z_1, \ldots, z_m)\big)_k, y_1, \ldots, y_{n-k})$. By $\mathbf{A5}_\infty$, $\mathbf{AV}s^\kappa \vdash x = \pm(\big((z_1, \ldots, z_m)\big)_{k \cdot m}, \big(y_1\big)_m, \ldots, \big(y_{n-k}\big)_m)$, and by $\mathbf{A6}_\infty$, $\mathbf{AV}s^\kappa \vdash x = \pm(\big(z_1\big)_k, \ldots \big(z_m\big)_k, \big(y_1\big)_m, \ldots, \big(y_{n-k}\big)_m) \equiv x'$.

  We distinguish two cases: Either, x_i is the only subterm of x with depth $d-1$ (i.e. $u = 1$) or there is a nonempty set X (of size u) of subterms with depth $d-1$. In the first case, all y_j have depth at most $d-2$. Since all z_j are subterms of x_i, all z_j also have depth at most $d-2$. Therefore, x' has a depth of at most $d-1$. In the latter case, since all z_j have depth at most $d-2$, and x' contains the set $X \setminus \{x_i\}$ of subterms with depth $d-1$. Hence, x' contains $u-1$ unique subterms with depth $d-1$. In both cases x' is smaller than x, hence x is not the smallest counterexample.

We conclude that there is no smallest term x, such that $\mathcal{AV}^\kappa \models x = y$, but $\mathbf{AV}s^\kappa \nvdash x = y$. □

Complete finite axiomatisation

The structure of the axiom scheme is quite simple. It is, therefore, no surprise that a finite set of axioms is sufficient to completely axiomatise $\mathcal{AV}^\kappa$. A problem we need to solve first, is the fact that there is an infinite number of tuple averaging functions, one for each arity. We therefore define an average with only one parameter, and allow this parameter to contain a collection of tuples, in Figure 4.4

Definition 4.7 (Signature Σ_{AV^κ})**.** We define a signature Σ_{AV^κ}, for $a \in A^\kappa$:

$$\varphi ::= a | \pm(\psi)$$

$$\psi ::= \psi, \psi | \varphi$$

Modulo associativity, a trivial bijection exists between Σ_{AVs^κ} and Σ_{AV^κ} terms. If we ignore the parentheses of pairing, as is common, then a string in Σ_{AVs^κ} is mapped to the same string in Σ_{AV^κ}. For simplicity, we make no distinction between the two signatures.

We define another operation, similar to counting, denoted by $\#(x)$. All terms that are not pairs have the same count (1). For any n, all nested pairings $x_1, \ldots, x_n$ also have the same count (n). To formulate axioms using averages, with pairing and counting, we need a new signature. All terms in Σ_{AV^κ} are also terms in the new signature:

Definition 4.8 (Signature $\Sigma^\#_{AV^\kappa}$)**.** We define the signature $\Sigma^\#_{AV^\kappa}$, for $a \in A^\kappa$:

$$\varphi ::= a | \varphi, \varphi | \pm(\varphi) | \#(\varphi)$$

(**A1a**) $x, y = y, x$	(**A1b**) $(x, y), z = x, (y, z)$
(**A2**) $\#(\pm(x)) = \#(\pm(y))$	(**A3**) $\#(x, y) = \#(x), \#(y)$
(**A4**) $\pm(\pm(x)) = \pm(x)$	(**A5**) $\pm(x) = \pm(y) \Rightarrow \pm(x, y) = \pm(x)$
(**A6**) $\#(x) = \#(y) \wedge \pm(x) = \pm(y) \Rightarrow x = y$	(**A7**) $\pm(x) = x$ for any $x \in A$

Figure 4.4: Axioms for tuple average ($\mathbf{AV}^\kappa$)

Using pairing and counting, we define the axioms $\mathbf{AV}^\kappa$.

The theory $\mathbf{AV}^\kappa$ is set up in such a way that terms in Σ_{AVs^κ} can only equal other terms in Σ_{AVs^κ}:

Proposition 4.11. *If $\mathbf{AV}^\kappa \vdash x = y$, then x is a term in Σ_{AVs^κ} if, and only if, y is a term in Σ_{AVs^κ}.*

Proof. Assume $\mathbf{AV}^\kappa \vdash x = y$, x in Σ_{AVs^κ}, but y not in Σ_{AVs^κ}. By pigeonhole principle, there are x' in Σ_{AVs^κ} and y' not in Σ_{AVs^κ}, such that $x' = y'$ follows from an axiom. It is immediate that that axiom is not **A1a**, **A1b**, **A2**, **A3**, **A4**, or **A7**. Axioms **A5** and $\mathbf{A6}_\infty$ have preconditions, the preconditions must have some terms x'' in Σ_{AVs^κ}, y'' not in Σ_{AVs^κ}, such that $\mathbf{AV}^\kappa \vdash x'' = y''$. □

The counting operator, #, is intended to count the number of elements in a nested pair:

Proposition 4.12. *Let x'_i and y'_j be Σ_{AVs^κ} terms for all $1 \leq i \leq n$, $1 \leq j \leq m$. If $\mathbf{AV}^\kappa \vdash \#(x'_1, \ldots, x'_n) = \#(y'_1, \ldots, y'_m)$ then $n = m$.*

Proof. The converse is easy to prove, if $n = m$, then $\mathbf{AV}^\kappa \vdash \#(x'_1, \ldots, x'_n) = \#(y'_1, \ldots, y'_m)$. This helps us, as it proves that any nested pair of size n of Σ_{AVs^κ} terms has the same count as any other nested pair of size n of Σ_{AVs^κ} terms. Thus the count of nested pairs of Σ_{AVs^κ} terms is invariant under all axioms except **A2** and **A3**. We need only prove that there are no $x'_1, \ldots, x'_n$ and $y'_1, \ldots, y'_m$, with $n \neq m$, such that they are provably equal, using only **A2** and **A3**. We can do structural induction over the counting operator, with respect to the two axioms. So if $n = 1$, then we must apply **A2**, thus $m = 1$. If $\#((s_1, \ldots, s_i), (s'_1, \ldots, s'_{i'})) = \#((t_1, \ldots, t_j), (t'_1, \ldots t'_{j'}))$, then $\#(s_1, \ldots, s_i)$ must equal $\#(t_1, \ldots, t_j)$, and $\#(s'_1, \ldots, s'_{i'})$ must equal $\#(t'_1, \ldots t'_{j'})$, and by hypothesis, $i = j$ and $i' = j'$, thus $i + i' = j + j'$. □

Using Propositions 4.11 and 4.12, we can prove that $\mathbf{AV}^\kappa$ is a conservative extension of $\mathbf{AV}s^\kappa$:

Lemma 4.13. *For all terms x, y in Σ_{AVs^κ}, $\mathbf{AV}s^\kappa \vdash x = y$ iff $\mathbf{AV}^\kappa \vdash x = y$.*

Proof. We first prove the right implication; if $\mathbf{AV}s^\kappa \vdash x = y$ then $\mathbf{AV}^\kappa \vdash x = y$.

It suffices to prove, for every axiom scheme in $\mathbf{AV}s^\kappa$, that they can be derived in $\mathbf{AV}^\kappa$.

Next, we prove the left implication; if $\mathbf{AV}^\kappa \vdash x = y$ then $\mathbf{AV}s^\kappa \vdash x = y$.

Take the derivation tree of $\mathbf{AV}^\kappa \vdash x = y$. The conclusion must be a term in Σ_{AVs^κ}. We distinguish the possible derivation trees by the last derivation rule, and prove that there is a tree $\mathbf{AV}s^\kappa \vdash x = y$ with the same conclusion.

By Proposition 4.11, it can be readily verified that substitution, reflexivity, symmetry and transitivity can be applied with the same premisses and conclusion in both theories. That leaves axiom, conditional axiom and context rule (see Definition 3.4), which we address below. The crucial part of the proof is in the context rule, as it is the only part where the conclusion is a term in Σ_{AVs^κ}, without its premisses necessarily being in Σ_{AVs^κ}.

ax The only two axioms that can be used to deduce $\mathbf{AV}^\kappa \vdash x = y$ are **A4** and **A7**.
Whenever **A7** applies, so does $\mathbf{A7}_\infty$.
Whenever we apply **A4** to $x = \pm(x_1, \dots x_n)$, we can apply $\mathbf{A6}_\infty$, with $m = 0$.

cond The only two conditional axioms that can be used to deduce $\mathbf{AV}^\kappa \vdash x = y$ are **A5** and **A6**.
To match **A5** in $\mathbf{AV}s^\kappa$, we need to prove that if $\pm(x_1, \dots, x_n) = \pm(y_1, \dots, y_m)$, then $\pm(x_1, \dots, x_n, y_1, \dots, y_m) = \pm(x_1, \dots, x_n)$:

$$
\begin{aligned}
\pm(x_1, \dots, x_n, y_1, \dots, y_m) &= \{\mathbf{A6}_\infty\} \\
\pm(x_1, \dots, x_n, \big(\pm(y_1, \dots, y_m)\big)_m) &= \{\text{Apply condition}\} \\
\pm(x_1, \dots, x_n, \big(\pm(x_1, \dots, x_n)\big)_m) &= \{\mathbf{A6}_\infty\} \\
\pm(\big(\pm(x_1, \dots, x_n)\big)_{n+m}) &= \{\mathbf{A5}_\infty\} \\
\pm(x_1, \dots, x_n)
\end{aligned}
$$

Since x and y are terms in Σ_{AVs^κ}, whenever axiom **A6** is applied to derive $x = y$, $\mathbf{AV}s^\kappa \vdash x = \pm(x) = \pm(y) = y$ (either using $\mathbf{A7}_\infty$ or $\mathbf{A6}_\infty$ with $n = 1$ and $m = 0$).

cont The only function in Σ_{AVs^κ} is averaging, $\pm$, so the context rule must apply to an application of averaging. Therefore, for some $z_1, \dots z_k$, $x = \pm(z_1, \dots, z_{i-1}, x', \dots, x'_n, z_i, \dots, z_k)$ and $y{=}\pm(z_1, \dots, z_{i-1}, y', \dots, y'_m, z_i, \dots, z_k)$, where $\mathbf{AV}s^\kappa \vdash x' = y'$. If x' and y' are terms in Σ_{AVs^κ}, the context rule can readily be applied in $\mathbf{AV}s^\kappa$, and by Proposition 4.11, we need only consider the case where neither x' and y' are terms in Σ_{AVs^κ}.

By Definition 4.8, we know that the only terms in $\Sigma^{\#}_{AV^\kappa}$ not in Σ_{AVs^κ} have counting (#) or pairing as a main operator. However, neither x' nor y' contain a counting operator, since x and y do not contain one. Therefore, we conclude that $x' \equiv x'_1, \dots, x'_n$ and $y' \equiv y'_1, \dots, y'_m$, where all x'_i and y'_i are terms in Σ_{AVs^κ}. There are only three axioms where the conclusion may be a pair. Two axioms (**A1a** and **A1b**) deal with the fact that pairing is commutative and associative, which is covered by $\mathbf{A1}_\infty$ in this context. The remaining axiom, $\mathbf{A6}_\infty$, can only be applied under two conditions. First, that $\mathbf{AV}^\kappa \vdash \#(x'_1, \dots, x'_n) = \#(y'_1, \dots, y'_m)$ which, by Proposition A.1, implies that $n = m$. Second, that $\mathbf{AV}^\kappa \vdash \pm(x'_1, \dots, x'_n) = \pm(y'_1, \dots, y'_m)$. Under

these two conditions, we can derive in $\mathbf{AV}s^{\kappa}$:

$$\begin{aligned}
\pm(z_1,\ldots,z_{i-1},x'_1,\ldots,x'_n,z_i,\ldots,z_k) &= \{\mathbf{A6}_\infty\}\\
\pm(z_1,\ldots,z_{i-1},\big(\pm(x'_1,\ldots,x'_n)\big)_n,z_i,\ldots,z_k) &= \{\text{Applyconditions}\}\\
\pm(z_1,\ldots,z_{i-1},\big(\pm(y'_1,\ldots,y'_m)\big)_m,z_i,\ldots,z_k) &= \{\mathbf{A6}_\infty\}\\
\pm(z_1,\ldots,z_{i-1},y'_1,\ldots,y'_m,z_i,\ldots,z_k)
\end{aligned}$$

Therefore, we conclude that for any derivation tree in $\mathbf{AV}^{\kappa}$, a tree with matching conclusions exists in $\mathbf{AV}s^{\kappa}$. □

As an immediate consequence of Lemma 4.13, we see that $\mathbf{AV}^{\kappa}$ sound and complete with respect to $\mathcal{AV}^{\kappa}$. If we take $\kappa = 3$, then $\mathcal{AV}^{\kappa}$ is the model of opinion averaging:

Corollary 4.14. $\mathbf{AV^3}$ *is a sound and complete axiomatisation of opinion averaging.*

Proof. Opinion means (Definition 4.4) are an instance of tuple means (Definition 4.5), hence Lemma 4.13 also applies to opinion means. □

4.4 Dilution, Fusion and Averaging of Opinions from Experiments

In this section, we use the two axiomatisations of opinion averaging ($\mathbf{AV}s^{\mathbf{3}}$ and $\mathbf{AV^3}$) to construct a complete axiomatisation of $\mathcal{EXP}{+}\mathcal{AV}s^3$; the model consisting of $\mathcal{EXP}$ and $\mathcal{AV}^3$. In particular, we provide an axiomatic relationship between the operators in $\mathcal{EXP}{+}\mathcal{AV}s^3$ and opinion averaging. Similar to Section 4.3, we provide these axioms in two flavours. First, we provide an axiom scheme, that formalises the relationship in a natural way. Second, we provide a finite axiomatisation, based on the former. Finally, we analyse the axiomatisations, their variations and their models. Moreover, we study the implications of the axioms regarding dilution, and propose alternative axioms, that capture more reasonable models.

Furthermore, we define our constants $\mathbf{0}$ and $\mathbf{1}$, as averages of *belief triples* in A^3. Let $\{\beta,\delta,\upsilon\} = A^3$, then we present a complete axiom scheme ($\mathbf{FDN}s + \mathbf{AV}s^3$) of $\mathcal{EXP}{+}\mathcal{AV}s^3$ in Figure 4.5.

The axiom scheme $\mathbf{FDN}s + \mathbf{AV}s^3$ is sound with respect to $\mathcal{EXP}{+}\mathcal{AV}s^3$:

Lemma 4.15. *If* $\mathbf{FDN}s + \mathbf{AV}s^3 \vdash x = y$ *then* $\mathcal{EXP}{+}\mathcal{AV}s^3 \models x = y$.

Proof. Verifying that the five axioms are true in the model is straightforward. □

And we prove completeness:

Lemma 4.16. *If* $\mathcal{EXP}{+}\mathcal{AV}s^3 \models x = y$ *then* $\mathbf{FDN}s + \mathbf{AV}s^3 \vdash x = y$.

Proof. We apply structural induction over the shape of x in $\Sigma_{EXP+AVs^3}$, to show that $\mathbf{FDN}s + \mathbf{AV}s^3 \vdash x = \pm(\beta,\ldots,\beta,\delta,\ldots,\delta,\upsilon,\ldots,\upsilon)$. If:

$$(\mathbf{D1}_\infty)\quad \mathbf{0} = \pm(\delta, \upsilon)$$
$$(\mathbf{D2}_\infty)\quad \mathbf{1} = \pm(\beta, \upsilon)$$
$$(\mathbf{D7}_\infty)\quad \pm(\left(\beta\right)_l,\ \left(\delta\right)_m,\ \left(\upsilon\right)_n) + \pm(\left(\beta\right)_{l'},\ \left(\delta\right)_{m'},\ \left(\upsilon\right)_{n'}) = \pm(\left(\beta\right)_{l\cdot n'+l'\cdot n},\ \left(\delta\right)_{m\cdot n'+m'\cdot n},\ \left(\upsilon\right)_{n\cdot n'})$$
$$(\mathbf{D11}_\infty)\quad \pm(\left(\beta\right)_l,\ \left(\delta\right)_m,\ \left(\upsilon\right)_n) \cdot \pm(\left(\beta\right)_{l'},\ \left(\delta\right)_{m'},\ \left(\upsilon\right)_{n'}) = \pm(\left(\beta\right)_{l\cdot l'},\ \left(\delta\right)_{l\cdot m'},\ \left(\upsilon\right)_{l\cdot n'+(m+n)\cdot(l'+m'+n')})$$
$$(\mathbf{D17}_\infty)\quad \overline{\pm(\left(\beta\right)_l,\ \left(\delta\right)_m,\ \left(\upsilon\right)_n)} = \pm(\left(\beta\right)_m,\ \left(\delta\right)_l,\ \left(\upsilon\right)_n)$$

Figure 4.5: Axiom scheme of Fusion, Dilution and Negation in terms of Averaging. Forms $\mathbf{FDN}s + \mathbf{AV}s^3$ together with $\mathbf{AV}s^{\mathbf{3}}$

$[x = \upsilon]$ then, by **A7**, $\mathbf{FDN}s + \mathbf{AV}s^3 \vdash \upsilon = \pm(\upsilon)$.

$[x = \mathbf{0}]$ then, by $\mathbf{D1}_\infty$, $\mathbf{FDN}s + \mathbf{AV}s^3 \vdash \mathbf{0} = \pm(\delta, \upsilon)$.

$[x = \mathbf{1}]$ then, by $\mathbf{D2}_\infty$, $\mathbf{FDN}s + \mathbf{AV}s^3 \vdash \mathbf{1} = \pm(\beta, \upsilon)$.

$[x = \pm(x_1, \ldots, x_n)]$ then, by the induction hypothesis, all x_i are equal to Σ_{AVs^κ} terms. Therefore x is a Σ_{AVs^κ} term. By Proposition 4.7, there exists some $z \equiv \pm(\beta, \ldots, \beta, \delta, \ldots, \delta, \upsilon, \ldots, \upsilon)$ with $\mathcal{AV}^\kappa \models x = z$, and by Lemma 4.13, $\mathbf{AV}s^\kappa \vdash x = z$ and $\mathbf{FDN}s + \mathbf{AV}s^3 \vdash x = z$.

$[x = x' + x'']$ then, by the induction hypothesis, $x' \equiv \pm(\left(\beta\right)_l,\ \left(\delta\right)_m,\ \left(\upsilon\right)_n)$ and $x'' \equiv \pm(\left(\beta\right)_{l'},\ \left(\delta\right)_{m'},\ \left(\upsilon\right)_{n'})$. Apply $\mathbf{D7}_\infty$.

$[x = x' \cdot x'']$ then, by the induction hypothesis, $x' \equiv \pm(\left(\beta\right)_l,\ \left(\delta\right)_m,\ \left(\upsilon\right)_n)$ and $x'' \equiv \pm(\left(\beta\right)_{l'},\ \left(\delta\right)_{m'},\ \left(\upsilon\right)_{n'})$. Apply $\mathbf{D11}_\infty$.

$[x = \overline{x'}]$ then, by the induction hypothesis, $x' \equiv \pm(\left(\beta\right)_l,\ \left(\delta\right)_m,\ \left(\upsilon\right)_n)$. Apply $\mathbf{D17}_\infty$.

Hence, every term can be reduced to its unique normal form (Definition 4.6). □

Complete finite axiomatisation

The complete axiom scheme $\mathbf{FDN}s + \mathbf{AV}s^3$ can be formulated as a finite set of axioms. In order to obtain a complete finite set of axioms for $\mathcal{EXP}+\mathcal{AV}s^3$, we need to rewrite the schemes $\mathbf{D7}_\infty$, $\mathbf{D11}_\infty$ and $\mathbf{D17}_\infty$ as a finite set of axioms, because we proved that $\mathbf{AV^3}$ is a complete finite axiomatisation of opinion averaging.

In order to construct such a finite axiomatisation, we introduce some more auxiliary operators. In particular, we introduce pairing sensitive fusion ($\boxplus$), pairing sensitive dilution ($\boxdot$) and the pairing sensitive inversion ($\sim$).

Definition 4.9 (Signature $\Sigma^{\square}_{EXP+AVs^3}$)**.** We define the signature $\Sigma^{\square}_{EXP+AVs^3}$ as:

$$\varphi ::= \upsilon \mid \mathbf{0} \mid \mathbf{1} \mid \varphi + \varphi \mid \varphi \cdot \varphi \mid \overline{\varphi} \pm (\psi)$$
$$\psi ::= \varphi \mid \psi, \psi \mid \psi \boxplus \psi \mid \psi \boxdot \psi \mid \sim(\psi)$$

(**D1**) $\mathbf{0} = \pm(\delta, \upsilon)$	(**D12**) $\pm(x) \cdot \pm(y) = \pm(x \boxdot y)$
(**D2**) $\mathbf{1} = \pm(\beta, \upsilon)$	(**D13**) $(x, x') \boxdot y = (x \boxdot y), (x' \boxdot y)$
(**D3**) $D(\beta)$	(**D14**) $x \boxdot (y, y') = (x \boxdot y), (x \boxdot y')$
(**D4**) $D(\delta)$	(**D15**) $\beta \boxdot \pm(y) = \pm(y)$
(**D5**) $D(\pm(x)) \Rightarrow D(\beta, x)$	(**D16**) $\delta \boxdot \pm(y) = u$
(**D6**) $D(\pm(x)) \Rightarrow D(\delta, x)$	(**D17**) $\upsilon \boxdot \pm(y) = u$
(**D7**) $\pm(x) + \pm(y) = \pm(x \boxplus y)$	(**D18**) $\overline{\pm(x)} = \pm(\sim(x))$
(**D8**) $(u, x) \boxplus y = y, (x \boxplus y)$	(**D19**) $\sim(x, y) = \sim(x), \sim(y)$
(**D9**) $x \boxplus (u, y) = x, (x \boxplus y)$	(**D20**) $\sim(\beta) = \delta$
(**D10**) $D(x) \wedge D(y) \Rightarrow (u, x) \boxplus y = y$	(**D21**) $\sim(\delta) = \beta$
(**D11**) $D(x) \wedge D(y) \Rightarrow x \boxplus (u, y) = x$	(**D22**) $\sim(\upsilon) = \upsilon$

Figure 4.6: Finite axiomatisation of Fusion, Dilution and Negation in terms of Averaging. Forms $\mathbf{FDN} + \mathbf{AV}^3$ together with $\mathbf{AV^3}$.

As the name suggests, the pairing sensitive operators behave like their original operator, except that they range over pairs, instead of opinions. The pairing sensitive variants implement the calculations under the brace in the axiom scheme $\mathbf{FDN}s + \mathbf{AV}s^3$, so they must treat the pairs correctly. Note that Lemma 4.13 proves that they do so. We furthermore introduce a unary operator ($D(\_)$) that tests whether an opinion is dogmatic, and an identity element of pairing. Both are used to provide a basis for an inductive definition of pairing sensitive fusion.

The following lemma states the relation between the two axiom systems.

Lemma 4.17. *For all terms x, y in $\Sigma_{EXP+AVs^3}$, $\mathbf{FDN}s + \mathbf{AV}s^3 \vdash x = y$ iff $\mathbf{FDN} + \mathbf{AV}^3 \vdash x = y$*

Proof. We first prove the right implication; if $\mathbf{FDN}s + \mathbf{AV}s^3 \vdash x = y$, then $\mathbf{FDN} + \mathbf{AV}^3 \vdash x = y$.

It suffices to show that $\mathbf{D7}_\infty$, $\mathbf{D11}_\infty$ and $\mathbf{D17}_\infty$ are derivable in $\mathbf{FDN} + \mathbf{AV}^3$.

Next, we prove the right implication; if $\mathbf{FDN} + \mathbf{AV}^3 \vdash x = y$, then $\mathbf{FDN}s + \mathbf{AV}s^3 \vdash x = y$.

It suffices to prove for x, y in Σ_{AVs^κ} that, whenever $\mathbf{AV}^\kappa \vdash x = y$ also $\mathbf{AV}s^\kappa \vdash x = y$. We cannot reuse the same approach as above, since there are equalities in $\mathbf{AV}^\kappa$, which are not equalities in $\mathbf{AV}s^\kappa$. Luckily, these equalities do not concern terms in $\Sigma_{EXP+AVs^3}$. Due to transitivity of equality, we assume, without loss of generality, that y is in the unique normal form. We can do structural induction over the term x. Hence it is either atomic, $x' + x''$, $x' \cdot x''$, $\overline{x'}$ or $\pm(x_1, \ldots, x_n)$, for $x', x'', x_1, \ldots, x_n$ in unique normal form. If the main operator x is atomic, then $\mathbf{FDN}s + \mathbf{AV}s^3 \vdash x = \pm(x) \equiv y$ is trivially true. If x is an average, we apply Lemma 4.13. If $x = x' + x''$, $x = x' \cdot x''$, or $x = \overline{x'}$ then we apply Lemmas A.6, A.7 or A.8 from Appendix A, respectively. □

As mentioned before, having a complete axiomatisation of a fraction of Subjective Logic, in this case $\mathcal{EXP}+\mathcal{AV}s^3$, may help us to analyse Subjective Logic. We present an analysis of dogmatic beliefs in Subjective Logic, an analysis of choosing different mappings π_c^{-1} and an analysis of the relation between dilution and trust chaining.

Dogmatic Opinions

There are, as stated before, no dogmatic belief triples in $\mathcal{EXP}+\mathcal{AVs}^3$. There are *dogmatic beliefs* in $\mathcal{BDU}_i$ and $\mathcal{BDU}_\gamma$, which posed problems in their axiomatisation. For that reason, we studied $\mathcal{EXP}$ and $\mathcal{EXP}+\mathcal{AVs}^3$ instead. The axiomatisations $\mathbf{FDN}s + \mathbf{AV}s^3$ and $\mathbf{FDN} + \mathbf{AV}^3$ are meant to completely axiomatise $\mathcal{EXP}+\mathcal{AVs}^3$, and successfully do so. An advantage of axiomatisations in general, is that they can easily be adopted to a different set of terms. We may analyse how the axioms of $\mathbf{FDN}s + \mathbf{AV}s^3$ (or $\mathbf{FDN} + \mathbf{AV}^3$) apply to dogmatic beliefs.

Definition 4.10 (Signature $\Sigma_{BDU+AVs^3}$)**.** Let the signature $\Sigma_{BDU+AVs^3}$ denote:

$$\varphi ::= \beta \mid \delta \mid \upsilon \mid \mathbf{0} \mid \mathbf{1} \mid \pm(\varphi, \ldots, \varphi) \mid \varphi + \varphi \mid \varphi \cdot \varphi \mid \overline{\varphi}$$

We can also apply $\mathbf{FDN}s + \mathbf{AV}s^3$ and $\mathbf{FDN} + \mathbf{AV}^3$ to $\Sigma_{BDU+AVs^3}$. Contrary to $\Sigma_{EXP+AVs^3}$, $\Sigma_{BDU+AVs^3}$ contains averages with zero parameters. Let $\mathbf{FDN}s + \mathbf{AV}s^3$ be the axiomatisation of $\Sigma_{BDU+AVs^3}$, since all axioms are positive, there is at least a model that adheres to $\mathbf{FDN}s + \mathbf{AV}s^3$.

The fusion of a dogmatic *trust opinion* and a non-dogmatic opinion, is the dogmatic opinion, via $\mathbf{D7}_\infty$ and $\mathbf{A5}_\infty$: $\mathbf{FDN}s + \mathbf{AV}s^3 \vdash \pm(\left(\beta\right)_l, \left(\delta\right)_m, \left(\upsilon\right)_n) + \pm(\left(\beta\right)_{l'}, \left(\delta\right)_{m'}, \left(\upsilon\right)_0) = \pm(\left(\beta\right)_{l\cdot 0+l'\cdot n}, \left(\delta\right)_{m\cdot 0+m'\cdot n}, \left(\upsilon\right)_{n\cdot 0}) = \pm(\left(\beta\right)_{l'}, \left(\delta\right)_{m'}, \left(\upsilon\right)_0)$. When we look at the fusion of two arbitrary dogmatic opinions, the results are surprising: $\mathbf{FDN}s + \mathbf{AV}s^3 \vdash \pm(\left(\beta\right)_l, \left(\delta\right)_m, \left(\upsilon\right)_0) + \pm(\left(\beta\right)_{l'}, \left(\delta\right)_{m'}, \left(\upsilon\right)_0) = \pm(\left(\beta\right)_{l\cdot 0+l'\cdot 0}, \left(\delta\right)_{m\cdot 0+m'\cdot 0}, \left(\upsilon\right)_{0\cdot 0}) = \pm()$, an empty average. We can view the empty average like a special element, similar to inconsistency (i) in $\mathcal{BDU}_i$, defined in Section 4.1. The surprising part is that even the fusion of complete (dis)trust with complete (dis)trust yields a contradiction, as $\mathbf{FDN}s + \mathbf{AV}s^3 \vdash \beta + \beta = \pm()$. That is similar to $(1,0,0) \oplus (1,0,0) = i$ in $\mathcal{BDU}_i$, and in Section 4.1 we argue why this is surprising. As one expects, if we fuse anything with contradiction, we get a contradiction, due to multiplication by zero. Dilution of one or two dogmatic opinions is not essentially different from non-dogmatic opinions. The dilution of contradictions are a more interesting case. If we have a contradiction on the right-hand side, we get: $\mathbf{FDN}s + \mathbf{AV}s^3 \vdash \pm(\left(\beta\right)_l, \left(\delta\right)_m, \left(\upsilon\right)_n) \cdot \pm() = \pm(\left(\beta\right)_{l\cdot 0}, \left(\delta\right)_{l\cdot 0}, \left(\upsilon\right)_{l\cdot 0+(m+n)\cdot(0+0+0)}) \equiv \pm()$. This is not in line with **B6**, as it dictates that $\upsilon \cdot \pm() = \upsilon$ nor **B10** as it dictates that $\delta \cdot \pm() = \upsilon$. If we have a contradiction on the left-hand side, we get: $\mathbf{FDN}s + \mathbf{AV}s^3 \vdash \pm() \cdot \pm(\left(\beta\right)_l, \left(\delta\right)_m, \left(\upsilon\right)_n) = \pm(\left(\beta\right)_{0\cdot l}, \left(\delta\right)_{0\cdot m}, \left(\upsilon\right)_{0\cdot n+(0+0)\cdot(l+m+n)}) \equiv \pm()$, again contradiction. Regardless of the left-hand side, the dilution will yield a contradiction, contrary to Proposition 4.1. So even if the recommending user A is a known liar (i.e. our opinion of him is δ), and we should reject whatever he says, as dictated by **B10**, when A claims a contradiction, our resulting opinion becomes a contradiction..

The axioms of $\mathbf{FDN}s + \mathbf{AV}s^3$ applied to $\Sigma_{BDU+AVs^3}$ do not yield the same equalities as the axioms of **BDU**. Furthermore, $\mathbf{FDN}s + \mathbf{AV}s^3$ is neither sound with $\mathcal{BDU}_i$ nor $\mathcal{BDU}_\gamma$. As $\mathbf{FDN}s + \mathbf{AV}s^3$ is an axiom scheme, we could trivially tweak cases where we have an empty average, or an average containing no uncertainty, to make it in line with $\mathcal{BDU}_i$. Similarly, we can add an axiom scheme for the fusion of two dogmatic opinions, taking the average of the two dogmatic opinions. In that case the axioms correspond to $\mathcal{BDU}_\gamma$. However, in itself this is not an interesting

exercise, as we can already see that neither axiomatisations will be so self-evident that it helps us decide which model corresponds to our interpretation of aggregation of dogmatic opinions.

Superficially, we might expect that $\mathbf{FDN} + \mathbf{AV}^3$ is no different from $\mathbf{FDN}s + \mathbf{AV}s^3$ over $\Sigma_{BDU+AVs^3}$. We, however, only proved the correspondence of $\mathbf{FDN}s + \mathbf{AV}s^3$ and $\mathbf{FDN} + \mathbf{AV}^3$ for Σ_{AVs^κ} terms. In $\mathbf{FDN} + \mathbf{AV}^3$, we have incidentally introduced dogmatic opinions, to help us with our axiomatisation. The fusion of a dogmatic opinion with any non-dogmatic opinion will yield the dogmatic opinion. Let x be a non-dogmatic opinion with n υ's in its unique normal form (Definition 4.6), and let y be a dogmatic opinion. By applying **D7** once, **D13** n times, **D10** once, and $\mathbf{A5}_\infty$ n times, we get that $\mathbf{FDN} + \mathbf{AV}^3 \vdash x + y = y$. If x and y are both dogmatic, we cannot apply any of **D13**, **D14**, **D10**, or **D11**. Therefore, the fusion of dogmatic x and y cannot be simplified into its unique normal form. In $\mathbf{FDN}s + \mathbf{AV}s^3$, the fusion of any dogmatic x and y was equal to $\pm()$, so for dogmatic x, x', y, y', $\mathbf{FDN}s + \mathbf{AV}s^3 \vdash x + y = x' + y'$, but $\mathbf{FDN} + \mathbf{AV}^3 \nvdash x + y = x' + y'$. Just as in $\mathbf{FDN}s + \mathbf{AV}s^3$, dilution of dogmatic opinions in no problem in $\mathbf{FDN} + \mathbf{AV}^3$. The dilution of a fusion of dogmatic terms can also not be simplified. By applying dilution and fusion recursively, we get more and more convoluted terms not in $\Sigma_{EXP+AVs^3}$. In a way, all these convoluted terms correspond to an intuition of inconsistency. It is possible to add axioms that equate these convoluted terms, without any influence on equality of $\Sigma_{EXP+AVs^3}$ terms. In effect, these axioms equate all the convoluted terms to inconsistency. It is also possible to add axioms from **BDU** about the dogmatic terms, without introducing contradictions. For example, we can add $D(x) \Rightarrow x + x = x$, or even $D(x) \wedge D(y) \Rightarrow x + y = \pm(x, y)$. In conclusion, the axiomatisations $\mathbf{FDN}s + \mathbf{AV}s^3$ and $\mathbf{FDN}s + \mathbf{AV}s^3$ can readily be modified to capture different models, such as $\mathcal{BDU}_i$ and $\mathcal{BDU}_\gamma$.

Experiments

One positive experiment is represented in the model as $(1/2, 0, 1/2)$. As we mentioned in Section 3.2, as far as fusion is concerned, we might have picked $(1/3, 0, 2/3)$, or any value bigger than zero. We have also showed that it does matter for dilution. Picking the experiments differently effectively changes the model. We analyse the changes caused by alternative choices for the experiments.

For example, $\mathbf{1} \cdot x$, represents a dilution, where a user A makes a claim x about B, and we had one positive experience with A. As $\mathbf{1}$ had a believe component of $1/2$, our belief about B is diluted by exactly half. The trust and the distrust we have in B is half of the (dis)trust A claims to have. One can successfully argue that half is too much, but also that half is not enough. one can change it either way by picking a different correspondence π_c^{-1} between experiments and Subjective Logic. Recall that: $\pi_c^{-1}(s, f) = (\frac{s}{s+f+c}, \frac{f}{s+f+c}, \frac{c}{s+f+c})$. If we map the positive experiment with parameter c, the ratio between u and b in $\mathbf{1}$ is c. Similarly, for the negative experiment $u : b = c$. By Proposition 4.7, if c is a rational number larger than zero, then there exists a representation as an average of β and υ for $\mathbf{1}$, and δ and υ for $\mathbf{0}$. Not surprisingly, the ratio between β's and υ's in such an average is c. If we want to change $\mathbf{FDN}s + \mathbf{AV}s^3$ (or $\mathbf{FDN} + \mathbf{AV}^3$) to represent $\mathcal{ECP}+\mathcal{AV}s^3(c)$, we simply replace $\mathbf{D1}_\infty$ and $\mathbf{D2}_\infty$ (or **D1** and **D2**) by $\pm(\left(\beta\right)_{c_d}, \ldots, \left(\upsilon\right)_{c_n})$, where $c = \frac{c_n}{c_d}$.

Now, we see that axiomatically τ_c is an isomorphism over fusion, as the ratios $\upsilon : \beta$ and $\upsilon : \delta$ remain constant:

$$\pm(\big(\beta\big)_{l\cdot c_n},\ \big(\delta\big)_{m\cdot c_n},\ \big(\upsilon\big)_{n\cdot c_d}) + \pm(\big(\beta\big)_{l'\cdot c_n},\ \big(\delta\big)_{m'\cdot c_n},\ \big(\upsilon\big)_{n'\cdot c_d}) = \{\mathbf{D7}_\infty\}$$
$$\pm(\big(\beta\big)_{l\cdot c_n\cdot n'\cdot c_d + l'\cdot c_n\cdot n\cdot c_d},\ \big(\delta\big)_{m\cdot c_n\cdot n'\cdot c_d + m'\cdot c_n\cdot n\cdot c_d},\ \big(\upsilon\big)_{n\cdot c_d\cdot n'\cdot c_d}) = \{\mathbf{A5}_\infty\}$$
$$\pm(\big(\beta\big)_{(l\cdot n' + l'\cdot n)\cdot c_n},\ \big(\delta\big)_{(m\cdot n' + m'\cdot n)\cdot c_n},\ \big(\upsilon\big)_{n\cdot n'\cdot c_d})$$

Above we proved that τ_c is an isomorphism on the model of Subjective Logic with fusion. Whereas here we proved that τ_c is *isomorphic* over the axioms of fusion and averaging. In the axiomatic variant above, the reason as to why τ_c is an isomorphism is clearer: As we proved in Proposition 4.7, every opinion can be represented as an average of β, δ, υ, changing the mapping results in a different but fixed ratio between β, δ and υ.

Given this intuition of the isomorphism, we can immediately see where dilution fails to be congruent. Given $x \cdot y = z$, we can see that the expression can be broken down to applications of **D15**, **D16** and **D17**. The ratio between β, δ and υ will not be respected. Let the fraction of υ increase, then the number of applications of **D17** goes up (making the fraction of υ in z increase) and the amount of υ in z provided by **D15** increases. The fact that it is no congruence therefore depends (at least partially) on the uncertainty-favouring nature of the dilution operator in Subjective Logic. In the next paragraph, we analyse the dilution operator and the fact that it favours uncertainty.

Dilution

Axioms **D15**, **D16** and **D17** provide a basis to derive equalities of dilution. If we have a dilution $x \cdot y$, then some information from y is retained but diluted by υ. The axiom **D15** allows the influence of y on the result, whereas **D16** and **D17** dilute with υ. Recalling Proposition 4.7, we can see that the ratio of the dilution is $b_x : d_x + u_x$.

Axiom **D17** is not self-evident, even if a trust chaining is a dilution, there is no apparent reason why the dilution ratio should be $b : d + u$. An immediate consequence of **D17** is that $\upsilon \cdot x = \upsilon$, also known as Propositions 4.1 and 4.5 in **BDU** and **EXP**, respectively. It states, in other words, that we completely disregard the claims of a stranger. Obviously, this is not a general truth, as we can easily imagine a situation where we (partially) believe a stranger. Moreover, dilution dictates that $(0.2, 0.75, 0.05)$ yields more information than $(0.19, 0, 0.81)$, when applied left of a dilution. Someone with whom you have many bad experiences and few good experiences is preferable (in Subjective Logic) to someone with whom you have slightly less good experiences, but no bad experiences at all. In Subjective Logic, the trustor is very paranoid about uncertainty. This observation is one of the central tenets of Section 5.2, where the argument set out in more detail. For now, we merely consider some alternatives for the axioms of dilution.

The first alternative is by ignoring the uncertainty component completely, when the opinion is left of a dilution. Axiomatically, we replace **D17** in $\mathbf{FDN} + \mathbf{AV}^3$ by a temporary axiom (**T6**): $(\upsilon, x)\boxdot\pm(y) = x\boxdot\pm(y)$. In the equation $\mathbf{FDN} + \mathbf{AV}^3[\mathbf{T6}] \vdash x \cdot y = z$, the uncertainty of x is ignored, and only the trust and distrust

components of x are relevant. In other words, the ratio of trust versus distrust in x determines how likely we think y is to be correct, more precisely, the dilution ratio is $b_x : d_x$. However, when there are neither trust nor distrust components in x, there is no well-defined ratio between them. An opinion x without trust or distrust equals υ. As we can see, $\upsilon \cdot x$ cannot be simplified into the unique normal form. Hence, a stranger making a claim leads to an expression that cannot be equated to an opinion. This is, of course, undesirable. However, it is not hard to see why the claim of a stranger leads to such problems. Consider the interpretation of beta distributions. The ratio $s : f$ equals the ratio $\beta : \delta$. In a beta distribution, the mode of the distribution is at $\frac{s}{s+f}$. We could have called the dilution axiomatised by $\mathbf{FDN} + \mathbf{AV}^3[\mathbf{T6}]$ mode-based dilution, as the mode of x determines the dilution ratio. In that light, it is not surprising that the mode-based dilution with a distribution without a mode (υ) is not consistent. Mode-based dilution solves the paranoid trustor problem.

The second alternative is to care about the mean instead of the mode. In that case we replace **D17** by a temporary axiom (**T7**): $\upsilon \boxdot \pm(x) = \upsilon, \pm(x)$. The dilution ratio of $x \cdot y$ is then $\beta_x + \upsilon_x : \delta_x + \upsilon_x$, or alternatively $\frac{\beta_x}{\upsilon_x} : \frac{\delta_x}{\upsilon_x}$. We can apply $\pi^{-1}((\beta_x, \delta_x, \upsilon_x)) = (\frac{\beta_x}{\upsilon_x}, \frac{\delta_x}{\upsilon_x},$ to get that the dilution ratio is $s_x + 1 : f_x + 1$, or $\alpha_x : \beta_x$. The mean of a beta distribution α, β is $\frac{\alpha}{\alpha+\beta}$. The dilution ratio, therefore, corresponds exactly to the mean. By increasing u (relative to **T6**), both sides of the dilution ratio are increased, making the ratio less sensitive for differences between b and d. Formulated geometrically, by increasing u, the beta distribution becomes more flattened, making it less sensitive for differences between b and d. The mean-based solution uses the expected ratio of successes and failures of the *recommender* to determine the validity of the recommendation.

4.5 AND and OR of Opinions from Experiments

The structure of this section is similar to that of Sections 4.3 and 4.4. First we provide an intuitive complete axiom scheme, then we construct a finite complete axiomatisation based thereupon.

In Section 4.4, we showed that we have a complete finite axiomatisation ($\mathbf{FDN} + \mathbf{AV}^3$) of consensus, discounting and complement of belief triples, using fusion, dilution and inversion of opinions (proven in Lemma 4.17). In this section, we extend the axiomatisation to encompass operators for multiplication and co-multiplication. We refer to the resulting model as $\mathcal{SL}$. The operators we add to the signature are called *AND* and *OR*, denoted $\_ \wedge \_$ and $\_ \vee \_$, respectively. We define the signature Σ_{SLs} as:

Definition 4.11 (Signature Σ_{SLs})**.**

$$\varphi ::= \upsilon \mid \mathbf{0} \mid \mathbf{1} \mid \pm(\varphi, \ldots, \varphi) \mid \varphi + \varphi \mid \varphi \cdot \varphi \mid \overline{\varphi} \mid \varphi \wedge \varphi \mid \varphi \vee \varphi$$

Before we provide the complete axiom scheme **SL**s of $\mathcal{SL}$, we look at three possible axioms. The first is commutativity of AND, (**N1**): $x \wedge y = y \wedge x$. The second is associativity of AND, (**N2**): $(x \wedge y) \wedge z = x \wedge (y \wedge z)$. Both statements are true, but we do not use them as axioms, as they are derivable from the axiom

$$(\mathbf{D23}_\infty)\ \pm((\beta)_l,\ (\delta)_m,\ (\upsilon)_n) \wedge \pm((\beta)_{l'},\ (\delta)_{m'},\ (\upsilon)_{n'}) = \pm((\beta)_{l\cdot l'},\ (\delta)_{m\cdot(l'+m'+n')+m'\cdot(m+n)},\ (\upsilon)_{l\cdot n'+n\cdot l'+n\cdot n'})$$

$$(\mathbf{D24}_\infty)\ x \vee y = \overline{\overline{x} \wedge \overline{y}}$$

Figure 4.7: Axiom scheme of AND and OR in terms of Averaging. Forms **SL***s* together with **FDN***s* + **AV***s*^{3}.

scheme we provide in **SL***s*. However, we introduce them as temporary axioms here, since they are evidently true, and, together with the De Morgan rule below, they suffice to prove an interesting duality property. We furthermore introduce one of the De Morgan rules, ($\mathbf{D24}_\infty$): $x \vee y = \overline{\overline{x} \wedge \overline{y}}$. By adopting this De Morgan rule, we merely need to define AND in terms of opinion averaging, as OR is stated in terms of negation and AND. Before we define AND in terms of opinion averaging, we note that the dual of these three statements can be derived from just axioms **N1**, **N2** and $\mathbf{D24}_\infty$:

Proposition 4.18. *In any axiom scheme where* **I1**, **N1**, **N2**, $\mathbf{D24}_\infty$ *are derivable, the statements* $x \vee y = y \vee x$, $(x \vee y) \vee z = x \vee (y \vee z)$ *and* $x \wedge y = \overline{\overline{x} \vee \overline{y}}$ *are also derivable.*

Proof.

$$\begin{aligned} x \vee y &= \overline{\overline{x} \wedge \overline{y}} = \overline{\overline{y} \wedge \overline{x}} = y \vee x \\ (x \vee y) \vee z &= \overline{\overline{(x \vee y)} \wedge \overline{z}} = \overline{\overline{x} \wedge \overline{y} \wedge \overline{z}} = \overline{\overline{x} \wedge \overline{(y \vee z)}} = x \vee (y \vee z) \\ x \wedge y &= \overline{\overline{\overline{x} \wedge \overline{y}}} = \overline{\overline{x} \vee \overline{y}} \end{aligned}$$

□

Let $\{\beta, \delta, \upsilon\} = A^3$, then we present a complete axiom scheme (**SL***s*) of $\mathcal{SL}s$ in Figure 4.7.

The axioms **N1** and **N2** are subsumed by axiom $\mathbf{D23}_\infty$ in **SL***s*:

Proposition 4.19. *In* **SL***s*, $x \wedge y = y \wedge x$ *and* $(x \wedge y) \wedge z = x \wedge (y \wedge z)$ *are true.*

Proof. Assume $x = \pm((\beta)_l,\ (\delta)_m,\ (\upsilon)_n)$, $y = \pm((\beta)_{l'},\ (\delta)_{m'},\ (\upsilon)_{n'})$ and $z = \pm((\beta)_{l''},\ (\delta)_{m''},\ (\upsilon)_{n''})$ are the unique normal forms of x, y and z. For commutativity, it suffices to verify that swapping l for l', m for m' and n for n' changes nothing. For associativity, it suffices to verify that there are $l \cdot l' \cdot l''$ instances of β, $m \cdot (l' + m' + n') \cdot (l'' + m'' + n'') + m' \cdot (l + n) \cdot (l'' + m'' + n'') + m'' \cdot (l + n) \cdot (l' + n')$ instances of δ and $l \cdot l' \cdot n'' + l \cdot n' \cdot l'' + n \cdot l' \cdot l'' + n \cdot n' \cdot l'' + n \cdot l' \cdot n'' + l \cdot n' \cdot n'' + n \cdot n' \cdot n''$ instances of υ in both averages. □

The axiom scheme **SL***s* is sound with respect to $\mathcal{SL}s$:

Lemma 4.20. *If* $\mathbf{SL}s \vdash x = y$ *then* $\mathcal{SL}s \models x = y$.

Proof. Verifying that both axioms are true in the model is straightforward. □

And we prove completeness:

(**D23**) $\pm(x \wedge y) = \pm(x) \boxtimes \pm(y)$	(**D24**) $x \vee y = \overline{\overline{x} \wedge \overline{y}}$	
(**D25**) $(\beta, x) \boxtimes y = y, (x \boxtimes y)$	(**D30**) $P_\delta(\beta, x) = \delta, P_\delta(x)$	(**D36**) $P_\upsilon(\beta, x) = \upsilon, P_\upsilon(x)$
(**D26**) $(\delta, x) \boxtimes y = P_\delta(y), (x \boxtimes y)$	(**D31**) $P_\delta(\delta, x) = \delta, P_\delta(x)$	(**D37**) $P_\upsilon(\delta, x) = \delta, P_\upsilon(x)$
(**D27**) $(\upsilon, x) \boxtimes y = P_\upsilon(y), (x \boxtimes y)$	(**D38**) $P_\delta(\upsilon, x) = \delta, P_\delta(x)$	(**D38**) $P_\upsilon(\upsilon, x) = \upsilon, P_\upsilon(x)$
(**D28**) $\beta \boxtimes y = y$	(**D33**) $P_\delta(\beta) = \delta$	(**D39**) $P_\upsilon(\beta) = \upsilon$
(**D29**) $\delta \boxtimes y = P_\delta(y)$	(**D34**) $P_\delta(\delta) = \delta$	(**D40**) $P_\upsilon(\delta) = \delta$
(**D30**) $\upsilon \boxtimes y = P_\upsilon(y)$	(**D41**) $P_\delta(\upsilon) = \delta$	(**D41**) $P_\upsilon(\upsilon) = \upsilon$

Figure 4.8: Finite axiomatisation of AND and OR in terms of Averaging. Forms **SL** together with $\mathbf{FDN} + \mathbf{AV}^3$.

Lemma 4.21. *If* $\mathcal{SL}s \models x = y$ *then* $\mathbf{SL}s \vdash x = y$.

Proof. We append two cases to the structural induction found in the proof of Lemma 4.17, one for AND and one for OR.

> $[x = x' \wedge x'']$ then, by induction hypothesis, $x' \equiv \pm(\left(\beta\right)_l, \left(\delta\right)_m, \left(\upsilon\right)_n)$ and $x'' \equiv \pm(\left(\beta\right)_{l'}, \left(\delta\right)_{m'}, \left(\upsilon\right)_{n'})$. Apply $\mathbf{D23}_\infty$ to $x' \wedge x''$.
>
> $[x = x' \vee x'']$ then apply $\mathbf{D24}_\infty$ and the cases for inversion and AND. □

Complete finite axiomatisation

In order to construct a finite axiomatisation, we introduce some more auxiliary operators. In particular, we introduce pairing sensitive AND ($\_ \boxtimes \_$), similar to the other pairing sensitive operations (e.g. pair sensitive fusion, $\_ \boxplus \_$).

Definition 4.12 (Signature $\Sigma^{\square}_{SLs}$)**.** We define the signature $\Sigma^{\square}_{SLs}$ as:

$$\varphi ::= \upsilon \mid \mathbf{0} \mid \mathbf{1} \mid \varphi + \varphi \mid \varphi \cdot \varphi \mid \overline{\varphi} \mid \varphi \wedge \varphi \mid \varphi \vee \varphi \mid \pm(\psi)$$

$$\psi ::= \varphi \mid \psi, \psi \mid \psi \boxplus \psi \mid \psi \boxdot \psi \mid {\sim}(\psi) \mid D(\psi) \mid \psi \boxtimes \psi$$

The finite axiomatisation **SL** is an extension of $\mathbf{FDN} + \mathbf{AV}^3$, and adds the axioms displayed in Figure 4.8.

The following lemma states the relation between the two axiom systems.

Lemma 4.22. *For all terms* x, y *in* Σ_{SLs}, $\mathbf{SL}s \vdash x = y$ *iff* $\mathbf{SL} \vdash x = y$

Proof. Rules $\mathbf{D24}_\infty$ and **D24** are identical, we can ignore the cases for OR.

Hence, to prove the right implication – if $\mathbf{SL}s \vdash x = y$, then $\mathbf{SL} \vdash x = y$ – it suffices to prove that $\mathbf{D23}_\infty$ is derivable from **SL**.

To prove the left implication – if $\mathbf{SL} \vdash x = y$, then $\mathbf{SL}s \vdash x = y$ – we reuse the proof of Lemma 4.17, and allow AND as an alternative in the structural induction.

In other words, it suffices to prove that if $\mathbf{SL} \vdash x = y$ and $x \equiv x' \wedge x''$, for y, x', x'' in the unique normal form, then $\mathbf{SL}s \vdash x = y$. Lemma A.9 – found in Appendix A – proves exactly that. □

5

Axiomatisation of Trust Operations

In the previous chapter, we used Subjective Logic as the semantics of trust networks. That means that we took consensus, discounting, multiplication, comultiplication and complement as correct implementations of *trust aggregation*, *trust chaining*, trust conjunction, trust disjunction and trust negation. In this chapter, we depart from *Subjective Logic*. First, in Section 5.1, we identify some issues with the axiomatisation, and thus with the model. Then, in Section 5.2, we drop (or weaken) the axioms that caused problems. The resulting axiomatisation is consistent with Subjective Logic, but does not contain the objectionable axioms identified in Section 5.1. Finally, in Section 5.3, we do without the relationship to Subjective Logic, and directly identify axioms.

5.1 Identifying Issues

In Section 3.2, we showed that π_1 is not the only isomorphism between beta distributions and belief triples with regards to trust aggregation (and consensus). When we restrict our view to trust aggregation, all choices for c are, therefore, equal[1]. The choice of c does, however, influence the other operators. There is no self-evident reason to prefer a certain value over another. The only reason we selected 1 was because it was notationally convenient. Without a form of justification, we cannot accept axioms that encode a specific choice of c.

As we saw in Section 3.2, we can view opinions as beta distributions. We can look at how the actual beta distributions transform over the axiomatised trust operations. In particular, we are going to sketch a scenario in Example 5.1 where we look at two statements equal in **SL**, that we do not expect to be equal when we look at them as distributions.

Example 5.1. Consider a machine B that is very unreliable and fails between 99.9 and 100% of the time (i.e. $0 \leq p \leq 0.001$ for reliability p of B). Further, consider a *user* A, with whom we had some dealings previously. Our experience with A is mildly negative, say one positive and two negative ($\mathbf{1} + \mathbf{0} + \mathbf{0}$). In this particular case, A is lying when he gives us his opinion on B. He claims that B is quite reliable, by stating that in his interactions with B he observed four successes and one failure ($\mathbf{1} + \mathbf{1} + \mathbf{1} + \mathbf{1} + \mathbf{0}$). Applying dilution, we would get opinion $(\mathbf{1}+\mathbf{0}+\mathbf{0})\cdot(\mathbf{1}+\mathbf{1}+\mathbf{1}+\mathbf{1}+\mathbf{0}) = \pm(\beta, \delta, \delta, \upsilon)\cdot\pm(\beta, \beta, \beta, \beta, \delta, \upsilon) = \pm(\big(\beta\big)_4, \big(\delta\big)_1, \big(\upsilon\big)_{19})$ If we want to study the probability we assign to B having a reliability between 0 and 0.1%, we should calculate the beta distribution. Since $(s, f) = \pi((1/6, 1/24, 19/24)) = (4/19, 1/19)$, we have a beta distribution given by $\alpha = s+1 = 23/19, \beta = f+1 = 20/19$.

[1] Note that in later versions of Subjective Logic, where the base rate is introduced (such as [JOO10]), this statement is no longer valid.

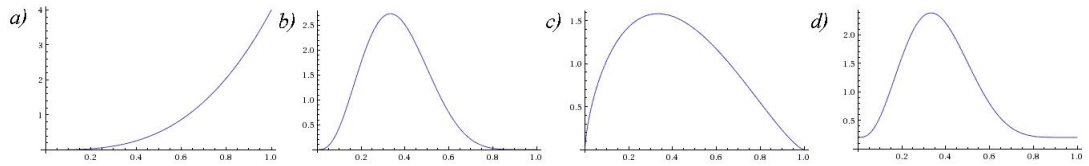

Figure 5.1: Figures a and b are given probability functions. Figure c is the SL style dilution of a and b. Figure d is the expected shape of the chain of a and b.

We take the definite integral of the probability density function between 0 and 0.01, which is:

$$\frac{\int_0^{0.001}(t^{4/19}(1-t)^{1/19}dt)}{\int_0^1(t^{4/19}(1-t)^{1/19}dt)} \approx 0.00000517$$

We assigned a near zero probability, around $5 \cdot 10^{-}6$, to the p-value of B being $0 \leq p \leq 0.001$. However, directly looking at the trust network (without applying dilution or discounting), one may conclude that the actual probability of that p-value is over 100 times larger. The user A is either lying, or telling the truth. We can calculate the expected reliability of A. The expected p-value of A, can be computed using π to be $2/5$. Machine A is expected to succeed 40% of the time, and thus expected to lie 60% of the time. If A lies, his claim is vacuous. We fall back on the basic assumption that all p-values are equally likely. We expect that 60% of the time, there is a 0.1% chance of $0 \leq p \leq 0.001$. In other words, we expect a probability of at least 0.0006 of B having a p-value in the first per mil.

The example shows that in Subjective Logic, it is possible for a recommender deemed unreliable (i.e. user A) to still significantly alter our expectations. The reader is invited to repeat the example for the alternative mappings $\pi_c()$, and mode-based and mean-based dilution as proposed in Section 4.4. We need not explicitly calculate the outcomes, as the underlying reason of the near-zero values at the extremes, is the fact that all Subjective Logic opinions have a particular shape, as can be seen in Figure 5.1. The problem in Example 5.1 is an immediate consequence of the fact that all beta distributions except the uniform distribution have at least one low tail. Probability distributions with heavy tails cannot be expressed as beta distributions, but we expect trust chaining via an unreliable source to have exactly such a distribution. Figure 5.1 shows the two distributions (c and d) of opinions that are the result of a dilution (of a and b). Graph c shows the result for dilution as it is in Subjective Logic, graph d shows a distribution with a more accurate shape. If we look at the range $0.9 \leq p \leq 1$, graph c has significant differences for 0.9 and 1, graph d does not. In graph b values at 0.9 and 1 are both near zero, and thus similar. If the user claiming b was truthful, then there is little difference between the density at 0.9 and 1, if the user was lying, then the distribution is uniform, each value has equal density. The resulting graph of dilution of b should therefore have little difference between masses at 0.9 and 1.

The axioms **B6** and $\mathbf{B6_{0,1}}$ should also not be accepted as axioms. The dilution $\upsilon \cdot x$ expresses that a user that you have no information of, stated a particular opinion. It is no more than normal to be very skeptical about the truth of the opinion, but there clearly is a non-zero probability that x is the true opinion of the stranger. Scenarios where it is better to use a stranger's *recommendation* are commonplace,

therefore always ignoring it should not be an axiom. The same argument holds for $\mathbf{B10_{0,1}}$ (and its temporary generalisation $\mathbf{T1_{0,1}}$), even though a user only behaved badly (once), we cannot exclude the possibility that he is behaving correctly now. It is, in other words, not self-evident that the recommendation of an unknown or unreliable user must be completely discarded.

We furthermore reevaluate associativity of dilution ($\mathbf{B3}$, $\mathbf{B3_{0,1}}$) with an example:

Example 5.2. Consider users A, B and C. You want to form an opinion on C. Consider the following two scenarios, in both scenario's you have bad experiences with A, leading you to an extremely skeptical opinion x:

- You go to A end ask him his opinion about C. His response is: "I don't have any personal opinion, but I've had some interactions with B that lead me to have opinion y about him, and B said he had opinion z about C," which is simply $y \cdot z$. Your opinion about C is therefore $x \cdot (y \cdot z)$. Since we are very skeptical about A speaking the truth, we are not at all sure what B really said, nor how reliable B is. As a consequence, the information we have about C is almost nothing. More precisely, the more skeptical we are about A, the closer our opinion gets to υ.
- You've had no personal interactions with B, who claimed to have opinion z about C. Your opinion about A is x, and A said he had opinion y about B, leading you to have opinion $x \cdot y$ about B. Your opinion about C is now $(x \cdot y) \cdot z$. User A claimed that his opinion about B is y. However, our opinion is that A almost never speaks the truth. We have almost no information about the behaviour of B, B is almost a stranger. Contrary to Subjective Logic, in this context we generally do not reject the opinions of strangers, so our opinion is close to $\upsilon \cdot z$. The more skeptical we are about A, the closer our opinion gets to $\upsilon \cdot z$.

In the first scenario, when A is lying we have no information about C, as we do not know what B said. In the second scenario, when A is lying, we still know what a stranger thinks of C. If $\mathbf{B6}$ and $\mathbf{B6_{0,1}}$ are rejected, then the results of the first and second scenario are not (generally) equal.

Finally, we look at trust conjunction (and via De Morgan at trust disjunction). Before doing so, recall that the expected value of a belief triple is $\frac{b+u}{1+u}$. If we provide an opinion mean $\pm(\left(\beta\right)_l, \left(\delta\right)_m, \left(\upsilon\right)_n)$, it immediately follows that its expected value is $\frac{l+n}{l+m+2n}$. So, the expected value of $\pm(\left(\beta\right)_l, \left(\delta\right)_m, \left(\upsilon\right)_n) \wedge \pm(\left(\beta\right)_{l'}, \left(\delta\right)_{m'}, \left(\upsilon\right)_{n'}) = \pm(\left(\beta\right)_{l \cdot l'}, \left(\delta\right)_{m+m'-m \cdot m'}, \left(\upsilon\right)_{l \cdot n' + l' \cdot n + n \cdot n'})$ is $\frac{(l+n) \cdot (l'+n')}{(l+m+n) \cdot (l'+m'+n') + l \cdot n' + l' \cdot n + n \cdot n'}$. However, the expected value of B succeeding and C succeeding, should be the product of B succeeding and C succeeding. In other words, the expected value should be $\frac{l+n}{l+m+2n} \cdot \frac{l'+n'}{l'+m'+2n'} = \frac{(l+n) \cdot (l'+n')}{(l+m+2n) \cdot (l'+m'+2n')}$. The expected value that AND predicts differs from the evident expected value[2].

[2] This issue has been identified and addressed, as multiplication and comultiplication have been updated in Subjective Logic in [JM05] to ensure that expected value of (co)multiplication is the correct expected value. The new definition of (co)multiplication yields the correct expected value, but still not the right probability distribution (Theorem 7.10). As neither definition yields the right probability distribution, we chose to remain close to a simpler and more intuitive definition of (co)multiplication.

5.2 Axiomatisation of Trust Opinions

In Section 5.1, we identified a collection of reasons why **SL** and **FDN** + **AV**^{3} are not valid axiomatisations. In fact, we saw that even the axiomatisation **EXP** contains some problematic axioms: $\mathbf{B3_{0,1}}$, $\mathbf{B6_{0,1}}$ and $\mathbf{B10_{0,1}}$.

Before we start the selection and analysis of axioms, we make an assumption about trust chaining. Without such an assumption, providing axioms of dilution would be difficult, and the axioms would be weaker. We assume that failed recommendations are distributed the same way as successful recommendations a priori, but carry no correlation with reality. In particular, we assert that receiving a known lie carries no information (which is expressed by total uncertainty, υ). We note that this assumption underlies all three axioms regarding dilution in this section. In Part II, we have sufficient formal machinery to look at alternatives of this assumption, for now it suffices to note that the assumption does not lead to inconsistencies (Proposition 8.4).

The axiomatisation should have *fusion*, *dilution*, *AND*, *OR* and *inverse*. We do not include opinion averaging, since the operator was only introduced as an auxiliary operator. The remaining signature is Σ_{ATC}:

Definition 5.1 (Signature Σ_{ATC})**.**

$$\varphi ::= \upsilon \mid \mathbf{0} \mid \mathbf{1} \mid \varphi + \varphi \mid \varphi \cdot \varphi \mid \overline{\varphi} \mid \varphi \wedge \varphi \mid \varphi \vee \varphi$$

By rejecting $\mathbf{B3_{0,1}}$, $\mathbf{B6_{0,1}}$ and $\mathbf{B10_{0,1}}$, Proposition 4.5 becomes invalid. However, the conclusion of that proposition appears self-evident. The recommender in the trust chain either speaks the truth, in which case υ describes the target, or the recommender is lying, in which case we have no information about the target, represented by υ. In either case, we end up with the *trust opinion* υ. Hence, we propose the weaker axiom $x \cdot \upsilon = \upsilon$ to replace the three axioms.

The De Morgan rule from Section 4.5 ($\mathbf{D24_\infty}$) appears self-evident. The axiom $\mathbf{D24_\infty}$ causes the issues with AND and OR identified in Section 5.1, and is rejected. The rejection of $\mathbf{D23_\infty}$ invalidates Proposition 4.19. However, the conclusion of that Proposition, commutativity and associativity of AND, appears self evident. We therefore accept **N1** and **N2**. These three axioms, together with double negation, are sufficient to derive their duals (Proposition 4.18).

We gather the self-evident axioms into one axiomatisation, **ATC**, in Figure 5.2. The signature of the **ATC** is again Σ_{EXP} as in Section 4.2. Recall that Σ_{EXP} has three constants, υ, **0**, **1**, and two binary operations, $\_ + \_$ and $\_ \cdot \_$. The constants are intended to keep their meaning, just as the fusion operator, but $x \cdot y$ should denote dilution in our axioms.

All axioms are axioms of **EXP**, but not vice versa. Let's go through the axioms of **ATC**, while keeping **EXP** (and **BDU**) in mind.

(**C1**) $x + \upsilon = x$	(**C8**) $\overline{\overline{x}} = x$
(**C2**) $x + y = y + x$	(**C9**) $\overline{\upsilon} = \upsilon$
(**C3**) $(x + y) + z = x + (y + z)$	(**C10**) $\overline{\mathbf{1}} = \mathbf{0}$
(**C4**) $x \cdot \upsilon = \upsilon$	(**C11**) $\overline{x + y} = \overline{x} + \overline{y}$
(**C5**) $x \cdot (y \cdot z) = y \cdot (x \cdot z)$	(**C12**) $\overline{x \cdot y} = x \cdot \overline{y}$
(**C6**) $x \wedge y = y \wedge x$	(**C13**) $x \wedge y = \overline{\overline{x} \vee \overline{y}}$
(**C7**) $(x \wedge y) \wedge z = x \wedge (y \wedge z)$	

Figure 5.2: Axiomatisation of Fusion, Dilution and Negation (**ATC**)

As with axiom **B5**, there is an identity element of fusion (**C1**), which intuitively represents the opinion derived from zero information. Fusion should represent the operation of combining two opinions into a new opinion, order should therefore not matter, as expressed by **C2** and **C3**.

If there is a user A claiming he has opinion υ about B, then it does not even matter whether A is lying. If A is lying, you have no information, if A is speaking the truth, you have no information. Therefore, υ is the right-zero element of dilution, as expressed by **C4**, similar to Propositions 4.1 and 4.5.

Although it seems initially surprising that associativity is rejected, while left commutativity (**C5**) is retained, it becomes clear after giving it some more thought. Consider users A, B and C in a trust chain. If A is lying, then we have opinion υ (no information) about C. If A is telling the truth, but B is lying, then we still have opinion υ about C. If both are telling the truth, then we have opinion z about C. The order of the two recommenders is therefore irrelevant, leading to the axiom of left commutativity.

The intuition behind commutativity (**C6**) and associativity (**C7**) of AND is immediate. Any operator that is not commutative and associative is not trust conjunction.

The intuition behind the inverse is swapping trust and distrust, or swapping p for $1 - p$. Double negation (**C8**) holds, as $1 - (1 - p) = p$. Our assumption was that all p are equally likely parameters for the *Bernoulli distribution* of a machine that we do not have any information about. For such a machine, there is no difference between p and $1 - p$, and swapping them is an identity operation. Hence $\upsilon = \overline{\upsilon}$, axiom **C9**. Axiom **C10** expresses the idea that the basic experiments, **0** and **1**, are each other's duals. Axiom **C11** expresses the idea that fusion does not favour trust or distrust. When you swap trust and distrust in two opinions, fuse them, and swap trust and distrust again, the result should equal the fusion of the two original opinions.

The asymmetric distribution rule of trust chaining **C12** states that $\overline{x \cdot y} = x \cdot \overline{y}$. The rule follows from the notion that if the recommender lies, the subject has no information, i.e. υ, and $\overline{\upsilon} = \upsilon$. Observe that $\overline{x \cdot y}$ is $\overline{y}$ if the recommender is truthful, and $\overline{\upsilon}$ otherwise. Further, $x \cdot \overline{y}$ is also $\overline{y}$ if the recommender is truthful, and υ otherwise.

Finally, the De Morgan rule (**C13**) is justified with the notion that if both B and C succeed, then it is not the case that B failed or that C failed, and vice versa.

(**K1**) $\mathbf{E}(\upsilon) = 1/2$	(**K7**) $\mathbf{W}(\upsilon) = 0$
(**K2**) $\mathbf{E}(\mathbf{0}) = 1/3$	(**K8**) $\mathbf{W}(\mathbf{0}) = 1$
(**K3**) $\mathbf{E}(\mathbf{1}) = 2/3$	(**K9**) $\mathbf{W}(\mathbf{1}) = 1$
(**K4**) $\mathbf{E}(x \cdot y) = \mathbf{E}(x) \times \mathbf{E}(y) + \mathbf{E}(\overline{x}) \times \mathbf{E}(\upsilon)$	(**K10**) $\mathbf{W}(x + y) = \mathbf{W}(x) + \mathbf{W}(y)$
(**K5**) $\mathbf{E}(x \wedge y) = \mathbf{E}(x) \times \mathbf{E}(y)$	(**K11**) $\mathbf{W}(\overline{x}) = \mathbf{W}(x)$
(**K6**) $\mathbf{E}(\overline{x}) = 1 - \mathbf{E}(x)$	

(**K12**) $\mathbf{W}(x \cdot y) < \mathbf{W}(y)$
(**K13**) $\mathbf{E}(x) > \mathbf{E}(y) \Rightarrow \mathbf{W}(x \cdot z) > \mathbf{W}(y \cdot z)$
(**K14**) $\mathbf{W}(y) > \mathbf{W}(z) \Rightarrow \mathbf{W}(x \cdot y) > \mathbf{W}(x \cdot z)$
(**K15**) $\mathbf{W}(y) > \mathbf{W}(z) \Rightarrow \mathbf{W}(x \wedge y) > \mathbf{W}(x \wedge z)$
(**K16**) $\mathbf{E}(x) = \mathbf{E}(y) \Rightarrow x \cdot z = y \cdot z$

Figure 5.3: Axiomatisation of Expected Value and Weight of Trust Opinions in **ATC**(**EVW**)

5.3 Axiomatisation of Expected Value and Weight

The axiomatisation **ATC** provided in the previous section cannot prove several true statements regarding (degree of) belief and uncertainty. In this section, we introduce two new concepts: Expected value and weight. The expected value and weight are set for beta distributions. For a beta distribution based on s successes and f failures, the expected value is $\frac{s+1}{s+f+2}$ and the weight is $s + f$.

Subjective Logic's belief triples can be mapped to beta distributions, which can be mapped to a pair of expected value and weight. However, unlike subjective logic (where two opinions are equal if their belief, disbelief and uncertainty are equal), we do not assume that if the expected value and weight are equal, then the opinions are equal. Situations as in Figure 5.1 are among the reasons we cannot assume that. Part d of Figure 5.1 must have an expected value and a weight. However, there exists a beta distribution with that expected value and weight. Since it is not a beta distribution, there are at least two distributions with the same expected value and weight.

The expected value has a very precise meaning, but (at least for now) weight only has a specific meaning for particular opinions. Namely for opinions represented by beta distributions, where it is $s + f$. Although we do not yet have a general, semantics, we do know some properties of the weight, which we can axiomatise.

We present the axiomatisation in Figure 5.3, and discuss the axioms below.

The axioms **K1**, **K2** and **K3** encode the relation between trust opinions based on direct interactions and beta distributions. Recall that the expected value of a beta distribution is $\frac{s+1}{s+f+2}$.

Axiom **K4** relies on the assumption (mentioned in Section 5.2) that failed recommendations carry no information. The axiom expresses that the expected value of a dilution is the weighted average of the stated opinion and υ, where the weight is determined by the probability that the recommendation is successful.

Axiom **K5** expresses the basic notion of conjunction, namely that the probability of A and B is the product of their probabilities (assuming A and B are independent). We need not formulate a similar axiom for disjunction due to De Morgan.

Inversion of an opinion leads to negation of the expected value, as expressed in **K6**.

The axioms **K7**, **K8** and **K9** define the weight of the basic components, together with **K10**. Together, they ensure the weight of an opinion corresponds to the number of interaction they represent, when they consist only of direct interactions (and possible inverses, via **K11**).

When a term contains dilution, we can observe three things. First, **K12**, receiving a trust opinion y from a recommender always carries less weight (carries more uncertainty) than having trust opinion y by direct interactions, since there is a positive probability that the recommendation is a lie. Second, **K13**, a trust opinion from a more reliable recommender carries more weight than one from a less reliable one. Third, **K14** a trust opinion resulting from a recommendation with more weight, itself carries more weight than a trust opinion resulting from a recommendation with less weight. The last observation also derives from the assumption that failed recommendations are distributed the same way as real recommendations, since this prevents recommenders from consistently assigning higher weights to fake recommendations. This means that a recommendation with high weight has the same probability of being true as one with a low weight, but it carries more weight.

Axiom **K15** covers the notion that more information regarding a subtarget leads to more information about the target. For example, assume I depend on a seller and delivery service for an interaction, then learning about the seller is also learning about 'the seller and delivery service'. Note that together with **C6**, the symmetrical case holds, and together with **C13** and **C8** the dual case (OR) holds.

Finally, axiom **K16** encodes the notion that if two equal recommendations are expected to be true with equal probability, the resulting trust opinion must be equal.

Part II

The Probabilistic Method

6

The Beta Model

Trust is closely related to probability. With all factors considered equal, a target with a higher probability of success is (or should be) more trusted than a target with a lower probability of success. In the case of *correctness trust models*, this is an important principle. In this part of the thesis, we formulate some probabilistic principles and derive definitions of the operators (trust aggregation, trust chaining and the logical trust operations) from these principles.

There is a sharp contrast between this part and Part I. In Part I, we formulated axioms regarding the operators, and studied the models that satisfy these axioms. Here, we provide the semantics of the operators in terms of probabilistic statements, and derive their models from the basic principles regarding the random variables in the probabilistic statements. This chapter serves multiple purposes, one of which is to provide a concrete example of such a technique, namely the Beta model. Chapters 7 and 8 use the same general technique, but introduce new random variables and their relation to other random variables.

The Beta model is not new. The idea of applying the *beta distribution* to trust was introduced in [MM02] and [JI02]. The philosophic view on the Beta model in those papers is, however, different from ours. There, the view is that people have trust opinions, people aggregate trust opinions, and a trust model must have a computational way of mimicking such aggregation. They identify that the beta distribution is an appropriate and effective basis for such computations. Such an approach leans towards cognitive trust models. Our philosophy is inspired by ElSalamouny et al. [ESN10, ElS11] (who have coined the term Beta model), where, rather than using probability theory as a tool to model trust, trust is defined in terms of probability theory. Trust opinions have a clear probabilistic meaning. That philosophy has the benefit that computations can be proven correct, and the benefit that the effects of changing the assumptions is immediate.

We are more explicit in defining trust in probability theory than ElSalamouny et al. In fact, we are the first to formulate the Beta model with such an explicit distinction between the probabilistic principles and the probabilistic statements that we are interested in. That distinction allows us to readily interpret the probabilistic statements as trust opinions in a trust model. In other words, we can keep the distance between probability theory and trust models small (formally, probabilistic statements and trust opinions are *isomorphic*).

Since we can keep the distance between probability theory and trust models small, we can achieve the following: First, we can analyse (existing) trust models to verify whether they adhere to the basic probabilistic principles of the Beta paradigm. Second, we can readily create a trust model based on computations derived from

the probabilistic principles of the Beta paradigm. In this chapter, the trust model extracted from the Beta paradigm is the Beta model. The Beta model is the only model (up to isomorphism) in the Beta paradigm, when we restrict ourselves to *trust aggregation*.

In Section 6.1, we introduce the notions of probability theory that are necessary to construct the Beta model, and its extensions with *trust chaining* and *logical trust operations*. Then, we provide the Beta model in Section 6.2.

6.1 Preliminaries

In the Beta paradigm, trust (Chapters 7 and 8) is probabilistic. We require the following concepts from probability theory (see e.g. [Bil95, Gut07]). Theorems and propositions are lifted from such textbooks without proof in this section. Readers with an understanding of probability theory may only find Theorem 6.3 and Definition 6.7, at the end of this section, of interest.

Definition 6.1 (σ-algebra, measure, probability measure)**.** Let Ω be a set of events. A set $\mathcal{F}$ of subsets of Ω is called a σ-algebra if the following three properties hold.

1. $\emptyset \in \mathcal{F}$.
2. If $A \in \mathcal{F}$ it follows that $\Omega \setminus A \in \mathcal{F}$.
3. If $A_1, A_2, \ldots \subset \mathcal{F}$ it follows that $\bigcup_n A_n \in \mathcal{F}$.

Let P be a map from $\mathcal{F} \to \mathbb{R} \cup \{\infty\}$. Then, this map is called a measure if

1. $P(\emptyset) = 0$.
2. $P(A) \geq 0$ for all $A \in \mathcal{F}$.
3. If $A_1, A_2, \ldots \subset \mathcal{F}$ such that $A_k \cap A_l = \emptyset$ for all $k \neq l$, it follows that $P(\bigcup_n A_n) = \sum_n P(A_n)$.

If P maps to $[0, 1]$ and $P(\Omega) = 1$, it is called a probability measure.

The tuple $(\Omega, \mathcal{F})$ from Definition 6.1 is called a measurable space. The triple $(\Omega, \mathcal{F}, P)$ is called a measure space. If P is additionally a probability measure, the triple is called a probability space.

Definition 6.2 (Random variable)**.** Let $(\Omega, \mathcal{F}, P)$ be a probability space and $(E, \mathcal{E})$ a measurable space. A mapping $X \colon \Omega \to E$ is a random variable, if

$$\{\omega \in \Omega | X(\omega) \in B\} \in \mathcal{F} \text{ for all } B \in \mathcal{E}.$$

When Ω and E are countable, the σ-algebras $\mathcal{F}$ and $\mathcal{E}$ can be assumed to be the power sets over Ω and E, respectively.

In probability theory, the expression $\{\omega \in \Omega | X(\omega) \in B\}$ is often abbreviated to $\{X \in B\}$.

Definition 6.3 (Probability space of a random variable)**.** Let $(\Omega, \mathcal{F}, P)$ be a probability space, $(E, \mathcal{E})$ a measurable space and $X \colon \Omega \to E$ a random variable. Then $P_X(B) := P(\{X \in B\}), B \in \mathcal{E}$ defines a probability measure P_X on $(E, \mathcal{E})$.

The expression $P(\{X \in B\})$ is usually shorthanded to $P(X \in B)$.

Definition 6.4 (Distribution of a random variable)**.** The probability measure P_X is called the distribution of the random variable X.

The probability space $(E, \mathcal{E}, P_X)$ is called discrete, if E is countable.

Definition 6.5 (Independence of random variables)**.** Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $X_1, \ldots, X_n$ be n random variables (over Ω) with values in the measurable spaces $(E_i, \mathcal{E}_i)$, $i \in \{1, \ldots, n\}$. The random variables $X_1, \ldots, X_n$ are independent when

$$P(X_1 \in B_1, \ldots, X_n \in B_n) = \prod_{i=1}^{n} P_{X_i}(B_i) \text{ for } B_i \in \mathcal{E}_i.$$

As shorthand notation we write $X \perp\!\!\!\perp Y$ when X and Y are independent.

Definition 6.6 (Conditional independence of variables)**.** Let $(\Omega, \mathcal{F}, P)$ be a probability space and let X, Y, Z be random variables (over Ω) with values in the measurable spaces $(E_i, \mathcal{E}_i)$, $i \in \{X, Y, Z\}$. Two random variables X and Y are conditionally independent given the variable Z if

$$P(X \in A, Y \in B | Z \in C) = P(X \in A | Z \in C) P(Y \in B | Z \in C).$$

for each $A \in \mathcal{E}_X, B \in \mathcal{E}_Y$ and $C \in \mathcal{E}_Z$.

As shorthand we write $P(X, Y|Z) = P(X|Z) \cdot P(Y|Z)$, $(X \perp\!\!\!\perp Y)|Z$ or $X \perp\!\!\!\perp Y|Z$. Note that the definition is equivalent to $P(X|Y, Z) = P(X|Z)$.

Theorem 6.1 (Law of total probability)**.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space, A and C events and let $B_1, \ldots, B_n$ be a partition in that probability space. Then*

$$P(A|C) = \sum_{i=1}^{n} P(A|B_i, C) P(B_i|C).$$

The law of total probability also holds for continuous random variables X, and Y with positive density functions f_X and f_Y, respectively.

$$f_Y(y) = \int_{-\infty}^{\infty} f_Y(y|X = x) \cdot f_X(x) \, \mathrm{d}x.$$

Theorem 6.2 (Bayes' law for conditional probabilities)**.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space and B and C events and let $A_1, \ldots, A_n$ be a partition in that probability space. Then*

$$P(A_j|B, C) = \frac{P(B|A_j, C) P(A_j|C)}{P(B|C)} = \frac{P(B|A_j, C) P(A_j|C)}{\sum_{i=1}^{n} P(B|A_i, C) P(A_i|C)}.$$

Note that in this form *Bayes' theorem* also holds for variables (instead of events). This is true for discrete random variables, continuous random variables as well as a mixture of discrete and continuous random variables. If continuous variables are involved, they need to have a positive density function.

Theorem 6.3 (Product distribution)**.** *Let X and Y be two independent continuous random variables, with positive probability density functions $f(x)$ and $g(x)$. Then the random variable U, with $U = X \cdot Y$, is a continuous random variable with probability density function h, with*

$$h(u) = \int_{-\infty}^{\infty} \frac{1}{|y|} \cdot f(\frac{u}{y}) \cdot g(y)\,\mathrm{d}y.$$

We call the distribution of U a product distribution.

An important distribution we refer to in the next sections is the beta distribution. More information about the distribution can be found in [JKB95].

Definition 6.7 (Beta distribution)**.** A *beta distribution* is a family of continuous probability distributions in the interval $[0, 1]$, parameterised by two positive parameters, $\alpha, \beta \geq 1$. The probability density function of a beta distribution with parameters α and β is

$$f_{\mathrm{B}}(x; \alpha, \beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\int_0^1 y^{\alpha-1}(1-y)^{\beta-1}\,\mathrm{d}y} \propto x^{\alpha-1}(1-x)^{\beta-1}.$$

The expression under the fractions is known as the beta function on α and β, and for positive integers α and β, the beta function fulfills $\mathrm{B}(\alpha, \beta) = \frac{(\alpha-1)!(\beta-1)!}{(\alpha+\beta-1)!}$.

The expected value of a beta distribution is well-known and simple:

Proposition 6.4. *The expected value of a beta distribution $f_{\mathrm{B}}(x; \alpha, \beta)$ is given* $\mathbf{E}(f_{\mathrm{B}}(x; \alpha, \beta)) = \frac{\alpha}{\alpha+\beta}$.

Note that this implies that $\mathbf{E}(\vartheta_{s,f}(x)) = \mathbf{E}(x^s \cdot (1-x)^f \cdot \mathrm{NF}) = \frac{s+1}{s+f+2}$, since $s = \alpha - 1$ and $f = \beta - 1$. Since we usually refer to beta distributions by the number of successes and failures, the latter format is used more often.

6.2 Formalisation

In this section, we formalise the assumptions that we have for trust in a system based on asymmetric interactions (like transactions in e-commerce systems), where expectations are clearly defined. First, we informally introduce our assumptions with motivations, illustrate it with an example, and then formally state the assumptions as relations between random variables. Finally, we look at the formal properties of the Beta model.

To reiterate some assertions of the Beta paradigm (see Section 2.4): Interactions are the building blocks in our trust analysis. Interactions are between a *subject* and a *target*. A subject forms a trust opinion about a target, before the subject interacts

with the target. The observed behaviour of the active party is objectively classified as successful (a *success*) or failed (a *failure*). Furthermore, the probability that the active party behaves well is determined by its *integrity* parameter p. A *user* will most likely exhibit non-probabilistic behaviour, and will therefore behave well in some situations and badly in others. However, we do not know the correlation between situations and behaviours, nor do we necessarily know the situation. In the light of this, we can view the integrity p as the chance that a user is in a situation where his behaviour is successful (or even where behaving well is in his best interest in some iterative game[1], as in [MRS03]). Lastly, we assume that p neither changes over time nor with respect to the environment. This assumption allows us to treat previous interactions in a mathematically coherent way, since all interactions are equally relevant for the current situation. In the model, a user will never know the integrity of another user, but will have an estimate based on these previous interactions.

Throughout Part II, we use a running example, introduced below, to illustrate the techniques we use. Recall that direct applicability to existing trust systems in practice is not one of the goals of this thesis. Hence, the running example is selected on the basis of effectively communicating underlaying intuitions, not on the basis of direct applicability to practice.

Running Example. An economy teacher wants to teach her students about e-commerce with the help of a turn-based game. To set up the game, the teacher secretly distributes a random value $p_i \in [0, 1]$ to each student c_i for $1 \leq i \leq 30$. The value p_i represents the integrity of each student, and, similar to the integrity of users on an e-commerce website, it is unknown to the other players. On an e-commerce system this parameter models how likely the outcome of an interaction is to be successful. Each turn of the game follows the following pattern. First, in the turn of student c_i (the subject), the teacher assigns another student c_j (the target) to c_i. Then, c_i has the choice between trusting or not trusting c_j. In case c_i chooses to trust c_j, c_i gains two points with probability p_j, i.e. with the probability corresponding to the other student's integrity parameter. With the remaining probability of $1-p_j$, c_i loses one point. If c_i chooses not to trust c_j, then he neither gains nor loses points. On an e-commerce platform winning points corresponds to a successful interaction (a success), losing points to a failed interaction (a failure). After every turn, the teacher updates the students' points, only revealing the outcome to c_i. Like in e-commerce, trusting someone with high integrity has a high probability to result in a successful interaction; trusting someone with a low integrity has a high probability to result in an unsuccessful interaction.

To formulate the assumptions of the Beta paradigm in a formal manner, we need to define interactions of users, integrity parameters of users, sets of interactions that users made in the past, and composite targets. The outcomes of interactions can be a success, denoted S, or a failure, denoted F. We are often interested in the previous interactions between a subject and a target, which we call an *interaction history* of the subject about the target. Furthermore, we take an interaction history to be a pair of natural numbers: the first number as representing the number of successes, the second number as representing the number of failures.

[1] Users expect to interact multiple times with other users, and even if betrayal is profitable on the short run, it may be more profitable to conform on the long run.

We first define a series of random variables. Let $\mathbf{A}$ denote a set of users. For $A, C \in \mathbf{A}$ and a set of events Ω, we then define:

- $E_C \colon \Omega \to \{\text{S}, \text{F}\}$ is a discrete random variable modelling the outcome of the corresponding interaction with target C.
- $R_C \colon \Omega \to [0,1]$ is a continuous random variable modelling the (hidden) *integrity parameter* of target C which defines the probability of success.
- $O_C^A \colon \Omega \to \mathbb{N} \times \mathbb{N}$ is a discrete random variable modelling the *interaction history* of subject A about target C, representing the past interactions (number of successes and failures) between A as passive party and C as active party.

Running Example. In the classroom game, E_C models the outcome of an interaction with student C. The variable R_C describes the secret parameter initially assigned by the teacher to C and O_C^A expresses how many times student A interacted successfully with student C, and how many times it was a failure.

A *trust opinion* is a distribution over the integrity parameter of a target, based on the interaction history about the involved active parties. Hence, if a subject A establishes a trust opinion about a target C, the probability density function is of the form $f_{R_C}(x|O_C^A, \varphi)$, where φ may express additional conditions.

Next, we provide the assumptions of the Beta model, in the shape of dependencies and independencies of random variables, as we have formulated in [MS13a]. For a more concise formulation of the (in)dependencies, we introduce sets of random variables, again for $A \in \mathbf{A}$:

$$\begin{aligned}\mathbb{E} &:= \{E_C : C \in \mathbf{A}\},\\ \mathbb{R} &:= \{R_C : C \in \mathbf{A}\},\\ \mathbb{O} &:= \{O_C^A : A, C \in \mathbf{A}\},\\ \mathbb{W} &:= \mathbb{E} \cup \mathbb{R} \cup \mathbb{O}.\end{aligned}$$

The size of the interaction histories is unknown. We therefore model it with a distribution λ, called the *entanglement*. Hence, $\lambda(4)$ is the probability that a particular subject has 4 interactions with a particular target. Let $c \in [0,1]$, $x_s, x_f \in \mathbb{N}$ and $\lambda \colon \mathbb{N} \to [0,1]$ be a probability distribution. For all users $A, C \in \mathbf{A}$ we set up the following dependency and independency relations as our assumptions:

D1 R_C is uniformly distributed on $[0,1]$.
If we know nothing about the integrity of C, we assert all values equally likely according to the principle of maximal entropy [Jay57]. The choice of the distribution of R_C is fairly inconsequential, as discussed in Remark 6.1, below.

D2 $P(E_C{=}\text{S} | R_C{=}c) = c$.
We assume that the probability of good behaviour of C is determined by an integrity parameter c.

D3 $P(O_C^A{=}(x_s, x_f) | R_C{=}c) = \binom{x_s+x_f}{x_s} c^{x_s}(1-c)^{x_f}\lambda(x_s + x_f)$.
We assume that the probability of A having x_s successes and x_f failures with

C is equal to the probability of A having x_s+x_f past interactions with C, and that the fraction x_s of successes is binomially distributed with rate c. That it is binomial implies that each past interaction was determined to be a success of failure with fixed probability c, independently of other interactions.

I1 For $W \in \mathbb{W}\backslash\{O_C^A\}$, it holds that $O_C^A \perp\!\!\!\perp W|R_C$.
The interaction history is completely determined by its size, and the probability of a success in a single interaction (by Dependency D3).

I2 For $W \in \mathbb{W}\backslash\{R_C\}$ and $\{C, D_0, \dots D_n\} = \mathbf{A}$, $R_C \perp\!\!\!\perp W|E_C, O_C^{D_0}, \dots, O_C^{D_n}$.
The only indicators of the integrity parameter of C, are interactions with it. That is, the collection of all interaction histories of arbitrary users with C, and most recent interaction E_C.

I3 For $W \in \mathbb{W}\backslash\{E_C\}$, it holds that $E_C \perp\!\!\!\perp W|R_C$.
The behaviour of C is completely determined by its integrity parameter (by Dependency D2).

Note that Independency I1 is necessary to be able to work with Dependency D3, when provided a term such as $P(O_C^A|R_C, O_D^B)$. To derive $P(O_C^A|R_C, O_D^B) = \binom{x_s+x_f}{x_s} \cdot c^{x_s} \cdot (1-c)^{x_f} \cdot \lambda(x_s+x_f)$, Dependency D3 is not sufficient by itself. Similar reasoning applies to the necessity of Independency I3, in the presence of Dependency D2.

We explicitly do not define the Beta model using the notion of beta distributions. The foundation of the Beta model is formed by Dependencies D1-D3 and Independencies I1-I3. The link to beta distributions is formally derived later, in Theorem 6.5 and Lemma 6.6.

Definition 6.8 (Beta model)**.** A trust model is said to be the *Beta model*, when it satisfies Dependencies D1-D3 and Independencies I1-I3.

Remark 6.1. The choice of the prior distribution in Dependency D1 can be altered without needing to overhaul the models and the conclusions drawn about them. Let $f(x)$ be a trust opinion about C in the Beta model, the trust opinion about C with an alternative prior distribution $g(x)$ (with support on $(0, 1)$) is proportional to $f(x) \cdot g(x)$. Hence, given an alternative prior $g(x)$, our model remains correct modulo a multiplicative factor $g(x)$. In practice, survey data can be used to establish a reasonable prior.

A trust opinion of A about C can now be seen as the probability density function given by $f_{R_C}(c|\varphi)$, where φ represents all knowledge of A about C, modulo the relations of the random variables. In the Beta model, φ is equal to O_C^A, for subject A and target C. In this case, we call $f_{R_C}(c|\varphi)$ a *simple trust opinion*, to be able to distinguish it from trust opinions involving *recommendations* (*chained trust opinions*) and trust opinions involving logical trust operations (*composite trust opinions*). Recall that for arbitrary A, C, we may shorthand the simple trust opinion $f_{R_C}(c|O_C^A{=}(s, f))$ to $\vartheta_{s,f}(c)$.

The Beta model is based on beta distributions [MS13a].

Theorem 6.5. *The simple trust opinion obtained from an interaction history with x_s successes and x_f failures, $\vartheta_{x_s,x_f}(c) = f_{R_C}(c|O_C^A{=}(x_s, x_f))$, is the beta distribution $f_{\mathrm{B}}(c; x_s + 1, x_f + 1)$.*

Proof.

$$
\begin{aligned}
& f_{R_B}(x|O_B^A{=}(m,n)) \\
=& \frac{P(O_B^A{=}(m,n)|R_B{=}x)\cdot f_{R_B}(x)}{\int_0^1 P(O_B^A{=}(m,n)|R_B{=}x')\cdot f_{R_B}(x')\,\mathrm{d}x'} \\
=& \frac{\binom{m+n}{m}x^m(1-x)^n\lambda(m+n)\cdot f_{R_B}(x)}{\int_0^1 \binom{m+n}{m}(x')^m(1-x')^n\lambda(m+n)\cdot f_{R_B}(x')\,\mathrm{d}x'} \\
=& \frac{x^m(1-x)^n}{\int_0^1 (x')^m(1-x')^n\,\mathrm{d}x'} \\
=& f_{\mathrm{B}}(x;m+1,n+1).
\end{aligned}
$$

□

Suppose there are two concurrently held trust opinions based on two different interactions with a single user. It is desirable to combine these two trust opinions into a single trust opinion based on both interactions. We introduce a trust aggregation operator to accomplish that:

Definition 6.9 (Aggregation of trust opinions)**.** The aggregation of trust opinion $f(c)$ and $g(c)$ is $\frac{f(c)\cdot g(c)}{\int_0^1 f(x)\cdot g(x)\,\mathrm{d}x} \propto f(c)\cdot g(c)$.

The trust aggregation operator correctly combines simple trust opinions:

Lemma 6.6. *Given trust opinions $f = \vartheta_{x_s,x_f}$ and $g = \vartheta_{y_s,y_f}$, the aggregate trust opinion $\frac{f(c)\cdot g(c)}{\int_0^1 f(x)\cdot g(x)\,\mathrm{d}x}$ is equal to $\vartheta_{x_s+y_s,x_f+y_f}$.*

Proof. Observe the following proportionalities:
$\frac{f(c)\cdot g(c)}{\int_0^1 f(x)\cdot g(x)\,\mathrm{d}x} \propto f_{\mathrm{B}}(c;x_s{+}1,x_f{+}1)\cdot f_{\mathrm{B}}(c;y_s{+}1,y_f{+}1) \propto f_{\mathrm{B}}(c;x_s{+}y_s{+}1,x_f{+}y_f{+}1)$
Since the left hand side and the right hand side are distributions, they are equal. □

Our assumptions regarding simple trust opinions are in line with the Beta model. They are in fact sufficient to derive it (Theorem 6.5). Hence, those assumptions can be seen as valid for the numerous models that use the Beta model as a foundation [Jøs97, TPJL06, Rie07].

6.3 Conclusion

The Beta model exists in many forms and formalisations. The view that trust opinions can be treated as formal entities with a specific probabilistic notion (rather than cognitive entities with a probabilistic approximation) is relatively new. The formalisation we provide in this chapter is unique in the sense that the separation between the semantics of trust opinions and trust aggregation and the assumptions on the domain is formulated so explicitly. In particular, the assertions on the domain (in the form of relations between random variables) are set up in such a way that the computations for simple trust opinions and trust aggregation follow as theorems (Theorem 6.5 and Lemma 6.6). Due to the clear separation between assumptions and computations, any disagreement on a particular computed trust opinion should be traced back to a disagreement on an assumption.

The main advantage of this set-up is that the other operations we are interested in (trust chaining, trust conjunction, trust disjunction and trust negation) can be derived in a similar fashion. In particular, models with these operations are conservative extensions of the Beta model with regards to the relations between the random variables, meaning that all dependencies and independencies remain valid. In the next two chapters, we will do exactly that. We introduce new random variables in order to define the other operations.

7

The Beta Model with Logical Trust Operations

In the previous chapter, we defined the Beta model. The Beta model reasons about *simple trust opinions* – trust opinions based on direct experience with the target. It is possible for a subject to make an assessment about the future behaviour of several targets, then such an assessment is called a *composite trust opinion*. In this chapter we extend the Beta model with composite trust opinions. We call that model the *Beta model with logical trust operations*.

In an interaction in the setting of the Beta model with *logical trust operations*, there are several parties that have an agreement. There is at least one target and there is one subject. Each target determines his own outcome to be a success or failure. Since the subject may be harmed if one of the targets fails, the subject forms a trust opinion about each of the targets before (potentially) interacting. To express composite trust opinions we denote the target in propositional logic, where atomic propositions represent successes or failures of individual targets. A target $A \wedge B$ succeeds iff both A and B succeed, similarly a target $A \vee B$ succeeds iff at least A or B succeeds. To illustrate the use of composite trust, consider the following example.

Example 7.1. Take an imaginary web service, CLOUD, that offers computational power to *users*, by CPU-scavenging in a similar fashion to BOINC [And04], i.e. CLOUD is a grid. A user that delegates a computation is a client, and a user that offers CPU cycles is a provider. Unlike BOINC, the CLOUD system is a commercial system, where clients pay for computations, and providers get paid for offering computational power.

The identity of the machines in CLOUD is public, and users can delegate computations to specific (groups of) machines. The infrastructure of CLOUD is open, which means that malicious users can easily join as a provider. Malicious users may sometimes take shortcuts in the computation, providing wrong results. Furthermore, non-malicious users may prematurely terminate a computation before a result is provided, for example, when the computer shuts down, restarts or drops the network. It may occur that a single computation is delegated to a group of computers working concurrently to reduce latency, each computer solving a part. It may also occur that a single computation is delegated to more than one (group of) provider(s), to still receive an answer when one of the (groups of) providers fails.

Now, suppose that a client, A, on CLOUD has an instance of an NP-complete problem, and sends the problem to a provider, D, and a copy of the problem to a

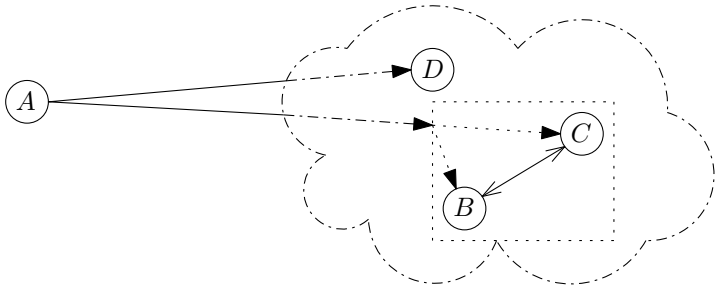

Figure 7.1: The two outgoing arrows from A are delegations for a computation. One for D, the other is split and runs concurrently on B and C.

pair of concurrent providers, B and C. See Figure 7.1 for a visual representation of the interaction.

In our terminology, clients are subjects, and providers are potential targets. A provider is successful if he delivers a correct result within a specified time frame. A provider fails if he returns a wrong result, returns it too late, or not at all. Since a client can quickly verify a (positive) solution to an NP-complete problem, correct and incorrect solutions can easily be distinguished. Hence, it suffices for the subject A to receive at least one correct result within the specified time frame from a target. If either the single provider D or both other providers B and C provide the correct result in time, the whole target's behaviour is considered good. We can denote this composite trust opinion as $D \vee (B \wedge C)$.

The subject not only wants to know the probability that the target succeeds, but also the uncertainty. If the probabilities b, c and d of B, C and D succeeding are independent, then one may anticipate that the expected probability of the target succeeding to be $d + b \cdot c - d \cdot b \cdot c$. We formally show the foresight on this trust opinion to be correct in Section 7.2.

Cloud computing is not the only practical trust systems where logical trust operations are relevant. In particular, trust conjunction is often relevant, as many e-commerce and e-service transactions involve several parties in such a way that the total transaction fails if any of the involved parties fail. We treat the logical trust operations as abstract entities, defined in probability theory, rather than the concrete relations in trust systems. To achieve this, we add new random variables to the Beta model, and add additional assumptions on these random variables, in Section 7.1

7.1 Formalisation

The formalisation of the Beta model with logical trust operations is similar to the formalisation of the Beta model in Section 6.2. The important differences are that targets are not necessarily users, but can be more complicated. This requires us to define extra random variables, on top of the random variables in the Beta model, namely R_T and E_T for such targets T (e.g. $T = B \vee C$) that are not users. The reason is that although a target T is not a user, we are still interested in forming a trust opinion on T, which requires R_T, and thus E_T. We add dependencies

specifically for the new random variables, and update the independencies from the Beta model to reflect the new random variables.

Before we introduce the formalisation, we extend the running example of the classroom game:

Running Example. After a number of turns, the students realise that the Beta model can be applied to construct a correct trust opinion about other students. Hence, the students make optimal choices. To keep the game interesting, the teacher introduces composition of targets in the following way: In the beginning of every turn, the teacher still assigns a subject $c_i \in S$, but assigns one or more targets $C \subseteq S \setminus \{c_i\}$. The teacher defines a propositional formula on top of these targets, say $c_j \wedge \neg c_k$. In that case, student c_i only gains the two points if c_j succeeds and c_k fails, and loses the point if either c_j fails or c_k succeeds. For the subject c_i to create the correct trust opinion, he needs to incorporate his opinion of all the targets in C; in the example $\{c_j, c_k\}$.

First we define the random variables. The set of users is again $\mathbf{A}$. The targets $\mathbf{T}$ are defined by $\varphi ::= A \,|\, \varphi \wedge \varphi \,|\, \varphi \vee \varphi \,|\, \neg\varphi$, for $A \in \mathbf{A}$. For $A, B \in \mathbf{A}$, $T \in \mathbf{T}$ and a set of events Ω, we define the following random variables.

- $E_T \colon \Omega \to \{\text{S}, \text{F}\}$ is a discrete random variable modelling the outcome of the corresponding interaction with target T.
- $R_T \colon \Omega \to [0,1]$ is a continuous random variable modelling the (hidden) integrity parameter of target T, defining the probability of success.
- $O_B^A \colon \Omega \to \mathbb{N} \times \mathbb{N}$ is a discrete random variable modelling the interaction history of A about B, representing the past interactions between A as passive party and B as active party.

It may be useful to be able to reason about the users that comprise the target. We introduce $\text{act}(\_)$, such that $\text{act}(T)$ is the set of users in T. For example, if $T = D \vee (B \wedge C)$, then $\text{act}(T) = \{B, C, D\}$.

For a more concise formulation of these (in)dependencies, we introduce sets of random variables.

$$\begin{aligned} \mathbb{E}^+ &:= \{E_T : T \in \mathbf{T}\}, \\ \mathbb{R}^+ &:= \{R_T : T \in \mathbf{T}\}, \\ \mathbb{O} &:= \{O_B^A : A, B \in \mathbf{A}\}, \\ \mathbb{W}^+ &:= \mathbb{E}^+ \cup \mathbb{R}^+ \cup \mathbb{O}. \end{aligned}$$

Note that $\mathbf{A} \subset \mathbf{T}$, so $\mathbb{E} \subset \mathbb{E}^+$, $\mathbb{R}^+ \subset \mathbb{R}$ and $\mathbb{W}^+ \subset \mathbb{W}$.

Let $x \in [0,1]$, $n, k \in \mathbb{N}$ and $\lambda \colon \mathbb{N} \to [0,1]$ be a probability distribution. For all $A, B \in \mathbf{A}$ and $S, T \in \mathbf{T}$ we set up the following dependency relations as our assumptions. Note that Dependencies D1_ℓ and D3_ℓ are the same as Dependencies D1 and D3 in Section 6.2:

D1_ℓ R_C is uniformly distributed on $[0,1]$.

D2_ℓ $P(E_T{=}\text{S} | R_T{=}p) = p$.
Similar to Dependency D2 from Section 6.2, extended to integrity and outcome of targets from $\mathbf{T}$, rather than targets from $\mathbf{A}$.

D3$_\ell$ $P(O_C^A{=}(x_s, x_f)|R_C{=}c) = \binom{x_s+x_f}{x_s} c^{x_s}(1-c)^{x_f}\lambda(x_s+x_f)$.

D4$_\ell$ $E_{S\wedge T} = \text{S}$ iff $E_S = \text{S}$ and $E_T = \text{S}$, for $\text{act}(S) \cap \text{act}(T) = \emptyset$.
We define conjunction of independent targets in such a way that the conjunction succeeds if both targets succeed.

D5$_\ell$ $E_{S\vee T} = \text{S}$ iff $E_S = \text{S}$ or $E_T = \text{S}$, for $\text{act}(S) \cap \text{act}(T) = \emptyset$.
We define disjunction of independent targets in such a way that the disjunction succeeds if at least one target succeeds.

D6$_\ell$ $E_{\neg T} = \text{S}$ iff $E_T = \text{F}$.
We define negation of a target in such a way that the negation of the target succeeds if the original target fails.

D7$_\ell$ There exist functions f, g, h, with $R_{S\wedge T} = f(R_S, R_T)$ and $R_{S\vee T} = g(R_S, R_T)$ when $\text{act}(S) \cap \text{act}(T) = \emptyset$, and $R_{\neg T} = h(R_T)$.
We assert that the integrity of a composite target is determined by the integrity of its constituent users.

I1$_\ell$ For $W \in \mathbb{W}^+ \backslash \{O_B^A\}$, it holds that $O_B^A \perp\!\!\!\perp W|R_B$.

I2$_\ell$ For $W \in \mathbb{W}^+ \backslash \{R_S : B \in \text{act}(S), R_S \in \mathbb{R}^+\}$ and $\{C, D_0, \ldots D_n\} = \mathbf{A}$, it holds that $R_C \perp\!\!\!\perp W|E_C, O_C^{D_0}, \ldots, O_C^{D_n}$.
Similar to Independency I2, however, the integrity of a user B is not independent of the integrity of a target S containing that user. For example, if $R_B = 0.2$, then $E_{B\wedge C} < 0.2$. If the user B is not a constituent of the target S, then their integrities are independent.

I3$_\ell$ For $W \in \mathbb{W}^+ \backslash \{E_S : A \in \text{act}(S), E_S \in \mathbb{E}^+\}$, it holds that $E_A \perp\!\!\!\perp W|R_A$.
Similar to Independency I3, however, the outcome of an interaction with the user A is not independent of the outcome of an interaction with a target S containing that user. For example, if $E_{A\wedge B} = \text{S}$, then $E_A = \text{S}$. If the user A is not a constituent of the target S, then the outcomes are independent.

Independency I3$_\ell$ can be generalised for composite targets.

Proposition 7.1. *For all* $W \in \mathbb{W}^+ \backslash \{E_S : \text{act}(T) \cap \text{act}(S) \neq \emptyset, R_S \in \mathbb{E}^+\}$*, it holds that* $E_T \perp\!\!\!\perp W|R_T$.

Proof. Apply structural induction. The base case precisely matches Independency I3$_\ell$. For the induction step use that, by definition of act(_), it holds that $\text{act}(T) \cup \text{act}(T') = \text{act}(T \wedge T') = \text{act}(T \vee T')$. □

The Beta model with logical trust operations satisfies Dependencies D1-D7$_\ell$ and Independencies I1$_\ell$-I3$_\ell$.

Definition 7.1 (Beta model with logical trust operations)**.** A trust model is said to be the *Beta model with logical trust operations*, when it satisfies Dependencies D1-D7$_\ell$ and Independencies I1$_\ell$-I3$_\ell$.

A trust opinion of A about T can now be seen as the probability density function given by $f_{R_T}(x|\varphi)$, where φ is a condition that represents all knowledge of A about all the constituents of T. In other words, the trust opinion of A about T is $f_{R_T}(x|O_B^A, O_C^A, \dots)$ for $B, C, \dots \in \mathrm{act}(T)$.

Our assumptions regarding simple trust opinions are in line with the beta model. Hence, those assumptions can be seen as valid for the numerous models based on the beta model [Jøs97, TPJL06, Rie07]. We extend the assumptions beyond simple trust opinions, by adding assumptions about composite trust opinions (Dependencies D4$_\ell$, D5$_\ell$ and D7$_\ell$). The additional dependencies are taken to be self-evident within the setting of the Beta paradigm. We see the four dependencies as a definition of *trust conjunction*, *trust disjunction* and *trust negation*. Under these assumptions, we show in Theorem 7.10 that composite trust opinions cannot generally be represented as beta distributions.

7.2 Composite Trust

In Example 7.1, we introduced the CLOUD grid. An example of a composite target was $D \vee (B \wedge C)$, where B, C and D are providers. The subject, A, has a (potentially empty) *interaction history* about B, C and D. In Example 7.2, we formally derive the trust opinion of A.

Example 7.2. The subject wants to form a trust opinion about $D \vee (B \wedge C)$, using only the interaction history of A about users B, C and D. The random variables O_B^A, O_C^A and O_D^A represent the interaction history of A about B, C and D. The random variable $R_{D\vee(B\wedge C)}$ represents the (unknown) integrity parameter of the target $D \vee (B \wedge C)$, and the random variable $E_{D\vee(B\wedge C)}$ represents the (unknown) outcomes of the next interaction with the target $D \vee (B \wedge C)$. We are interested not just in the probability that the next outcome of the target is a success ($E_{D\vee(B\wedge C)}$), but also in additional information, i.e. the random variable $R_{D\vee(B\wedge C)}$. Figure 7.2 depicts the relation between the users and the involved random variables. As stated in Section 7.1, given failures and successes of past interactions $(b_\mathrm{S}, b_\mathrm{F}, c_\mathrm{S}, c_\mathrm{F}, d_\mathrm{S}, d_\mathrm{F})$, the query for the trust opinion is of the shape $f_{R_{D\vee(B\wedge C)}}(x|O_B^A = (b_\mathrm{S}, b_\mathrm{F}), O_C^A = (c_\mathrm{S}, c_\mathrm{F}), O_D^A = (d_\mathrm{S}, d_\mathrm{F}))$. In other words, the trust opinion represents the probability distribution of a random variable that predicts the probability that the target succeeds.

Whenever a subject wants to compute a composite trust opinion about a target, he chooses the correct conditions and the correct random variable to form a distribution over, as illustrated in Example 7.2. Therefore, we can assume, without loss of generality, that we are given the term representing the probability distribution, and we want to compute an explicit probability density function.

We are interested in a random variable R_T, where T is not a single user (unless the subject wants a simple trust opinion). However, we have not provided direct relations between R_T and observation histories O_B^A or integrity parameters of single users R_A. The only random variable that we can immediately relate R_T to is E_T.

For more concise notation, we note the following lemma, based on the product distribution:

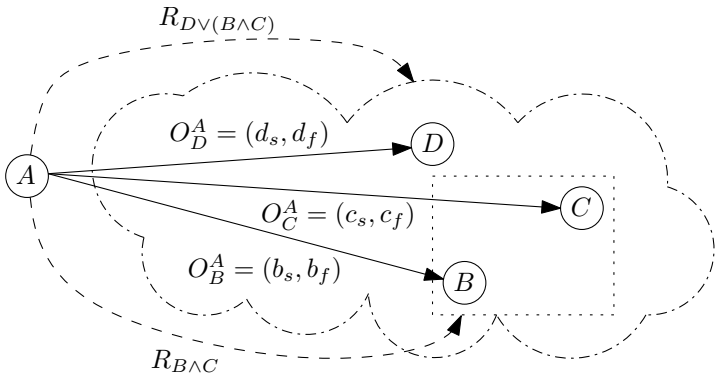

Figure 7.2: Solid arrows represent interaction histories. Dashed arrows represent composite trust opinions. Arrows are labelled with the relevant random variables.

Lemma 7.2. *If S and T do not share any users, then $R_{S\wedge T} = R_S \cdot R_T$.*

Proof. The product $R_S \cdot R_T$ of two random variables is defined as $(R_S \cdot R_T)(\omega) := R_S(\omega) \cdot R_T(\omega)$.

By Dependency D2, for all x, it holds that

$$P(E_{S\wedge T} = \mathrm{s} | R_{S\wedge T} = x) = x.$$

And, using Proposition 7.1 as well as Dependencies D2 and D4$_\ell$ we obtain

$$\begin{aligned}
&P(E_{S\wedge T} | R_S = y, R_T = z) \\
=&P(E_S = \mathrm{s}, E_T = \mathrm{s} | R_S = y, R_T = z) \\
=&P(E_S = \mathrm{s} | E_T = \mathrm{s}, R_S = y, R_T = z) \cdot P(E_T = \mathrm{s} | R_S = y, R_T = z) \\
=&P(E_S = \mathrm{s} | R_S = y) \cdot P(E_T = \mathrm{s} | R_T = z) \\
=&y \cdot z.
\end{aligned}$$

Assume, without loss of generality, that $R_S(\omega) = y$ and $R_T(\omega) = z$. By Dependency D7$_\ell$, there is a function f such that $x = P(E_{S\wedge T} = \mathrm{s} | R_{S\wedge T} = x) = P(E_{S\wedge T} = \mathrm{s} | f(R_S, R_T) = x)$. That implies that $x = f(y, z)$, and thus $P(E_{S\wedge T} = \mathrm{s} | f(R_S, R_T) = f(y, z)) = f(y, z)$. Now, since $R_S(\omega) = y$ and $R_T(\omega) = z$, we have

$$\begin{aligned}
&f(y, z) \\
=&P(E_{S\wedge T} = \mathrm{s} | f(R_S, R_T) = f(y, z)) = f(y, z) \\
=&P(E_{S\wedge T}) \\
=&P(E_{S\wedge T} | R_S = y, R_T = z) \\
=&y \cdot z.
\end{aligned}$$

Thus $R_S \cdot R_T = f(R_S, R_T) = R_{S\wedge T}$. □

A similar proof exists for $R_{S\vee T} = R_S + R_T - R_S \cdot R_T$, using independency over union rather than intersection, and $R_{\neg T} = 1 - R_T$.

It is no coincidence that, say $R_{S\vee T}$ is shaped $R_S + R_T - R_S \cdot R_T$, where $P(E_S \cup E_T | R_S, R_T) = P(E_S | R_S) + P(E_T | R_T) - P(E_S | R_S) \cdot P(E_T | R_T)$; the same computation as for the associated random variable of the integrity parameter. Of course, this generalises over more logical operations than just conjunction, disjunction and negation:

Remark 7.1. We have only allowed composition via conjunction, disjunction and negation. There is, however, no formal reason not to introduce arbitrary logical operators, such as the binary xor, or the ternary if-then-else or "majority in three". Although our technical results are limited to conjunction, disjunction and negation, our methodology is certainly not.

To have an additional logical operator, we merely need to add dependencies, that define when $E_{X(S_0,S_1,S_2)}$ holds, and that $R_{X(S_0,S_1,S_2)}$ is functionally dependent on R_{S_0}, R_{S_1} and R_{S_2}. We can assign a probability x to each row in the truth table of X, e.g. $x(S_0 = \text{S}, S_1 = \text{S}, S_2 = \text{F}) = R_{S_0} \cdot R_{S_1} \cdot (1 - R_{S_2})$. Now, $R_{X(S_0,S_1,S_2)}$ is equal to the sum of all $x(S_0 = s_0, S_1 = s_1, S_2 = s_2)$, such that $X(s_0, s_1, s_2)$ holds. Note that we can apply algebra of random variables to transform the expression[Spr79].

To work out some examples, in the truth table of "or", $X(\text{S}, \text{F})$, $X(\text{F}, \text{S})$ and $X(\text{S}, \text{S})$ all hold, so $R_{S \vee T} = R_S \cdot (1 - R_T) + (1 - R_S) \cdot R_T + R_S \cdot R_T$, which, via algebra of random variables, equals $R_S + (1 - R_S) \cdot R_T = R_S + R_T - R_S \cdot R_T$. Similarly, for if-then-else, denoted $S_1 \lhd S_0 \rhd S_2$, becomes $R_{S_1 \lhd S_0 \rhd S_2} = R_{S_0} \cdot R_{S_1} + (1 - R_{S_0}) \cdot R_{S_2}$. Finally, for "majority in three", denoted M_3, becomes $R_{M_3(S_0,S_1,S_2)} = R_{S_0} \cdot R_{S_1} \cdot (1 - R_{S_2}) + R_{S_0} \cdot (1 - R_{S_1}) \cdot R_{S_2} + (1 - R_{S_0}) \cdot R_{S_1} \cdot R_{S_2} + R_{S_0} \cdot R_{S_1} \cdot R_{S_2}$.

The De Morgan rules hold for the logical trust operations.

Proposition 7.3. *The random variables $R_{S \vee T}$ and $R_{\neg(\neg S \wedge \neg T)}$ are equal and the random variables $R_{S \wedge T}$ and $R_{\neg(\neg S \vee \neg T)}$ are equal.*

Proof. It suffices to perform some basic algebra on the random variables (see e.g. [Spr79]) as:

$$R_{S \vee T} = R_S + R_T - R_S \cdot R_T = 1 - (1 - R_S) \cdot (1 - R_T) = R_{\neg(\neg S \wedge \neg T)}$$

□

We can derive the probability density function of $R_{S \wedge T}$, for independent S and T, under any condition φ.

Theorem 7.4. *If S and T do not share any users, then*

$$f_{R_{S \wedge T}}(x|\varphi) = \int_x^1 \frac{1}{y} \cdot f_{R_S}(\frac{x}{y}|\varphi) \cdot f_{R_T}(y|\varphi) \, \mathrm{d}y.$$

Proof. Apply Theorem 6.3 and Lemma 7.2. It suffices to verify the integral bounds. $f_{R_S}(\frac{x}{y}|\varphi) = 0$ for $0 > \frac{x}{y}$ and $1 < \frac{x}{y}$, so we can ignore cases where $y < x$ and $y > 1$. □

The probability density function of $R_{\neg T}$ can also be derived under any condition φ:

Proposition 7.5. *For any target T,*

$$f_{R_{\neg T}}(x|\varphi) = f_{R_T}(1 - x|\varphi)$$

Proof. Since $R_{\neg T} = 1 - R_T$, if $R_{\neg T} = x$ then $R_T = 1 - x$. □

The probability density function of $R_{S\vee T}$, for independent S and T, under any condition φ, can be derived via the De Morgan rules (Proposition 7.3), the computation for trust conjunction (Theorem 7.4) and the computation for trust negation (Proposition 7.5).

Corollary 7.6. *If S and T do not share any users, then*

$$f_{R_{S\vee T}}(x|\varphi) = \int_{1-x}^{1} \frac{1}{y} \cdot f_{R_S}(1 - \frac{1-x}{y}|\varphi) \cdot f_{R_T}(1-y|\varphi)\,\mathrm{d}y.$$

Theorem 7.4, Corollary 7.6 and Proposition 7.5 are sufficient to derive trust opinions about arbitrary targets (where no active parties appear more than once), given arbitrary interactions with the active parties.

Corollary 7.7. *For every (finite) target where no users appear more than once, an explicit function for the trust opinion can be computed by the subject.*

Proof. Apply structural induction over the shape of the target. The base case (simple trust opinions) is proven in Theorem 6.5. To prove the induction step, take Theorem 7.4, Corollary 7.6 or Proposition 7.5 as rewrite rules from left to right. □

In Example 7.3, we derive an explicit formula for the trust opinion of $B \wedge C$, and look at some of its properties.

Example 7.3. Assume that the subject, A, wants to establish a trust opinion about the target, $B \wedge C$. In the past, A has interacted as a passive party with B several times; five times B behaved well, and once badly. Furthermore, A has interacted with C, too; four times C behaved well, and twice badly. The trust opinion of A about $B \wedge C$ is $f_{R_{B\wedge C}}(x|O_B^A = (5,1), O_C^A = (4,2))$. Using Theorem 7.4, the trust opinion can be computed as

$$\int_x^1 \frac{1}{y} \cdot f_{R_B}(\frac{x}{y}|O_B^A = (5,1), O_C^A = (4,2)) \cdot f_{R_C}(y|O_B^A = (5,1), O_C^A = (4,2))\,\mathrm{d}y.$$

By Independency $\mathrm{I2}_\ell$, we obtain

$$\int_x^1 \frac{1}{y} \cdot f_{R_B}(\frac{x}{y}|O_B^A = (5,1)) \cdot f_{R_C}(y|O_C^A = (4,2))\,\mathrm{d}y.$$

Which by Definition 6.7 and Theorem 6.5 is equal to

$$\int_x^1 \frac{\frac{1}{y} \cdot (\frac{x}{y})^6 \cdot (1-\frac{x}{y})^2 \cdot y^5 \cdot (1-y)^3}{\mathrm{B}(5,1) \cdot \mathrm{B}(4,2)}\,\mathrm{d}y.$$

The formula can be formulated without an integral (using e.g. Mathematica), and instead using some combinatorial functions, so that it reduces to

$$\frac{x^2 \cdot (1-x)^4 \cdot \Gamma(2) \cdot \Gamma(3) \cdot {}_2F_1(2,3;5;\frac{x-1}{x})}{\Gamma(5) \cdot \mathrm{B}(5,1) \cdot \mathrm{B}(4,2)}.$$

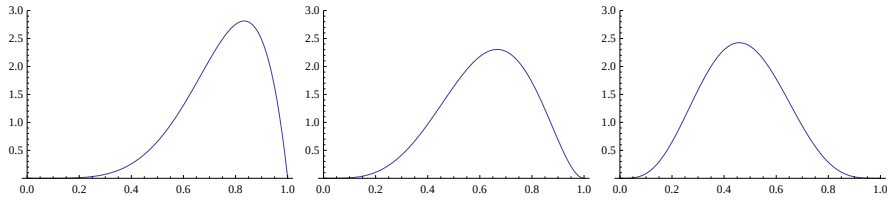

Figure 7.3: From left to right: trust opinion about B, about C and about $B \wedge C$.

where Γ is the gamma function, B the beta function (not to be confused with the beta distribution) and ${}_2F_1$ a hypergeometric distribution. This, in turn, simplifies to

$$2205x^4(1 + 4x - 5x^2 + 2x(2 + x)\log(x)).$$

The conjunction operation is depicted graphically in Figure 7.3. The rightmost distribution is the conjunction of the other two distributions. Recall that the abscissa depicts the integrity parameter of the targets in question. Thus, the more mass is on the right-hand side of the graph, the bigger the probability that the target has a high integrity. As we can see, both active parties (B and C) have a relatively high integrity, but their conjunction ($B \wedge C$) does not.

The expected value of the trust opinion about a target is equal to the probability that the target succeeds. Computation of the expected value of $f_{R_{B\wedge C}}$ yields $^{15}/_{32}$. The expected value for the single user B to succeed, f_{R_B}, is $^3/_4$ and for C to succeed, f_{R_C} is $^5/_8$. Not coincidentally, the expected value for $f_{R_{B\wedge C}}$ is the product of that of f_{R_B} and f_{R_C}, namely $^{15}/_{32} = {}^3/_4 \cdot {}^5/_8$.

As we suspected in the beginning of this chapter, and seen for a specific case in Example 7.3, the expected behaviour of a conjunction of targets, is equal to product of the expected behaviour of both targets.

Corollary 7.8. *If S and T do not share any active parties, then*

$$\mathbf{E}(R_{S\wedge T}) = \mathbf{E}(R_S) \cdot \mathbf{E}(R_T).$$

Proof. Immediate consequence of Lemma 7.2. □

Similarly $\mathbf{E}(R_{\neg T}) = 1 - \mathbf{E}(R_T)$ and $\mathbf{E}(R_{S\vee T}) = \mathbf{E}(R_S) + \mathbf{E}(R_T) - \mathbf{E}(R_S) \cdot \mathbf{E}(R_T)$.

Although the derivation in Example 7.3 seems asymmetrical with respect to S and T, commutativity and associativity hold.

Corollary 7.9. *Conjunctions and disjunctions of independent trust opinions are commutative and associative. Thus $R_{S\wedge T} = R_{T\wedge S}$, $R_{S\vee T} = R_{T\vee S}$, $R_{(S\wedge T)\wedge U} = R_{S\wedge(T\wedge U)}$ and $R_{(S\vee T)\vee U} = R_{S\vee(T\vee U)}$.*

Proof. Immediate consequence of Lemma 7.2. □

In Example 7.3, we have shown a specific composite trust opinion to be

$$f_{R_{B\wedge C}}(x|O_B^A = (5,1), O_C^A = (4,2)) = 2205x^4(1 + 4x - 5x^2 + 2x(2 + x)\log(x)).$$

Now, one can wonder whether there exists a *beta distribution* with a probability density function of that shape. It is important to realise that if (composite) trust opinions are closed under the logical trust operations, then there must be such a beta distribution. We prove that, in general, such a beta distribution does not exist:

Theorem 7.10. *A composite trust opinion need not be representable by a beta distribution.*

Proof. The expression $2205x^4(1 + 4x - 5x^2 + 2x(2 + x)\log(x))$, is a composite trust opinion, but not a polynomial. The probability density function of a beta distribution is always a polynomial (see Definition 6.7). Hence, that composite trust opinion is not based on a beta distribution. □

From Theorem 7.10, we can conclude that every trust model in which the trust opinions are (*isomorphic* to) beta models violates at least one of the assumptions. A famous example is Subjective Logic [Jøs97] (the binomial version without base rate), other examples include CertainLogic [RHMV11]. As the methodology of this paper is inspired by Subjective Logic, Dependencies D1, D2 and D3 are in line with the assumptions in Subjective Logic. Furthermore, the Independencies $\mathrm{I1}_\ell$, $\mathrm{I2}_\ell$, and $\mathrm{I3}_\ell$ are also based on (non-formal formulations in) Subjective Logic. By the pigeon hole principle, Dependency $\mathrm{D4}_\ell$ for conjunctions (or Dependency $\mathrm{D5}_\ell$ for disjunctions or Dependency $\mathrm{D6}_\ell$ for negation) or Dependency $\mathrm{D7}_\ell$ must be violated. Dependency $\mathrm{D4}_\ell$ states that $E_{S\wedge T} = \mathrm{s}$ iff $E_S = \mathrm{s}$ and $E_T = \mathrm{s}$ (for independent S and T), and Dependency $\mathrm{D7}_\ell$ asserts that the integrity of a composite target is determined by the integrity of the active parties. In other words, Dependencies $\mathrm{D4}_\ell$-$\mathrm{D7}_\ell$ are a formalisation of the natural semantics of the logical trust operations.

Thus, Subjective Logic and CertainTrust either contradict their own fundamental assumptions (the assumptions of the Beta model), or have an unreasonable definition of the logical trust operations (one where either "$E_{S\wedge T}$ iff E_S and E_T" or "a targets behaviour is determined by its constituents" does not hold). We will observe in the next chapter, that trust chaining is also not closed over beta distributions, whereas existing models assume they are. Therefore, we postpone a more detailed analysis to Section 8.4.

7.3 Conclusion

We have applied techniques from the Beta model to composite trust opinions; trust opinions based on trust conjunction, trust disjunctions or trust negation. Thus, we have derived an explicit definition of a trust opinion of the shape "Can I trust that both A and B will behave according to agreement?" Of course, more general statements exist, where for "A and B" any propositional formula can be substituted and our result also generalises to encompass these as well. By deriving a computation for the logical trust operations from the assumptions of the Beta model together with the natural semantics of the logical trust operations, we effectively obtain a correctness trust model in the Beta paradigm that has trust aggregation and the logical trust operations.

We have furthermore proven some properties about composite trust opinions. First, the trust opinion about a target $S \wedge T$ has the expected value $s \cdot t$, where s and t are the expected values of the trust opinion about S and T. (Similarly, for $S \vee T$, it is $s + t - s \cdot t$.) Second, that the expected algebraic properties – commutativity, associativity, De Morgan and double negation – of the logical operators hold when interpreted as logical trust operations. Third, a composite trust opinion is in general not a beta distribution. Hence, no trust model with elements isomorphic to beta distributions can satisfy all our assumptions, which implies they either contradict their own foundation or the natural semantics of the logical trust operations.

8

Beta Models with Trust Chaining

In interactions over the Internet, the information which a subject has about past behaviour of a target is limited. Hence it might be beneficial to ask for the help of third parties. Third party statements about the target are called *recommendations* (see also Section 2.3), hence we call these third parties *recommenders*. Trust opinions constructed with the help of recommendations are called *chained trust opinions*. In this paper, we formally study the implications of such recommendations.

To allow trust opinions to incorporate third parties, some modern trust models use the Beta model as a foundation, and increase the model's expressivity and its (practical) applicability by including recommendations. We say that a model which uses the Beta model as a foundation is in the *Beta paradigm*. Many models in the Beta paradigm that support *trust chaining* are ad-hoc. By ad-hoc models, we understand models designed according to intuition or statistical effectiveness in practice, rather than formal correctness (like formal correctness of *trust aggregation* in the Beta model). The notion that current models with trust chaining are ad-hoc, is supported by the fact that the research community has not yet settled on one trust model [KNS08], not even under the assumption that the trust model is in the Beta paradigm [JMP06].

Rather then proposing a new model in the Beta paradigm, we rigorously prove properties of trust chains valid in all models in the Beta paradigm. We refer to the collection of all correctness trust models in the Beta paradigm that support trust chaining as the *Beta family with trust chaining*. We show the following properties for the Beta family with trust chains. Chained trust opinions are modular (Proposition 8.8 and Theorem 8.9), meaning that complex trust opinions can be constructed from simpler ones. Every trust model makes implicit or explicit assumptions about how a recommender lies or about the number of interactions between *users* (Corollary 8.10). Chained trust opinions resulting have a different shape from the trust opinions in the Beta model (Theorem 8.11). Furthermore, Subjective Logic, an expressive ad-hoc extension of the Beta model, is not in the Beta family with trust chains (Corollary 8.13). The same conclusion can be derived for models similar to Subjective Logic, such as TRAVOS [TPJL06] and CertainTrust [Rie07] (Corollary 8.12).

In Section 8.1, we formalise the notion of recommendations and add it to the Beta model, effectively formalising all models in the Beta family with trust chains. Then, in Section 8.2, we study the most basic trust chains in the Beta family with trust chains. In Section 8.3, we prove that all models in the Beta family with trust chains have the property that trust opinions can be constructed modularly from the most basic trust chains. Finally, in Section 8.4, we characterise trust models

in the Beta family with trust chains, and show that existing models based on the Beta model are not in the Beta family with trust chains.

8.1 Formalisation

According to the Beta model, a subject A constructs his trust opinion using only his own information, when planning to interact with a target C. Depending on the constructed trust opinion, A chooses to interact or not. Suppose that A wants to make a more informed decision. Then, the subject A may ask a third party, a recommender B, for advice. A recommender could provide an honest recommendation, or lie. Chained trust opinions are based on the notion that a trust opinion on the recommender B is a valid measure for the likelihood that B provides an honest recommendation about C. More formally:

Definition 8.1 (Chained trust opinions)**.** Every recommender (like every target) has an integrity parameter that determines the probability of a successful interaction. In case of a successful interaction, their recommendation is their trust opinion about the target. Chained trust opinions are trust opinions based on recommendations from recommenders.

We add recommendations to the classroom game:

Running Example. After students figured out the *logical trust operations*, the teacher decided to modify the game differently. To keep the game interesting, as well as make it a more realistic emulation of e-commerce, the teacher adds recommendations in the following way: In the beginning of every turn, the teacher not only assigns a subject $c_i \in S$ and a target $c_j \in S$, but also a set of recommenders $R \subseteq S \setminus \{c_i, c_j\}$ if c_i has never interacted with c_j. Every recommender $c_k \in R$ has to honestly provide their past interactions with c_j with probability p_k, or construct and provide a fake past history with c_j with probability $1-p_k$. Again, students with a high integrity p_k are more likely to provide the past interactions rather than fake interactions. For a subject to construct the most accurate trust opinion, he needs to incorporate his opinion of c_k and the recommendation by c_k, for all $c_k \in R$.

To formally model recommendations in the Beta model, we introduce an additional random variable to the random variables E_C, R_C and R_AC from the Beta model.

- $S_C^B \colon \Omega \to \mathbb{N} \times \mathbb{N}$ is a discrete random variable modelling recommendations of the recommender B about the target C, representing the alleged past interactions between B as passive party and C as active party.

We also introduce additional sets of random variables:

$$\begin{aligned} \mathbb{S} &:= \{S_C^B : B, C \in \mathbf{A}\}, \\ \mathbb{W}_{\mathbb{S}} &:= \mathbb{W} \cup \mathbb{S}. \end{aligned}$$

Before introducing the relations between random variables, we introduce a collection of functions $\chi^B \colon [0,1] \times \mathbb{N} \times \mathbb{N} \to (\mathbb{N} \times \mathbb{N} \to [0,1])$, for $B \in \mathbf{A}$. A function χ^B

is called a *lying strategy*. For $b \in [0,1]$ and $w_s, w_f \in \mathbb{N}$, $\chi^B(b, w_s, w_f)$ is a probability distribution representing the probability that recommender B states (y_s, y_f). Game-theoretically, χ^B should be seen as the strategy of B, and $\chi^B(b, w_s, w_f)$ should be seen as the mixed move of B.

The dependencies and independencies are based on those in the Beta model. In fact, Dependencies D1-D3 are lifted directly from the Beta model, to become Dependencies D1_S-D3_S. Independencies I1-I3 are identical when restricted to random variables from the Beta model, i.e. in $\mathbb{W}$. Independencies I1-I3 are extended to Independencies I1_S-I3_S to deal with recommendations.

Let $a, b, x \in [0,1]$, $n, k \in \mathbb{N}$ and $\lambda\colon \mathbb{N} \to [0,1]$ as well as $\chi^B\colon [0,1] \times \mathbb{N} \times \mathbb{N} \to (\mathbb{N} \times \mathbb{N} \to [0,1])$, where $B \in \mathbf{A}$ be probability distributions. For all users $A, B, C \in \mathbf{A}$ we set up the following additional dependency and independency relations as our assumptions.

D1_S R_C is uniformly distributed on $[0,1]$.

D2_S $P(E_C{=}\text{S} | R_C{=}c) = c$.

D3_S $P(O_C^A{=}(x_s, x_f) | R_C{=}c) = \binom{x_s+x_f}{x_s} c^{x_s}(1-c)^{x_f}\lambda(x_s + x_f)$.

D4_S $P(S_C^B{=}(w_s, w_f) | E_B{=}\text{S}, O_C^B{=}(w_s, w_n)) = 1$.
Assumes that good behaviour of B implies that the recommendation of B corresponds to his interaction history with C.

D5_S $P(S_C^B{=}(y_s, y_f) | E_B{=}\text{F}, R_B{=}b, O_C^B{=}(w_s, w_f)) = \chi^B(b, w_s, w_f)(y_s, y_f)$.
Asserts that the lying strategy determines the fake recommendations of recommender B.

I1_S For $W \in \mathbb{W}_\mathbb{S} \backslash \{O_C^A, S_C^A\}$, it holds that $O_C^A \perp\!\!\!\perp W | R_C$.
Similar to Independency I1, except recommendations (other than from A about C) are also independent from the interaction history between A and C, under a fixed integrity parameter of C.

I2_S For $W \in \mathbb{W} \backslash \{R_C\}$ and $\{C, D_0, \ldots D_n\} = \mathbf{A}$, $R_C \perp\!\!\!\perp W | E_C, O_C^{D_0}, \ldots, O_C^{D_n}$.
Similar to Independency I2, except all recommendations are also independent from the integrity parameter of C, under the outcome of C and the interaction histories with C.

I3_S For $W \in \mathbb{W}_\mathbb{S} \backslash (\{E_B\} \cup \{S_D^B : D \in \mathbf{A}\})$, it holds that $E_B \perp\!\!\!\perp W | R_B$.
Similar to Independency I3, except recommendations (other than those from B) are also independent from the outcome of the interaction with B, under the integrity parameter of B.

I4_S For $W \in \mathbb{W}_\mathbb{S} \backslash \{S_C^B\}$, it holds that $S_C^B \perp\!\!\!\perp W | E_B{=}\text{F} \cap R_B \cap O_C^B$.
The choice of B for making fake recommendations about C is completely determined by $\chi^B(b, n, m)$ in Dependence D5_S.

We define that models in the Beta family with trust chains adhere to Dependencies $D1_S$–$D5_S$ and Independencies $I1_S$–$I4_S$, and support this definition below:

Definition 8.2 (Beta family with trust chains)**.** A model is said to be in the *Beta family with trust chaining*, when it satisfies Dependencies $D1_S$–$D5_S$ and Independencies $I1_S$–$I4_S$.

There are models that are inspired by the Beta model, and that include an operator $\otimes$ dealing with recommendations, but that are not models in the Beta family with trust chains. We argue that such models either are not Beta models or that $\otimes$ is not a trust chaining operator. If a model violates any of the Dependencies $D1_S$–$D3_S$ or Independencies I1–I3, it is not a Beta model. We distinguish the possible violations of an assumption for each remaining assumption separately. If a model violates

$D4_S$, then the model does not support trust chaining.

$D5_S$, then another assumption must also be violated. This is since under Dependencies $D1_S$–$D4_S$ and Independencies $I1_S$–$I4_S$ there exists a χ^B such that $\chi^B(b, w_s, w_f)(y_s, y_f) = P(S_C^B{=}(y_s, y_f)|O_C^B{=}(w_s, w_y), R_B = b, E_B = \text{F})$.

$I1_S$, then the model either violates Independency I1, or it assumes that some S_D^C are dependent with O_C^A given R_C. This is not in the spirit of the Beta paradigm, where we assert that such types of dependencies cannot be identified. Hence, the outcomes of the interactions between A and C should depend only on C.

$I2_S$, then the model either violates Independency I2, or it assumes that some R_C are dependent with S_E^D given all observations of C. This is not in the spirit of the Beta paradigm. The collection of all interactions with C should be an optimal estimator for R_C.

$I3_S$, then the model either violates Independency I3, or it assumes that some E_C are dependent with S_E^D (for $D \neq C$) under all observations of C, which is not in the spirit of the Beta paradigm. The probability of success of an interaction (given the integrity) should not be influenced by recommendations of others.

$I4_S$, then in this model recommenders differentiate their strategy either on information they cannot know (e.g. interactions that the recommender did not participate in) or on information that is irrelevant for the recommendation (e.g. his opinion on yet another user).

Not every model in the Beta family with trust chains is formalised our way. A model is already in the Beta family with trust chains when the assumptions can be reformulated to fit the assumptions up to isomorphisms.

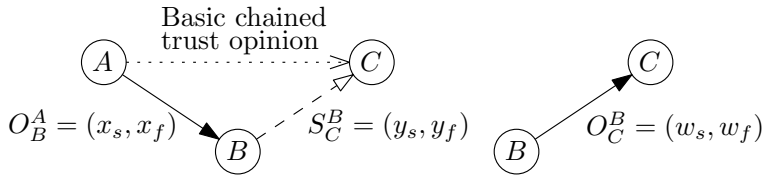

Figure 8.1: Left: The view of subject A about target C, including the recommendation S_C^B from B about C. Right: The view of recommender B about target C.

8.2 Basic Trust Chains

The most basic scenario that involves trust chains, involves exactly one recommendation. This recommendation is given about a target with which the subject has no prior interactions. In other words, the recommendation is the only source of information that a subject has. This scenario is called *basic trust chaining*. It is studied in this section. In Section 8.3, we then prove that more complicated scenarios can be reduced to scenarios with basic trust chains.

We define the basic trust chain as the simplest case of trust chaining, namely a trust chain involving only one recommender. The basic trust chain is also depicted in Figure 8.1.

Definition 8.3 (Basic trust chain)**.** A basic trust chain consists of three users: the subject A, the recommender B, and the target C. The subject has an *interaction history* $x = (x_s, x_f)$ with the recommender. The recommender provides a recommendation $y = (y_s, y_f)$ about the target and, in reality, has an interaction history $w = (w_s, w_f)$ with the target. The trust opinion of subject A about target C with recommendations by recommender B is the *chained trust opinion*. It is depicted in Figure 8.1.

Running Example. In the classroom game, basic trust chains appear when the teacher assigns only one recommender. Then, the subject is $c_i \in S$, the target is $c_j \in S \setminus \{c_i\}$ and the set of recommenders is $\{c_k\} \subset S \setminus \{c_i, c_j\}$.

We may now formulate the basic chained trust opinion of A about C with recommendations given by B as $f_{R_C}(c|O_B^A{=}(x_s, x_f), S_C^B{=}(y_s, y_f))$. In other words, to formulate a trust opinion about the target, the subject uses its interaction history about the recommender as well as the (possibly fake) recommendation given by the recommender. If A has never directly interacted with B, the pair (x_s, x_f) equals $(0, 0)$.

Theorem 8.1. *Dependencies $D1_S$–$D5_S$ and Independencies $I1_S$–$I4_S$ are sufficient to derive the basic chained trust opinion of A about C with recommendations by B as: $f_{R_C}(c|O_B^A{=}(x_s, x_f), S_C^B{=}(y_s, y_f)) =$*

$$\mathbf{eq^1}_{y_s,y_f}(c) \cdot \mathbf{eq^2} + \sum_{w \in O_C^B} (\mathbf{eq^1}_{w_s,w_f}(c) \cdot \mathbf{eq^3} \cdot (1 - \mathbf{eq^2})), \tag{8.1}$$

where,

$$\begin{aligned}
\mathbf{eq^1}_{\varphi_s,\varphi_f}(c) &= f_{\mathrm{B}}(c;\varphi_s+1,\varphi_f+1),\\
\mathbf{eq^2} &= \frac{\mathbf{eq^4}\cdot(x_s+1)}{\mathbf{eq^4}\cdot(x_s+1)+\sum_{w'\in O_C^B}\mathbf{eq^5}_{w'}\cdot(x_f+1)},\\
\mathbf{eq^3} &= \frac{\mathbf{eq^5}_{w_s,w_f}}{\sum_{w'\in O_C^B}\mathbf{eq^5}_{w'_s,w'_f}},\\
\mathbf{eq^4} &= \lambda(y_s+y_f)\cdot\binom{y_s+y_f}{y_s}\cdot\frac{y_s!y_f!}{(y_s+y_f+1)!}\\
\mathbf{eq^5}_{\varphi_s,\varphi_f} &= \int_0^1 \chi^B(b,\varphi_s,\varphi_f)(y_s,y_f)\cdot f_{\mathrm{B}}(b;x_s+1,x_f+2)\ \mathrm{d}b\\
&\quad\cdot\lambda(\varphi_s+\varphi_f)\cdot\binom{\varphi_s+\varphi_f}{\varphi_s}\cdot\frac{\varphi_s!\varphi_f!}{(\varphi_s+\varphi_f+1)!}
\end{aligned}$$

Proof. We short-hand (ψ_s,ψ_f) to ψ, for arbitrary ψ. The equations $\mathbf{eq^1}$–$\mathbf{eq^5}$ represent the following probabilities:

$$\begin{aligned}
\mathbf{eq^1}_{\varphi}(c) &= P(R_C{=}c|O_B^A{=}x,S_C^B{=}y,E_B{=}u,O_C^B{=}w),\\
\mathbf{eq^2} &= P(E_B{=}\mathrm{S}|O_B^A{=}x,S_C^B{=}y),\\
\mathbf{eq^3} &= P(O_C^B{=}w|O_B^A{=}x,S_C^B{=}y,E_B{=}\mathrm{F}),\\
\mathbf{eq^4} &= P(S_C^B{=}y|O_B^A{=}x,E_B{=}\mathrm{S}),\\
\mathbf{eq^5}_{\varphi} &= P(S_C^B{=}y,O_C^B{=}\varphi|O_B^A{=}x,E_B{=}\mathrm{F}).
\end{aligned}$$

The proof that $\mathbf{eq^1}$–$\mathbf{eq^5}$ actually represent these probabilities can be found in Section B.1 in Appendix B. The correctness of Formula (8.1) follows from the fact that $\mathbf{eq^1}$–$\mathbf{eq^5}$ compute the above probabilities, given that, for all $W\in\mathbb{W}_{\mathbb{S}}$: $S_C^B \perp\!\!\!\perp W|E_B{=}\mathrm{S}\cap O_C^B$ follows from Dependency $\mathrm{D4}_S$. □

Although Formula (8.1) may seem complicated, it can abstractly be viewed as a (infinite) weighted sum of *beta distribution*s:

Proposition 8.2. *For every* entanglement *and* lying strategy*, a basic chained trust opinion is a weighted sum of beta distributions.*

Proof. If we collect factors that do not contain the variable c (i.e. $\mathbf{eq^2}$-$\mathbf{eq^5}$) in the scalars k and k_{w_s,w_f}, Formula (8.1) simplifies to

$$k\cdot c^{y_s}(1-c)^{y_f}+\sum_{w_s,w_f\in\mathbb{N}\times\mathbb{N}} k_{w_s,w_f}c^{w_s}(1-c)^{w_f}. \tag{8.2}$$

□

Furthermore, for some specific models in the Beta family with trust chains, the formula significantly simplifies. Particularly, for a lying strategy that consists of constructing truthful recommendations (see dash-dotted graph in Figure 8.2), the trust opinion is a beta distribution:

Proposition 8.3. *Let* $\chi^B(b,w_s,w_f)(y_s,y_f)=1$ *iff* $(w_s,w_f)=(y_s,y_f)$*, then the trust opinion from Formula* (8.1) *simplifies to* $\vartheta_{y_s,y_f}(c)$*.*

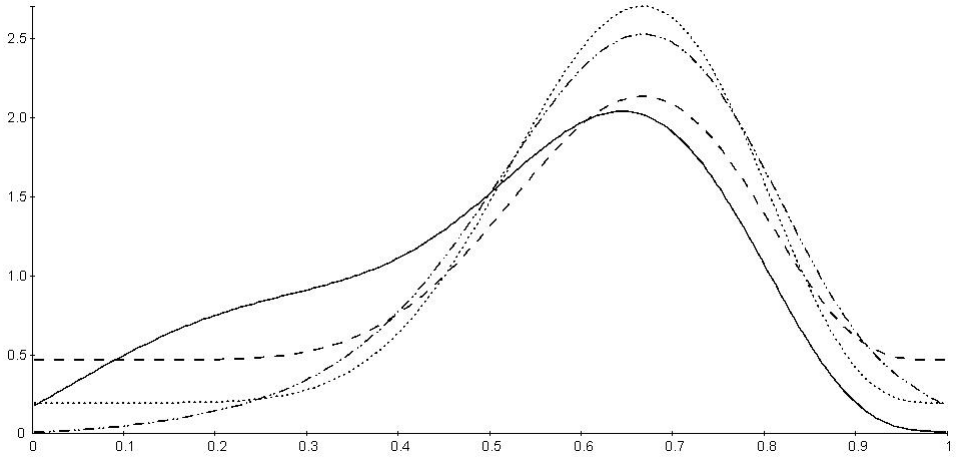

Figure 8.2: The same trust chain, $x = (6, 5)$ and $y = (8, 4)$, with different lying strategies. Solid: lies opposite of his true opinion. Dashed: lies independent of the his true opinion. Dash-dotted: lies similar to his true opinion. Dotted: lies with a positive bias.

Proof. Fill the choice of χ^B in, in Formula (8.1). □

Taking an arbitrary entanglement λ and a lying strategy that consists of constructing completely informationless recommendations (see dashed graph in Figure 8.2), the trust opinion is a weighted sum of a beta distribution and the uniform distribution:

Proposition 8.4. *Let* $\chi^B(b, w_s, w_f)(y_s, y_f) = \frac{1}{y_s+y_f+1}$ *iff* $w_s + w_f = y_s + y_f$*, then the trust opinion from Formula* (8.1) *simplifies to* $\frac{x_s+1}{x_s+x_f+2}\vartheta_{y_s,y_f}(c) + \frac{x_f+1}{x_s+x_f+2}$.

Proof. Fill the choice of χ^B in, in Formula (8.1). □

An immediate consequence of Theorem 8.1 and Proposition 8.2 is that a model that supports basic chained trust opinions, makes assumptions about the entanglement and lying strategies.

Corollary 8.5. *It is not possible to compute basic chained trust opinions without knowledge of the entanglement* λ *and the lying strategy* χ^B*.*

Proof. Propositions 8.3 and 8.4 assume different lying strategies. The resulting simplifications of Formula (8.1) are different. Hence the choice of χ^B matters. □

Running Example. In terms of the classroom game, the corollary states that it is relevant how many turns have been played and how students lie. If a recommendation states "8 successes and 2 failures", but each student has played 9 turns, the recommendation is clearly fake. Suppose, a student c_k provides a recommendation to c_i that is likely to be fake. If c_k and c_i are good friends outside of the game, c_k might have a lying strategy of creating fake recommendations that strongly resemble the truth. Otherwise, c_k provides recommendations unrelated to the truth. Then, it is wise for c_i to rely on the recommendation of his friend, but not on recommendations of other arbitrary classmates.

Corollary 8.5 implies that without assumptions on λ and χ^B, no model can provide trust opinions. Therefore, any trust model in the Beta family with trust chains either implicitly or explicitly makes assumptions about numbers of interactions and about the lying strategy of recommenders. We believe that making implicit assumptions about lying strategies is inappropriate, as it obfuscates the analysis of a model or hides undesirable consequences of a model. Hence, we suggest that new proposals for models in the Beta family with trust chains explicitly (and formally) provide the *lying strategy* of the recommenders.

Corollary 8.6. *For every entanglement λ and lying strategy χ^B, the subject can calculate the basic chained trust opinion.*

Proof. Apply Formula (8.1), with the relevant instantiations of λ and χ^B. □

Thus, when the number of turns in the classroom game is known, and it is known what kind of lying strategy each student has, the subject can correctly compute the trust opinion, whenever the teacher assigns only one recommender.

A positive consequence of Corollary 8.6 is that defining the entanglement and the lying strategy is sufficient to explicitly define a model in the Beta family with trust chains. Not only is it mathematically possible, but we have developed a tool named *Canephora*[1] that can compute basic chained trust opinions, when χ^B and λ are provided. The tool is a proof of concept, that creating a model in the Beta family with trust chains is merely a matter of defining an entanglement and lying strategies. It is a prototype that allows the numerical comparison between different models (i.e. different choices of entanglements and lying strategies). We explain the tool in further detail in Section 8.2.1.

In Section 8.3, we see that defining the entanglements and the lying strategies is sufficient to explicitly define models in the Beta family with trust chains (not just models restricted to basic trust chains).

Determining the entanglement λ is usually simpler than finding the lying strategy. In many e-commerce systems, the number of interactions between users is known to the system. For example, eBay knows if a product is sold, even if it does not know whether the transaction was a success for the subject. Or in the classroom game, the teacher announces the number of turns, explicitly providing λ. Even if the entanglement is unknown, by restricting the choices of χ^B, the entanglement λ can be eliminated from Formula (8.1).

Lemma 8.7. *There exist lying strategies, where the entanglement has no impact on the basic chained trust opinion.*

Proof. Consider the basic chained trust opinion given by Formula (8.1). For all $b \in \mathbb{R}$, and $w_s, w_f, y_s, y_f \in \mathbb{N}$ such that $w_s + w_f \neq y_s + y_f$, take $\chi^B(b, w_s, w_f)(y_s, y_f) = 0$. Then, $\lambda(\varphi_s + \varphi_f)$ cancels out of $\mathbf{eq^5}$ unless $\varphi_s + \varphi_f = y_s + y_f$. In the reduced term, we can substitute $\lambda(\varphi_s + \varphi_f)$ for $\lambda(y_s + y_f)$. Then $\lambda(y_s + y_f)$ is a scalar that appears in every summand in the numerators and denominators of $\mathbf{eq^2}$ and $\mathbf{eq^3}$. Thus λ cancels out of Formula (8.1). □

[1] `http://satoss.uni.lu/software/canephora`

Running Example. If a recommender makes a recommendation of which the size was impossible (or very unlikely), a student can identify the recommendation as a fake (or likely a fake). If all students take care never to fall into the pitfall of sizing fake recommendations according to a different distribution than the real interactions, sizing becomes irrelevant. Hence, the entanglement cancels out.

8.2.1 Canephora

The *Canephora* tool is both a prototype for a trust system with trust chaining and a tool for analysis of lying strategies. The high level functionality of the Canephora tool is that it takes a lying strategy, an entanglement and a set of parameters including the trust opinion regarding the recommender and the recommendation, and provides the chained trust opinion. The tool computes Formula (8.1), for given lying strategy and entanglement, via approximation.

Canephora uses several windows. The main window, depicted in Figure 8.3, allows the user to specify the calculation. The second window, depicted in Figure 8.4 shows all the results in a single window, allowing easier comparison between results. Each time a graph is computed, it opens a new window with an overview of data relevant to the computation, as depicted in Figure 8.5.

In the main window (Figure 8.3), we identify the following fields from top to bottom: The first field is a slider labelled speed, which trades off speed and accuracy, in a way explained later. The next two fields select the entanglement. Most entanglements come in families (such as the poisson distribution) with a single parameter. Hence, the user can select the family of the distributions, and the parameter. The next field is a number that defines the number of samples taken in the integrals. The number is a trade off between speed and precision. The next three fields contain each contain two numeric inputs: The first pair of numeric inputs represents the interaction history between the subject and the recommender. The second pair of numeric inputs represents the recommendation made by the recommender about the target. The third pair of numeric inputs represents the interaction history between the subject and the target. By setting this pair to $(0, 0)$, the basic chained trust opinion is obtained. The last field selects the lying strategy.

The second window (Figure 8.4) shows all graphs that have been calculated (and are not closed). Each graph in the window is controlled by a specific window for an individual graph. From the windows of the individual graphs, the colour and stroke can be selected. Closing the window of the graph removes the graph from this window.

The window for the individual graphs (Figure 8.5) is opened when the calculate button is hit on the main window. When the calculation is finished (the calculation happens in an individual thread, so the program does not block), this window shows the graph, and relevant information. The relevant information contains the expected value, the mode, the variance, different notions of entropy, a beta distribution approximating the graph, the calculation time, and the information used to calculate this graph.

We assert that the tool is an effective prototype for trust systems with basic trust chaining partially because of the data being calculated and displayed. Another rea-

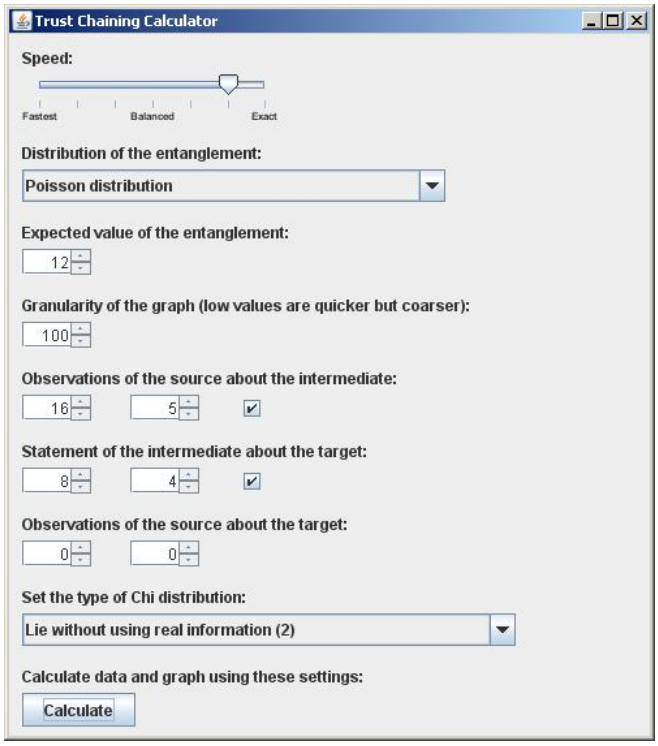

Figure 8.3: The main window of Canephora.

son is that, not only, are several useful entanglements and lying strategies included in the tool, but the tool has an interface for defining arbitrary entanglements and lying strategies. The combination of these two factors imply that the theoretical proof that a Beta model with trust chains can be defined by selecting an entanglement and lying strategies, also holds in practice. The tool does not yet support more involved trust chains, but Section 8.3 implies that involved trust chains can be constructed from basic trust chains.

The tool must approximate the results, since computing the results exactly is (in general) intractable. Formula (8.1) contains definite integrals and infinite summations, that contain the entanglement and the lying strategy, hence, these integrals and infinite summations cannot be reduced in a general way. Definite integrals can be approximated using the midpoint rule, where the quality of the approximation improves as the number of midpoints increases. The number of midpoints can be selected by the user. The infinite summations sum probabilities to probabilities, meaning that each term is non-negative, and that the sum converges to a value equal to at most 1. That means that there are less than k summands exceeding $1/k$. We can approximate the summation by ignoring sufficiently small summands. Rather than ignoring small summands, we can also do a Monte-Carlo simulation. Depending on the precision setting, more or less summands are ignored, and the Monte-Carlo simulation plays a big role only in rough simulations.

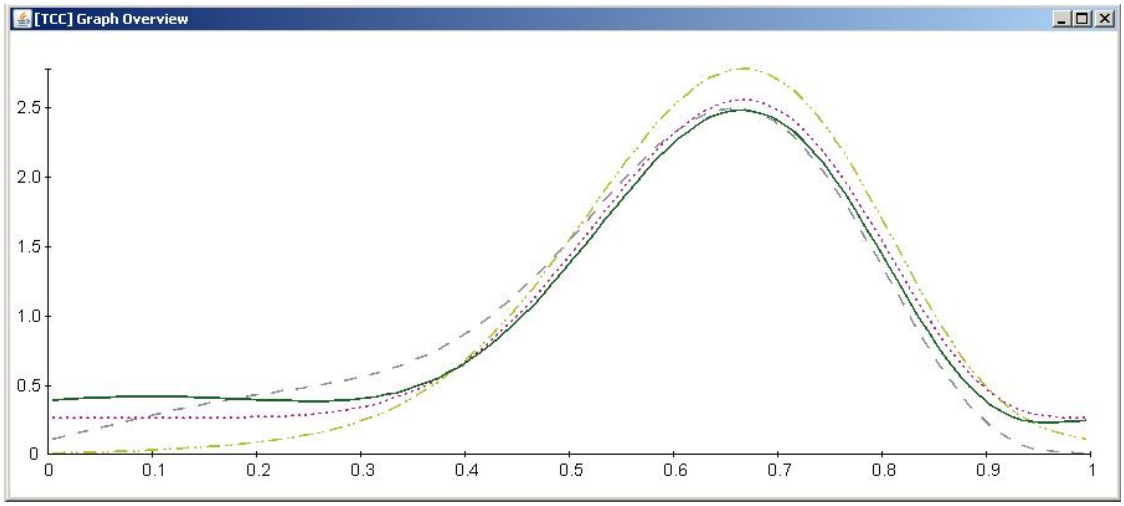

Figure 8.4: Overview of all relevant graphs.

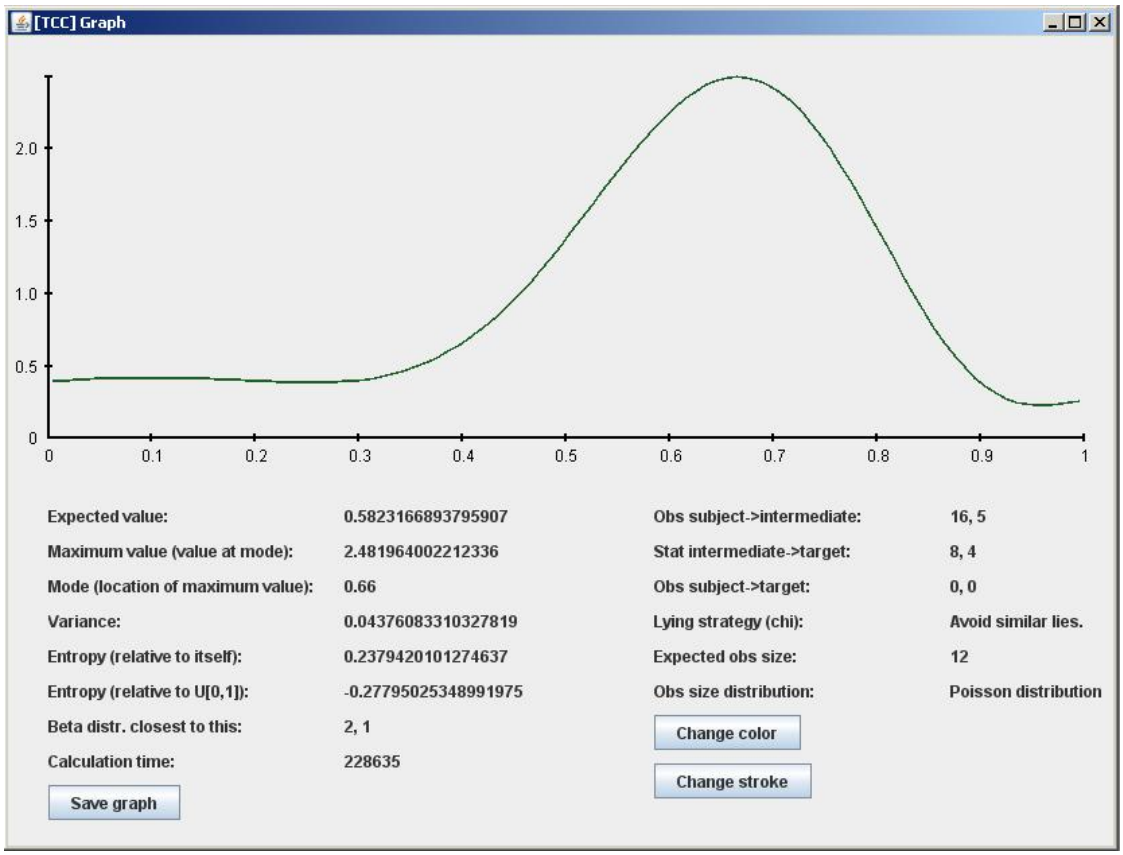

Figure 8.5: A window with a graph and relevant information.

8.3 Modular Construction of Trust Opinions

In Section 8.1, the assumptions of the Beta model were formally extended to include trust chaining. We have formally derived a parameterised trust opinion in the case of basic trust chains. However, it is possible that a subject receives more than one recommendation, or that the subject also has a *simple trust opinion* of the target. Recall trust aggregation from Definition 6.9. We first prove that a basic chained trust opinion can be aggregated with a simple trust opinion. Later, we prove that more complicated trust opinions can also be aggregated with basic trust opinions. The notion that aggregation of these trust opinions is possible, is called *modularity*.

Running Example. Imagine that the subject c_i constructs a trust opinion about the target c_j based on his past interactions (z_s, z_f) with c_j. However, the teacher also provides a recommender c_k, with which the subject has an interaction history of (x_s, x_f). The student c_k himself, gives the recommendation (y_s, y_f) about c_j. From the Beta model, the subject can construct his (simple) trust opinion based on (z_s, z_f). From Section 8.2, the subject can construct his (basic chained) trust opinion based on (x_s, x_f) and (y_s, y_f). The subject wants to construct a trust opinion based on (x_s, x_f), (y_s, y_f) and (z_s, z_f). We prove the subject merely needs to aggregate both trust opinions – to multiply the corresponding distributions.

Many trust models in the Beta family with trust chains (such as Subjective Logic) assert modularity. A priori, it is not obvious that the assertion of modularity is justified. In fact, the notion of *endogenous filtering* (Section 2.3) is based on the notion that modularity does not hold. The idea of endogenous filtering is to compare the contents of a recommendation about a target to what the subject already knows about the target. That means that the recommendation(s) must be evaluated in the context of the trust opinion of the subject. Modularity states the exact opposite of that. Namely that trust chains can be evaluated modularly, without context, and aggregated directly.

We distinguish two types of modularity. First modularity between a simple trust opinion and a basic chained trust opinion. Second, modularity between arbitrary (chained) trust opinions. The first notion is essentially the dual of the notion in endogenous filtering that recommendation should be compared to direct experience; the second is the dual of the notion that recommendations should be compared to each other. Since we prove both types of modularity, neither type of endogenous filtering should be applied to Beta family with trust chains.

In Chapter 10, we introduce a concrete model with trust chains (called the Default Model), rather than discuss a family of models with trust chains. Using a concrete model, we can provide an insight into why endogenous filtering should not be applied, and do so in Example 10.1 in Section 10.1.4. Due to the lack of a concrete model at this stage, we cannot provide a concrete intuition. However, we do have formal proofs for the entire family of models, namely Proposition 8.8 and Theorem 8.9.

We start by proving modularity between a simple trust opinion and a basic chained trust opinion:

Proposition 8.8. *For all models in the Beta family with trust chains, the chained trust opinion* $f_{R_C}(c|O_B^A{=}(x_s,x_f),S_C^B{=}(y_s,y_f),O_C^A{=}(z_s,z_f))$ *is the aggregate of the simple trust opinion* $f_{R_C}(c|O_C^A{=}(z_s,z_f))$ *and the basic chained trust opinion* $f_{R_C}(c|O_B^A{=}(x_s,x_f),S_C^B{=}(y_s,y_f))$.

Proof. We require Independency $\mathrm{I1}_S$ and Dependence $\mathrm{D1}_S$.

$$
\begin{aligned}
&f_{R_C}(c|O_B^A{=}(x_s,x_f),S_C^B{=}(y_s,y_f),O_C^A{=}(z_s,z_f))\\
=&\frac{P(O_B^A{=}(x_s,x_f),S_C^B{=}(y_s,y_f),O_C^A{=}(z_s,z_f)|R_C{=}c)\cdot f_{R_C}(c)}{P(O_B^A{=}(x_s,x_f),S_C^B{=}(y_s,y_f),O_C^A{=}(z_s,z_f))}\\
\overset{\mathrm{I1}_S}{=}&\frac{P(O_B^A{=}(x_s,x_f),S_C^B{=}(y_s,y_f)|R_C{=}c)\cdot P(O_C^A{=}(z_s,z_f)|R_C{=}c)\cdot f_{R_C}(c)}{P(O_B^A{=}(x_s,x_f),S_C^B{=}(y_s,y_f),O_C^A{=}(z_s,z_f))}\\
\overset{\mathrm{D1}_S}{\propto}&\frac{P(O_B^A{=}(x_s,x_f),S_C^B{=}(y_s,y_f)|R_C{=}c)\cdot f_{R_C}(c)P(O_C^A{=}(z_s,z_f)|R_C{=}c)\cdot f_{R_C}(c)}{P(O_B^A{=}(x_s,x_f),S_C^B{=}(y_s,y_f))\cdot P(O_C^A{=}(z_s,z_f))}\\
=&f_{R_C}(c|O_C^A{=}(z_s,z_f))\cdot f_{R_C}(c|S_C^B{=}(y_s,y_f),O_B^A{=}(x_s,x_f))
\end{aligned}
$$

□

Similar to Proposition 8.8, we can even prove that modularity holds for all trust opinions. Let φ be a collection of basic trust chains and potentially the interaction history between the target and the subject. In other words, for some n, let φ be given by:

$$[O_C^A{=}(z_s,z_f),]O_{B_1}^A{=}(x_s^1,x_f^1),S_C^{B_1}{=}(y_s^1,y_f^1),\ldots,O_{B_n}^A{=}(x_s^n,x_f^n),S_C^{B_n}{=}(y_s^n,y_f^n).$$

Theorem 8.9. *For all models in the Beta family with trust chains, the trust opinion* $f_{R_C}(c|O_B^A{=}(x_s,x_f),S_C^B{=}(y_s,y_f),\varphi)$ *is the aggregate of the trust opinion* $f_{R_C}(c|\varphi)$ *and the basic chained trust opinion*$f_{R_C}(c|O_B^A{=}(x_s,x_f),S_C^B{=}(y_s,y_f))$.

Proof. The only step of the proof in Proposition 8.8 that cannot be replicated (with φ substituted for $O_C^A{=}(z_s,z_f)$) is the application of Independency $\mathrm{I1}_S$. Thus:

$$
\begin{aligned}
&P(O_B^A{=}(x_s,x_f),S_C^B{=}(y_s,y_f),\varphi|R_C{=}c)\\
\overset{?}{=}&P(O_B^A{=}(x_s,x_f),S_C^B{=}(y_s,y_f)|R_C{=}c)\cdot P(\varphi|R_C{=}c)
\end{aligned}
$$

The proof obligation can be reduced (with Independencies $\mathrm{I1}_S$ and $\mathrm{I4}_S$) to $P(\varphi|R_C{=}c,E_C{=}u,O_C^B{=}(w_s,w_f),R_B{=}b)=P(\varphi|R_C{=}c)$, which follows from Independencies $\mathrm{I2}_S$ and $\mathrm{I3}_S$. See Section B.2 from Appendix B for details. □

From Theorem 8.9, we can conclude that the subjects can compute a trust opinion based on their own history with the target, as well as on recommendations of an arbitrary number of other users, provided that the subject can compute basic chained trust opinions for all recommendations. More generally, Theorem 8.9 allows us to generate the following structures $S(\lambda,\theta)=(P,O,g\colon P\to O,c_{\lambda,\theta}\colon P\times P\to O,a\colon O\times O\to O)$, where P is the set of interaction histories, O is the set of opinions, g is the function that maps interaction histories to simple trust opinions, $c_{\lambda,\theta}$ is the function that generates basic chained trust opinions (for entanglement λ and assignment of lying strategies to users θ) , and a represents aggregation of trust opinions. Depending on the choice of the entanglement and the assignment of lying strategies, the structures $S(\lambda,\theta)$ (generally) differ.

8.4 Analysis of the Models

The results from the last sections allow us to study the conditions that all trust opinions in all models in the Beta family with trust chains must adhere to. If an existing trust model violates these conditions, it is therefore not in the Beta family with trust chains. Which, in turn, means that these trust models either break an assumption of the Beta model (i.e. are not in the Beta paradigm), or its operator dealing with recommendations does not actually model trust chains according to Definition 8.2.

First, we point out that the work in Sections 8.2 and 8.3 captures all models in the Beta family with trust chains up to isomorphism:

Corollary 8.10. *Every model in the Beta family with trust chains is* isomorphic *to a structure $S(\lambda, \theta)$ for an entanglement λ and an assignment of lying strategies θ.*

Proof. The corollary is a direct consequence of Corollary 8.6 and Theorem 8.9. □

A consequence of the corollary is that if a model is in the Beta family with trust chains, there is a formulation of the model where the entanglement and the assignment of lying strategies are explicitly provided. This entails that if a formulation of a model does not explicitly mention the assignment of lying strategies, it is not an appropriate formulation as it obfuscates the lying strategies.

Furthermore, we prove a restriction on the shape of chained trust opinions. Before we do so, we define an exception to the restriction.

Definition 8.4 (Trivial lying strategy)**.** A lying strategy χ^B is trivial, when $\chi^B(b, w_s, w_f)(y_s, y_f) = 1$ iff $(w_s, w_f) = (y_s, y_f)$.

The trivial lying strategy is for the *recommender* to state the truth when lying. No serious trust model asserts that even unreliable recommenders always state the truth.

Theorem 8.11. *A basic chained trust opinion in* any *model in the Beta family with trust chains is in general not a* beta distribution*, except in the trivial case.*

Proof. Expression (8.2) from Proposition 8.2 can only represent a beta distribution $f_{\mathrm{B}}(c; S+1, F+1)$, if it can be simplified to $h \cdot c^S(1-c)^F$ for some $S, F \in \mathbb{N}$ and $h \in \mathbb{R}^+$. Rearranging the coefficients specifying how the constants depend on $\mathbf{eq^2}$ (henceforth let $p = \mathbf{eq^2}$), we therefore have to prove the following:

$$p \sum_{\ell} a_\ell c^\ell + (1-p) \sum_{m} b_m c^m = c^S(1-c)^F.$$

For $n < S$ or $n > S+F$, there is no summand c^n. Hence, we obtain that $pa_n + (1-p)b_n$ has to equal 0 for $n < S$ and $n > S+F$. This linear equation can only hold for a specific value of p, unless $a_S = b_S = 0$. As we prove the general case, we may not restrict p to a specific value, seeing that $\mathbf{eq^2}$ depends on x and y. Thus, $a_S = b_S = 0$. Hence the sum reduces to

$$p \sum_{\ell=S}^{S+F} a_\ell c^\ell + (1-p) \sum_{m=S}^{S+F} b_m c^m = c^S(1-c)^F.$$

The sum $\sum_{\ell=S}^{S+F} a_\ell c^\ell$ is proportional to a beta distribution, specifically $\mathbf{eq}^{\mathbf{1}}_y(c)$. The only beta distribution with maximal exponent $S+F$, and minimal exponent S is $f_{\mathrm{B}}(c; S{+}1, F{+}1)$. Thus, $\mathbf{eq}^{\mathbf{1}}_y(c) = f_{\mathrm{B}}(c; S{+}1, F{+}1)$, and $S = y_s$ and $F = y_f$. That means that the recommendation, $f_{\mathrm{B}}(c; S+1, F+1)$, equals the resulting opinion, meaning that the equation can only hold if we are in the exceptional case. □

Therefore, any model that represents all its chained trust opinions as beta distributions, is not in the Beta family with trust chains. The sole exception is the model where trust chaining returns the recommendation without modification, $S \otimes T = T$. Such a model (implicitly) asserts that all recommendations are always completely accurate. This model is neither theoretically interesting, as it can be formulated without trust chaining altogether, nor suitable in practice, as recommendations are not always completely accurate.

Corollary 8.12. *CertainTrust [Rie07] and TRAVOS [TPJL06] are not in the Beta family with trust chains.*

Proof. In CertainTrust and TRAVOS the result of a trust chain is always (isomorphic to) a beta distribution. □

TRAVOS is an interesting case, as the authors set out to do essentially the same as is done in this paper. Similar to this paper, they treat the Beta model formally (using random variables for the integrity, for the outcomes and the recommendations) and study the relation between honest recommendations and fake recommendations. However, TRAVOS asserts that the result of a trust chain (in their case called reputation) is a beta distribution. A similar argument holds for Subjective Logic:

Corollary 8.13. *Subjective Logic [Jøs97] is not in the Beta family with trust chains.*

Proof. In Subjective Logic the result of a trust chain is always (isomorphic to) a beta distribution. □

Hence, Subjective Logic breaks an assumption of the Beta model (on which it is based), or its operator dealing with recommendations (called discounting) does not actually model trust chaining. Both can be argued, since in Subjective Logic discounting is based on fuzzy logic, rather than distributions over integrity parameters, yet trust opinions and trust aggregation (called fusion) are based on the Beta model (i.e. based on distributions).

It is possible to alter Subjective Logic to incorporate a trust chaining operator such that it is isomorphic to a structure $S(\theta, \chi)$. However, the property of Subjective Logic that a trust opinion equates to a belief triple will no longer hold. Rather, a trust opinion will equate a weighted sum of belief triples, e.g. $\sum_i k_i(b_i, d_i, u_i)$. The fusion (trust aggregation) of two trust opinions $\sum_i k_i(b_i, d_i, u_i)$ and $\sum_j k'_j(b'_j, d'_j, u'_j)$ will then be $\sum_{i,j} k_i \cdot k_j((b_i, d_i, u_i) \oplus (b'_j, d'_j, u'_j))$, where $\oplus$ denotes unaltered fusion of belief triples from Subjective Logic. There are several valid variations for transitive trust operators (trust chains), and Proposition 8.4 shows that the operator need not be complicated. In Chapter 10, we formally introduce such a model.

8.5 Conclusion

We study a family of models based on the Beta distributions: the Beta family with trust chains. The models in that family are very similar to the Beta model, but more expressive. In particular, they can express trust chaining.

An important property, proven for all models in the Beta family with trust chains, is that trust chaining operations are modular (Proposition 8.8 and Theorem 8.9). So complicated trust opinions can be constructed by aggregating simpler trust opinions. Many existing trust models have asserted this property, which we now proved.

Another commonly asserted property in models inspired by the Beta model, is that all trust opinions can be represented as beta distributions. This property is disproved for models in the Beta family with trust chains (Theorem 8.11). This result implies in particular that Subjective Logic, TRAVOS and CertainTrust are not in the Beta family with trust chains (Corollaries 8.13 and 8.12).

We have proven that, up to isomorphism, every trust model in the Beta family with trust chains implicitly or explicitly makes assumptions about lying strategies and (except in special cases) about the entanglement (Corollary 8.10). Conversely, we have shown that, up to isomorphism, all trust models in the Beta family with trust chains can be constructed by selecting lying strategies and an entanglement (Corollary 8.10). Moreover, we have created a tool (Canephora) that calculates *chained trust opinions*, when instantiations of an entanglement and lying strategies are provided.

9

Quantifying Information from Recommendations

A challenge, when chaining trust opinions, is to know how *recommenders* act when they give fake *recommendations*, i.e., to know their lying strategy. We discuss this issue in detail in Section 9.2.

Part of the challenge of analysis of lying strategies, is to deduce which information is being leaked by which strategy. To reason about information leakage, we use techniques similar to those applied in differential privacy (e.g. in [Dwo06]). First, we illustrate the influence of lying strategies on a simplified example inspired by differential privacy:

Example 9.1. Assume a recommender witnesses a coin flip and the subject does not. If the subject wants to determine the result of the coin flip, he asks the recommender. Given the recommender's *integrity* parameter p, we know that with probability p he tells the truth about the outcome of the coin flip. However, with probability $1-p$, the recommender is not bound to the truth. In fact, then his goals is to provide us with as little overall information about the coin flip as possible. We have depicted the scenario in Figure 9.1. In the figure, c represents the bias of the coin, p the integrity parameter of the recommender and a and b the probability that he reports the correct result of the coin flip, provided he may lie. More specifically, if the recommender may lie and he sees heads ($F{=}h$), then he says heads ($S{=}h$) with probability a and tails ($S{=}t$) with probability $1-a$. If the recommender may lie and he sees tails ($F{=}t$), then he says tails ($S{=}t$) with probability b and heads ($S{=}h$) with probability $1-b$.

The resulting probabilities are summarised in Figure 9.2. The rows in the table correspond to the actual value of the coin flip, the columns to the statement of the recommender. The probability that the coin flip is heads ($F{=}h$), given that the recommender stated heads ($S{=}h$), is shown in the top left entry of the table. It is computed by adding the probability that the recommender is truthful (p) to the probability that the recommender is allowed to lie, but still chooses to state heads ($(1-p)a$). The remaining entries of the table are deduced similarly.

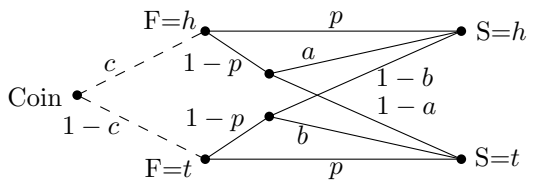

Figure 9.1: A coin flip and the choices of a recommender.

Actual ╲ Stated	S=h	S=t
F=h	$p + (1-p)a$	$(1-p)(1-a)$
F=t	$(1-p)(1-b)$	$p + (1-p)b$

Figure 9.2: Probabilities of a recommender's statements after a coin flip.

If the values of the row representing $F{=}h$ are equal to the values of the row representing $F{=}t$, the probability of a statement of the recommender is not correlated with the actual value of the flip. In this case the statements of the recommender do not leak any information. The two rows are equal if and only if $a+b = \frac{2p-1}{p-1}$. Since a, b and p represent probabilities, their values are in $[0,1]$, and thus the expression $\frac{2p-1}{p-1}$ has to be non-negative, for the equation to hold. This in turn limits the probability of the integrity parameter p to be in $[0, 1/2]$. If p takes values in $(1/2, 1]$, the two rows are not equal (i.e. $P(S{=}h|F{=}h) \neq P(S{=}h|F{=}t)$). Since it is not the case that $P(S{=}h|F{=}h) = P(S{=}h)$, there must be some correlation between statements (S) and actual values (F) and information leakage is unavoidable.

The intuition behind the example is straightforward. As long as we are allowed to lie at least half the time, we can lie exactly half the time. Then, since there are only two options (state the actual value, or state the opposite), we avoid giving away information. What our strategy should be in case information leakage is unavoidable has not yet been defined. In fact, what information and information leakage is, has not been formally defined. We define a measure of information (entropy) and information leakage in Section 9.3.

In Section 9.2, we revisit the game from Example 9.1 with a formal notion of entropy. More importantly, we propose a series of generalisations of that game. The final game presented in Section 9.2 reflects the game that the worst-case recommender plays in trust chains.

The utility function of the games in Section 9.2 is based on entropy, however, we have not defined which random variable we want entropy of. We show that there are different random variables to study information leakage of, and which are more interesting than others.

9.1 Entropy

We have identified that in Example 9.1, an optimal solution may not always exist. To quantify concepts such as information and information leakage, we introduce the notion of entropy, as in [McE01]:

Definition 9.1 (Entropy)**.** The entropy H of a discrete random variable X with possible values $x_1, ..., x_n$ for $n \in \mathbb{N}$ is given by $H(X) = \mathbf{E}(I(X))$, where $\mathbf{E}$ is the expected value and $I(X)$ is the random variable denoting the information content of X. Let p denote the probability mass function of X, then the entropy is:[1]

$$H(X) = \sum_{i=1}^{n} p(x_i)\, I(x_i) = \sum_{i=1}^{n} p(x_i) \log \frac{1}{p(x_i)}.$$

[1] In our considerations the base of the logarithm is not important.

If $p(x_i)$ is equal to 0 for some $i \in \{1, \ldots, n\}$ and $n \in \mathbb{N}$, then $p(x_i)\log(\frac{1}{p(x_i)})$ is taken to be 0.

Entropy can be extended to differential entropy, for continuous random variables X ranging from a to b, with probability density function f_X

$$h(X) = \int_a^b f_X(x)\log(\frac{1}{f_X(x)})\,\mathrm{d}x.$$

We extend the notion of entropy to a notion of conditional entropy as in [McE01].

Definition 9.2 (Conditional entropy)**.** The entropy H of a discrete random variable X given Y with possible outcomes $x_1, \ldots, x_n$ and $y_1, \ldots, y_m$ for $n, m \in \mathbb{N}$ is given by $H(X|Y) = \mathbf{E}(H(X|Y = y))$, where $\mathbf{E}$ is the expected value for Y and $H(X|Y = y)$ is the random variable denoting the information content of X under the condition $Y = y$. Let p and q denote the probability mass function of $H(X|Y)$ and Y, respectively, then the conditional entropy is:

$$H(X|Y) = \sum_{j=1}^{m} q(y_j)H(X|Y = y_j) = \sum_{j=1}^{m} q(y_j)\sum_{i=1}^{n} p(x_i|y_j)\log\frac{1}{p(x_i|y_j)}.$$

Conditional entropy can be extended for continuous random variables X and Y, by replacing the occurrences of entropy with differential entropy, i.e. $h(X|Y) = \mathbf{E}(h(X|Y = y))$, and by replacing sums with integrals.

Definition 9.3 (Information leakage)**.** Information leakage of a random variable Y about X is defined as $H(X) - H(X|Y)$ (or $h(X) - h(X|Y)$ for continuous X). If the base of the logarithm is 2, information leakage is given in bits.

Note that tables such as Figures 9.2, 9.3 and 9.4 only exist to provide an intuitive insight in information leakage. These tables have the property that the information leakage is 0 if and only if all rows are equal. Since information leakage not only covers the case that information leakage is 0, there is no formal reason to introduce these tables.

Finally, we define relative entropy, or Kullback-Leibler divergence:

Definition 9.4 (Relative Entropy)**.** Let X and Y be discrete random variables with the same set of outcomes $\{k_1, \ldots k_n\}$ for $n \in \mathbb{N}$. Let p and q be the probability mass function of X and Y, then the relative entropy between X and Y is:

$$D_{KL}(X||Y) = \sum_{i=1}^{n} p(k_i)\log(\frac{p(k_i)}{q(k_i)}).$$

The relative entropy is used to measure the information lost when Y is used to approximate X.

We use the notion of information leakage to define information games played between the subject and the recommender in Section 9.2 Exactly what this entropy represents, and which random variables we want to measure entropy of is discussed in Section 9.3

9.2 Information Games

For a concrete model of trust chaining, the *lying strategy* is required (see Corollary 8.5). Rather than assuming such a lying strategy (or using statistical data), in this section, we propose to apply game theory to find rational lying strategies. A comprehensive introduction into game theory can be found in [OR94]. We do not introduce game theory in its full generality. The games in this section (and their analysis) can be understood without prior understanding of game theory.

We first introduce a general game between the recommender and the subject. We then provide a series of increasingly complex instances of the game. The most complex instance will cover the basic trust chaining described in Definition 8.3 and explain how fake recommendations might, in fact, provide information.

The family of games has the property that they are zero-sum games, which means that the loss of one party (e.g. the recommender) is equal to the gain of the other party (e.g. the subject).

Definition 9.5 (The zero-sum recommendation game)**.** The *zero-sum recommendation game* is a family of games, parameterised with p, between the recommender and the subject where each player has exactly one move. First, the recommender makes a recommendation regarding a value. With a probability p, the recommender has no choice but to state the real value. With a probability $1 - p$, the recommender picks and states a value of choice. Then the subject must guess the value. In these games, the recommender tries to minimise information leakage whereas the subject chooses a strategy to obtain a maximal amount of information.

A mixed strategy allows a *user* to perform actions with a chosen probability. If the set of possible values in the zero-sum recommendation game is X, then a mixed strategy is a probability distribution over X, such that each element has a certain probability (density) of being recommended. In each game, it is evident what the pure options for the recommender are, and how this relates to the mixed strategy of the recommender.

In the zero-sum recommendation game, the recommender tries to minimise information leakage. This choice of strategy reflects a worst-case recommender. This assumption is debatable, since the recommender might actually attempt to skew the subject's opinion about the target positively (advertise the target) or negatively (smear the target). However, our choice provides an absolute minimum.

In Example 9.1 the optimal strategy was not to provide no information (set $a + b = 1$;[2]) when the recommender can lie, but to offset the number of times the recommender must tell the truth (set $a + b = \frac{2p-1}{p-1}$). In the following, we model the introductory example in the context of information games.

Game 9.1. (Coin flipping) In the coin flip game an recommender picks a probability distribution over recommendations, by picking two parameters a and b. This is his mixed move. Upon receiving the recommendation $S = s$, the subject assigns his estimated probability values to $F{=}h$ and $F{=}t$. Since the best estimate

[2]If $a + b = 1$ and the recommendation is fake, then the subject has no information. However there is a probability that the recommendation is not fake. Thus information is leaked, when the probability is non-zero.

of the value of the flip under recommendation s is $P(F|S=s)$, the entropy of the estimate is $H(F|S=s)$. The subject tries to minimise this entropy whereas the recommender wants to maximise it.

Recall that the goal of the recommender is to leak as little information as possible. Hence, if possible, the recommender chooses a and b in such a way that the two rows in Figure 9.2 are equal, thus setting $a+b=\frac{2p-1}{p-1}$. For $p > 1/2$, no solution exist, and we propose the use of information leakage, from Definition 9.2, to evaluate the choice of a and b. A good lying strategy (choice of a and b) minimises the information leakage of the statement S about the coin flip F. As the recommender does not control F, he can only minimise the information leakage, $H(F) - H(F|S)$, by choosing S. Effectively, the recommender must maximise $H(F|S)$, the conditional entropy of F under S.

As a formula, the expected conditional entropy is:

$$\begin{aligned} H(F|S) = {} & P(F{=}h,S{=}h)\log_2\left(\frac{P(S{=}h)}{P(F{=}h,S{=}h)}\right) + P(F{=}h,S{=}t)\log_2\left(\frac{P(S{=}t)}{P(F{=}h,S{=}t)}\right) \\ & + P(F{=}t,S{=}h)\log_2\left(\frac{P(S{=}h)}{P(F{=}t,S{=}h)}\right) + P(F{=}t,S{=}t)\log_2\left(\frac{P(S{=}t)}{P(F{=}t,S{=}t)}\right). \end{aligned}$$

We can instantiate the formula with values from Figure 9.2. Each cell gives the conditional probability $P(S|F)$, so we need to multiply the cells with $P(F)$, since $P(F,S) = P(S|F) \cdot P(F)$. The probability $P(F{=}h) = c$ and $P(F{=}t) = 1-c$. So we can instantiate $P(F{=}h, S{=}h) = c(p+(1-p)a)$, $P(F{=}h, S{=}t) = c(1-p)(1-a)$, $P(F{=}t, S{=}h) = (1-c)(1-p)(1-b)$, and $P(F{=}t, S{=}t) = (1-c)(p+(1-p)b)$, keeping in mind that $P(S{=}h) = P(F{=}h, S{=}h) + P(F{=}t, S{=}h)$. The recommender sets a and b (knowing c and p), maximising $H(F|S)$. The term reduces to $c\log_2(\frac{1}{c}) + (1-c)\log_2(\frac{1}{1-c})$, when we set $a+b=\frac{2p-1}{p-1}$, which is the a priori entropy of a coin flip $H(F)$. That means that the difference $H(F|S) - H(F) = 0$, and there is no information leakage, as expected in the analysis above without using entropy. Now, we not only have a formal way of expressing lack of information leakage, but also a quantification if information leakage does occur. If $p > 1/2$, the maximal expected conditional entropy occurs when the recommender picks $a = 0$, $b = 0$.

A real recommender in a trust chain has more than two options. Therefore, we now look at a simple zero-sum recommendation game where the recommender has k options:

Game 9.2. (The k-sided die) The setup of the game is similar to the one of Game 9.1. The only difference is that the move of the recommender is a distribution over k recommendations and the subject picks an estimate by assigning a probability to k different outcomes. In the game, the subject (recommender) tries to minimise (maximise) the entropy of his estimate, corresponding to the conditional entropy $H(X|S)$.

Rather than a coin flip, we can imagine a k-sided die. Instead of F, ranging over $\{h,t\}$, we have a random variable X ranging over $\{x_1,\ldots,x_k\}$. The fairness of the die is captured in $c_1,\ldots,c_k$, with $c_1+\ldots+c_k=1$, where c_i is the probability of $X=x_i$. The lying strategy is captured in $a_{i,j}$, $1 \le i,j \le k$, representing the probability of stating x_j ($S{=}x_j$), when x_i occurred ($X{=}x_i$). For $k=2$, this game is not essentially different from Game 9.1. Simply let heads and tails correspond

Is ╲ Stated	$S{=}x_1$	$S{=}x_2$	$S{=}x_3$
$X{=}x_1$	$p+(1-p)a_{1,1}$	$(1-p)(1-a_{1,2})$	$(1-p)(1-a_{1,3})$
$X{=}x_2$	$(1-p)a_{2,1}$	$p+(1-p)(1-a_{2,2})$	$(1-p)(1-a_{2,3})$
$X{=}x_3$	$(1-p)a_{3,1}$	$(1-p)(1-a_{3,2})$	$p+(1-p)(1-a_{3,3})$

Figure 9.3: Probabilities of recommender's statements after die throw.

to x_1 and x_2, use X rather than F, and let $a = a_{1,1}$, $b = a_{2,2}$ and $c = c_1$. Let $k = 3$, and we obtain Figure 9.3.

Again, the recommender tries to set the a values, such that all rows are equal:

$$\begin{aligned} a_{3,1} &= a_{2,1} = a_{1,1} + \frac{p}{1-p}, \\ a_{3,2} &= a_{1,2} = a_{2,2} + \frac{p}{1-p}, \\ a_{3,3} &= 1 - a_{3,2} - a_{3,1} = 1 - a_{1,1} - a_{2,2} - 2\frac{p}{1-p}. \end{aligned}$$

Like before, we have that any optimal solution must satisfy a restraint on the diagonal, $a_{1,1} + a_{2,2} + a_{3,3} = 1 - 2\frac{p}{1-p}$. Simple induction shows that the sum can be generalised for any k, to $a_{1,1} + \ldots + a_{k,k} = 1 - (k-1)\frac{p}{1-p}$.The right-hand side of the equation, $1 - (k-1)\frac{p}{1-p}$, is positive only when $p \leq 1/k$. A solution without information leakage, therefore, exists only if $p \leq 1/k$.

If $p > 1/k$, the recommender wants to minimise information leakage, thus maximise the conditional entropy. The conditional entropy $H(X|S)$ is:

$$\sum_{i,j} c_i((1-p)a_{i,j}\,[+p]_{\text{if } i=j}) \log_2 \left(\frac{c_j p + \sum_{1 \leq h \leq k} c_h (1-p) a_{h,j}}{c_i((1-p)a_{i,j}\,[+p]_{\text{if } i=j})} \right).$$

In the case $k = 2$, $p > 1/2$, Game 9.1 is actually an instance of this game. We know that $a_{1,1} + a_{2,2} = 0$ is the optimal strategy for the recommender. The obvious generalisation to cases $k \geq 3$ does not hold. For specific choices of p and c_i, a numerical analysis can show optimal solutions.

In trust chains that we study, targets are not k-sided dice. Rather, targets have an integrity parameter which determines the outcome of interactions. The recommendation, in the trust chain, regards this integrity parameter. So rather than a die with k sides, a target is a user with one of k possible integrity parameters:

Game 9.3. (The k-limited target.) Assume there is a target whose integrity parameter is an element of a set of values $\mathcal{X}$ (of cardinality k). A target with integrity x performs an action, which is a success, S, with probability x, and a failure, F, with probability $1 - x$. We use a random variable X to model the integrity parameter, E to model the action of the target.

Here, two games are possible. If we are interested only in the integrity parameter, the game is identical to Game 9.2. Alternatively, if the subject tries to estimate the outcome of the next interaction E with the target, rather than the integrity

Is ╲ Stated	$S{=}x_1$	$S{=}x_2$	$S{=}x_3$
$E{=}\textsc{s}$	$px_1 + \overline{p}(x_1a_{1,1} + x_2a_{2,1} + \ldots)$	$px_2 + \overline{p}(x_1a_{1,2} + x_2a_{2,2} + \ldots)$	$\ldots$
$E{=}\textsc{f}$	$p\overline{x_1} + \overline{p}(\overline{x_1}a_{1,1} + \overline{x_2}a_{2,1} + \ldots)$	$p\overline{x_2} + \overline{p}(\overline{x_1}a_{1,2} + \overline{x_2}a_{2,2} + \ldots)$	$\ldots$

Figure 9.4: Probabilities of recommender's statements about a target.

parameter X, then the move of the subject is to pick S or F (and the move of the recommender remains unchanged). The alternative is essentially saying that the integrity of the target is irrelevant by itself, but only relevant because it determines the probability of success, which is deemed inherently interesting. In the variant where we are inherently interested in the integrity, the utility function is still $H(X|S)$, in the version where we are inherently interested in the outcome of the next interaction, it is $H(E|S)$.

Using $\overline{\varphi}$ for $1-\varphi$, we obtain Figure 9.4. If the recommender picks $a_{i,j}$ as before; $a_{i,j} = a_{i,i} + \frac{p}{1-p}$ and $a_{1,1} + \ldots + a_{k,k} = 1-(k-1)\frac{p}{1-p}$, then it is easy to see that both rows are equal. In other words, if the subject gains no information about X, he gains no information about E. If $p \leq 1/k$, no information need be leaked about X, and thus nor about E. The converse does not necessarily hold, as even if $p > 1/k$.

The conditional entropy of E under S, $H(E|S)$ is:

$$\sum_i P(S{=}x_i,E{=}\textsc{s})\log_2\left(\frac{P(S{=}x_i)}{P(S{=}x_i,E{=}\textsc{s})}\right) + \sum_i P(S{=}x_i,E{=}\textsc{f})\log_2\left(\frac{P(S{=}x_i)}{P(S{=}x_i,E{=}\textsc{f})}\right).$$

where $P(S{=}x_i, E{=}\textsc{s}) = \sum_j x_j c_j a_{j,i}$ and $P(S{=}x_i, E{=}\textsc{f}) = \sum_j (1-x_j)c_j a_{j,i}$. Although $H(X) - H(X|S) = 0$ implies $H(E) - H(E|S) = 0$, $H(X|S) \neq H(E|S)$.

We have two variants of the k-limited target game, differing in their utility function, $H(X|S)$ versus $H(E|S)$. Which utility function is more suitable depends on the interest of the subject. Note, however, that $H(X|S)$ is a more discriminatory measure, as $H(X) - H(X|S) = 0$ implies $H(E) - H(E|S) = 0$. That means that if the values of p and all $a_{i,j}$ are set in such a way that all rows are equal in Figure 9.3 then both rows in Figure 9.4 are equal. However, the converse does not hold. For example, if, for every x_i there exists x_j such that $x_i = 1 - x_j$, then a trivial optimal solution exists for values of p up to $1/2$. By restricting the options to saying x_i or x_j, the solution is the same as in Game 9.1, $a_{i,i} + a_{j,j} = \frac{2p-1}{p-1}$, with $a_{i,j} = 1 - a_{i,i}$ and $a_{j,i} = 1 - a_{j,j}$. This strategy leaks a lot of information about X, as it restricts it to two values, but no information about E.

The choice of random variable to measure information over both influences the optimal strategy, and the measuring of suboptimal strategies. Contrary to the definition of Game 9.3, the recommender does not know the integrity parameter, but rather has a certain trust opinion on the target. The shape of the recommendation is, therefore, not an claim of an integrity parameter, but a claim of a trust opinion.

Game 9.4. (Basic trust chaining game) A game similar to Game 9.3 can be defined for basic trust chaining, defined in Definition 8.3. The set of values is a subset of $\mathbb{N} \times \mathbb{N}$, since not every pair (s, f) is necessarily a valid recommendation. In particular, if $\lambda(s + f) = 0$ (i.e. the probability of $s + f$ interactions is zero), then

(s, f) is not a valid recommendation. The set of possible recommendations is therefore $\{(s, f)|s, f \in \mathbb{N}, \lambda(s + f) > 0\}$.

If the size of the set of possible recommendations is k, then there is an obvious optimal lying strategy for $p < 1/k$, for reasons detailed in Game 9.2. If the subject is interested in finding out the real *interaction history*, O, of the recommender the optimal strategy works only in these cases. However, the subject is not interested in O, but in information about the target (e.g. X or E from Game 9.3). Since X and E are less discriminative than O, similar to how E was less discriminative than X in Game 9.3, there may exist optimal strategies for X and E which are not optimal for O. We can immediately observe that if the recommender only leaks information about the size of his real opinion, he leaks no information about the target. That means that for X and E, we merely require two rows to be equal, when the size of the recommendations are equal (e.g. we can distinguish $(1, 0)$ from $(1, 1)$ but not from $(0, 1)$). Hence, we can identify a class of optimal lying strategies for $p < \frac{1}{k+1}$, where k is the maximal value for $\lambda k > 0$.

Depending on the exact choice of information of interest (e.g. for E), the class of optimal lying strategies may be even larger.

We can find the optimal strategy for both players in all instances of Game 9.1, find the optimal strategy of the recommender in Games 9.2, 9.3 and 9.4 under some circumstances, and deduce the optimal strategy of the subject in Games 9.1–9.4 given the strategy of the recommender.

The zero-sum information game is chosen in such a way that the optimal strategy for the recommender is the worst-case scenario for the subject. Formally, we can state that the Nash equilibrium of Game 9.4 is interesting to find, since it provides the optimal strategy of the subject in the worst-case scenario. If, in reality, the recommender has another goal than to minimise information leakage, then the subject is guaranteed to gain information. This follows from the fact that if the opponent unilaterally deviates from the strategy in the Nash equilibrium, the opponent loses utility, which means you gain utility in a zero sum game.

In the section, we saw that the notion of information leakage was to general to be used in the definition of the zero-sum recommender games. We needed to reason about the information leakage of a particular random variable.

9.3 Utility

We saw that the choice of representation of information has a real impact on the strategies of the recommender. The choice is, in a sense, subjective. We look at four different choices. The first choice being the most straightforward: the number of experiments directly represents the amount of information. The second and third have been mentioned in Section 9.2, namely the entropy of the integrity parameter and the outcomes of interactions. Finally, we introduce an information measure that combines the positive aspects of these measures.

9.3.1 Interactions

The most straightforward measure of information is the number of interactions required for the probability distribution. In the case that the trust opinion is a *beta distribution*, the number of interactions required is obvious. It is less obvious if the trust opinion is a sum of different beta distributions, as is common for chained trust opinions (Theorem 8.11).

We need a measure that provides the right answers for beta distributions, that provides a unique answer for a distribution, and that provides intuitive answers for values between two beta distributions. The uniqueness requirement is obvious, but not trivial. Considering that $\beta(1,1) = 0.5 \cdot \beta(2,1) + 0.5 \cdot \beta(1,2)$, the measure must be equal for both sides.

We can construct a measure based on the notion that for every s, f, there are a, b such that $\beta(s+1, f+1) = a \cdot \beta(s+2, f+1) + b \cdot \beta(s+1, f+2)$. Using that notion, it is possible to convert every weighted sum of beta distributions $\sum_i w_i \cdot \beta(s_i+1, f_i+1)$, with $s_i + f_i \leq n$, into a sum $\sum_j v_j \cdot \beta(s'_j+1, f'_j+1)$ with $s'_j + f'_j = n$. For example, $0.5 \cdot \beta(1,1) + 0.5 \cdot \beta(2,1) = 0.75 \cdot \beta(2,1) + 0.25 \cdot \beta(1,2)$. We can use the total difference $\sum_j |\frac{1}{n+1} - v_j|$ as a basis for the measure, as it has the property that it is invariant over increases of n. More precisely, we need to multiply that with $\frac{n+1}{2}$, to normalise it to number of experiments, making $\frac{n+1}{2} \cdot \sum_j |\frac{1}{n+1} - v_j|$. Although this measure adheres to the three properties we required, it does have some counterintuitive properties. Furthermore, it can only be applied to finite weighted sums of beta distributions.

There is a better measure that is congruent to measuring the number of experiments, based on entropy measures, which we explore in the next section.

9.3.2 Entropy

A standard way of representing information content is entropy. Contrary to the method based on interactions mentioned above, it generalises neatly to other distributions. Furthermore, the games from Section 9.2 are designed to operate using entropy measures. If we want to use the notion of entropy to represent the information content of a trust opinion, then we need to select a random variable. We have seen that selecting the relevant random variable is neither obvious nor inconsequential. There are two obvious candidates, the (continuous) random variable representing the integrity of the target (i.e. the random variable that is distributed), or the (discrete) random variable representing the next outcome of the target. We evaluate both candidates below and then study a combined measure.

Entropy of Integrity If we are interested in information regarding the reliability of target C, the natural choice is to look at the entropy of the random variable representing that; R_C. However, we may not actually be interested in the exact value of R_C. Whether R_C is equal to 0.501 or 0.502 may not be important in reality, yet information-wise these two values are treated as completely different values. Intuitively, using R_C is too discriminative.

An interesting scenario arises, if we pick R_C to be the measure of information. If a

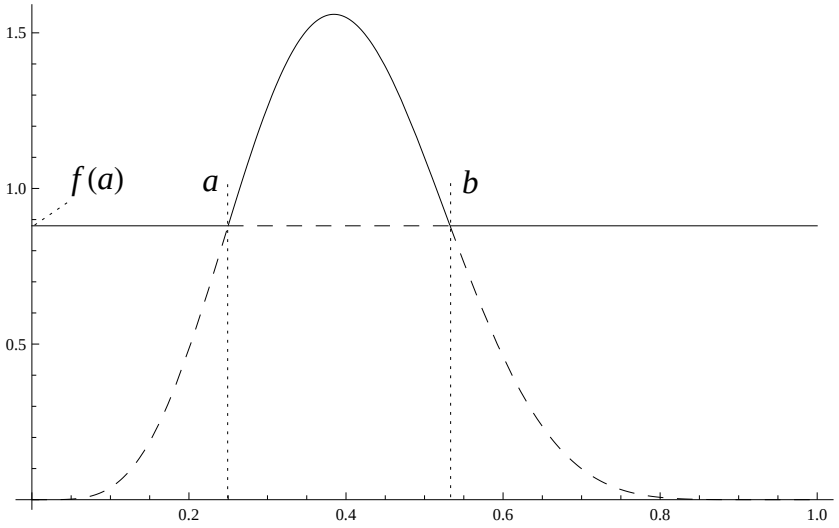

Figure 9.5: A beta distribution $\beta(6, 9)$ multiplied with 0.5. The line $f(a)$ has the property that $\max(f(a), \beta(6, 9))$ has a surface area of 1 in $[0, 1]$.

recommender is forced to tell the truth, T, with a certain probability, p, and can lie, L, otherwise, the information content of his recommendation can be characterised as $p \cdot T + (1 - p) \cdot L$. In Figure 9.5, we can see an example, where $p = 0.5$ and the truth T is $\beta(6, 9)$, and the lying strategy L is chosen to maximise the total entropy of the graph. If we can algebraically compute this graph, we can find the optimal lying strategy in a generalised case of Game 9.4. For us, an open question is whether we can generally find exact values for a and b, which is required for finding L.

Intuitively, the entropy of R_C is high for "flatter" graphs (e.g. the uniform distribution has maximal entropy). Graphs that are strongly concentrated around a certain value have low entropy, thus more information than flatter graphs. This correlates with what we expect from a measure of information. Hence, the entropy of integrity is a suitable measure of entropy, although not perfect as it may be considered too discriminative.

Entropy of Outcome of Interaction Naively, we may argue that we are not interested in the (exact) value of R_C. Rather that we are ultimately interested in E_C, the probability of success in an interaction, since the point of creating a trust opinion is for the subject to decide whether or not to enter the next interaction, which is determined by the likelihood of success. If we are interested in information regarding the next outcome of C, we should study the entropy of the random variable representing the next outcome; E_C. Recall that $P(E_C = \mathrm{s}|R_C{=}0.5) = 0.5$, meaning that the outcome is equivalent to a fair coin flip, which has 1 bit entropy. However, R_C is not a fixed number, but has a distribution, hence we should take the expectation of the entropy of E_C: $\int_0^1 f_{R_C}(x) \cdot H(E_C|R_C = x)\,\mathrm{d}x$. If R_C happens to be somewhere around 0.5, then increasing the number of interactions is increasing the entropy too (decreasing the information). This measure of entropy on the distribution based on $(0, 0)$ gives 0.721348 bits of entropy, but $(100, 100)$ gives 0.996402 bits of entropy. We lose $0.996402 - 0.721348 = 0.275054$ bits of in-

formation (about the next outcome), when we have more interactions. According to [ZMLM11] (as well as intuition), the utility (i.e. information measure) should increase when more data is available, however we lost bits when we accrued 200 data points.

Despite the intuition that E_C is more relevant for our interests than R_C, the measure performs unsatisfactory. Distributions based on many interactions can perform much worse than the uniform distribution. The measure should reflect that gaining one success and one failure increases information, which this measure does not.

Combined Measure of Entropy Now we combine the intuitive notion that E_C is more important than R_C, with a suitable information measure (as R_C provides), with the notion from Section 9.3.1.

The problem with studying the expected entropy of E_C (i.e. $\mathbf{E}(h(E_C))$), is that it is strongly biased around 0.5. Let B_x and B_y be *Bernoulli distributed* random variables with parameters x and y, respectively. In other words, B_x is E_C given $R_C = x$ and B_y is R_C given $R_C = y$. Then we can study the expected relative entropy between B_x and B_y, given x, y as outcomes of R_C, $\int_0^1 \int_0^1 f_{R_C}(x) \cdot f_{R_C}(y) \cdot D_{KL}(B_x||B_y)\,\mathrm{d}x\,\mathrm{d}y$. Essentially, this measures the entropy of the outcome of the interaction with a randomly (according to the distribution f_{R_C}) selected reliability x, relative to the outcome of an interaction with the true machine with unknown reliability y (distributed with f_{R_C}). A concrete interpretation would be to say that there is a true integrity x, and an approximated integrity y, and we want to know the relative entropy of between these two values. The probability density that the true integrity is x with probability density $f_{R_C}(x)$, and the approximation is y with probability density $f_{R_C}(x)$.

This entropy measure does not have the same problem as the entropy measure based on R_C, that it is too discriminative (e.g. seeing 0.501 as completely different from 0.502). And it also does not have the same problem as the entropy measure directly based on E_C, that it does not correlate with the number of interactions. In fact, an important characteristic of this approach is that correlates perfectly with the number of experiments. The beta distributions based on $(5, 5)$ and on $(10, 0)$ have the same information content in this measure. Moreover, the entropy of the beta distribution based on s successes and f failures equals $\frac{1}{s+f+2}$. Recall that entropy is the converse of information, so if $s + f$ increases, we expect the entropy to decrease towards 0. The entropy of an arbitrary probability distribution can now trivially be transformed to an equivalence in interactions by taking the multiplicative inverse and subtracting two. Hence, this entropy measure combines the best of both worlds. It does, however, have the peculiar side effect that the uniform distribution does not provide the minimal entropy (but $f_{\mathrm{B}}(0, 0)$ does).

Whether the expected relative entropy between two different interactions distributed via the relevant distribution is superior to simply taking the entropy of the distribution is debatable. Both measures have the right formal properties to function as an measure for information leakage, making the choice subjective.

9.4 Conclusion

On the basis of the fact that lying strategies are the big unknown in trust chains, we studied these lying strategies in this chapter.

In Section 9.2, we introduced four games with increasing complexity. Each of these games was a zero-sum recommender game. The first game is a simple game where the recommender tries to hide the result of a coin-flip, despite the fact that he must tell the truth a fraction of the time. The fourth game is sophisticated enough to encompass a recommender that tries to hide information on the target, despite the fact that he must provide his trust opinion a fraction of the time. We noticed that there is a generic class of optimal solutions for any notion of information about the target. We also noticed that for cases outside of that class, the optimal solution depends on the measure of information.

Therefore we studied several measures of information in Section 9.3. The first measure had no basis in entropy and did not generalise to all trust opinions, and therefore was not a suitable for the games, despite its intuitive merits. The second measure was the entropy of the integrity parameter. This measure provided intuitively satisfying results for trust opinions, but was deemed too discriminative. The third measure was the entropy of the outcome of the next interaction. The subject is interested mostly in the next interaction, however, the measure is not sufficiently discriminative and provided unsatisfactory measurements for trust opinions. The last measure combines the positive aspects of the three different methods.

Future work We have not found the general optimal strategy for the recommender in Games 9.2-9.4. It is interesting to find the optimal solutions for both the measure based on entropy of integrity and the combined measure, as described in Section 9.3. Given the optimal strategy for the recommender, the subject can assign χ^B to equal that optimal strategy. The subject is guaranteed not to lose information when the recommender changes strategy. Formulated differently, the subjects choice maximises the worst-case information gain.

Another interesting direction is to study different types of games. If we assert a specific goal for the recommender (e.g. to advertise specific targets), the game becomes asymmetrical. There could be scenarios with implicit cooperation, where the recommender provides a overly positive recommendation close to his real opinion when lying, and the subject accepts the recommendation with high confidence, since it is at least close to the truth when false.

Finally, we can model attacks on systems with recommendations on recommenders. We can assume a system where recommenders are recommended themselves (like the WoT). There can be a cluster of malicious recommenders that all recommend each other positively. Their goal is to make themselves credible. Such a goal can be encoded as a utility function, to define games over lying strategies with.

10

A Generic Extension of the Beta Model

In Chapter 6, we presented the Beta model. The Beta model is a *trust model* that only supports the *trust aggregation* operation. We have added the *logical trust operations*, in Chapter 7, which immediately yields a new model, namely the Beta model with logical trust operations. In Chapter 8, we added trust chaining to the Beta model. Such an extension yielded more than one possible model. We provided the entire family of Beta models with trust chains, and a straightforward way of selecting such a model, namely setting the *entanglement* and lying strategies. In this chapter, we select a particular model of trust chaining and extend the Beta model with both the logical trust operations and trust chaining.

The formulae for the logical trust operations (e.g. in Theorem 7.4) hold under arbitrary circumstances. The shape of the composite trust opinions, therefore, does not alter when we allow trust chaining. Together with the modularity results (e.g. in Theorem 8.9), this seems to imply that the two models (the *Beta model with logical trust operations* and a member of the *Beta family with trust chaining*) can be merged seamlessly. One of the goals of this chapter, is to show the relative ease of merging the two extensions of the Beta model into one model, called the Default Model. The construction of the Default Model is formal, meaning that we show how to formally derive a model with non-trivial expressivity. Additionally, it carries the implication that formal trust models can be applied to practice, rather than just exist as theoretical entities of which we proved their existence.

Besides showing the bearing of our theory on practice, a goal of this chapter is to combine the methodology of Part I, i.e. the axiomatic method, with the methodology of Part II, i.e. the probabilistic method. In particular, in Chapter 5, the axioms for dilution – the operator for trust chaining – encoded the assumption that lies contain no information. It is immediate that for arbitrary lying strategies, lies do contain information – in fact, they do, whenever lies correlate (positively or negatively) with the truth. However, for some specific lying strategies, characterised in Section 10.1.3, the assumption that lies contain no information is warranted. The axioms for dilution also encoded the assumption that every statement is equally likely a lie. Again, this is not the case for arbitrary lying strategies, but is the case for some lying strategies. The model that we study, the Default Model, selects a *lying strategy* with both properties.

The Default Model is chosen for its mathematical properties. One property, as just mentioned, is that lies carry no information. Moreover, every statement is equally likely to be a lie. Another property is that the entanglement has no bearing on the model. In other words, the model correctly models situations with arbitrary entanglements. Yet another property is that the representation of trust opinions is relatively simple. Most generally, a basic trust chain of finite summations of

beta distributions may yield an infinite summation of beta distributions. In our case, however, these summations remain relatively simple to express, because they are finite.

We divide this chapter in three sections. First, in Section 10.1 we define the model, and study its properties using techniques from Chapter 9. Then, we look at alternative representations in Section 10.2. Finally, in Section 10.3, we look at the application of the axioms from Chapter 5.

10.1 The Default Model

The model that we select is called the Default Model. The Default Model is chosen for its mathematical properties, as mentioned before. In order to rigourously discuss the mathematical properties, we need to formally define the Default Model. In this section, we refine the techniques introduced for other models, and use it to define the Default Model.

This section is subdivided in four parts. First, we discuss the syntax of the expressions in the Default Model, and the corresponding trust networks that may appear in the model as a consequence. Second, we reuse techniques from Chapters 6, 7 and 8 to provide the tools to define the semantics. Third, we provide the semantics of the expressions in the Default Model, based on the informal requirements on the model. Fourth, we analyse the model using techniques from Chapter 9.

10.1.1 Syntax

In Part I, we dealt with an axiomatic approach to trust. In that part, signatures were one of the central parts of the analysis. A signature defines the set of relevant expressions, and therefore defines which trust networks we reason over. The set of expressions that we allow in the model, closely matches those described in Section 10.2. The main difference is that trust chaining must be a postfixing operation, for reasons detailed later.

It is important that, in this analysis, we do not unnecessarily exclude certain expressions (i.e. certain trust networks). Assert that there is a trust network for which we can compute its *trust opinion* using techniques from Part II, but the expression that represents the trust network is not included. Then there exists a more general model based on the same techniques. Ideally, we want to have the most general model based on the techniques outlined in Part II.

On the other hand, we do not want to unnecessarily include certain expressions. In particular, expressions that represent a trust network where *users* appear multiple times[1] are undesirable, as e.g. the requirement of independence in Section 7.1. Further, expressions that represent a trust network with a *recommendation* should not allow this recommendation to be anything other than a *simple trust opinion* as

[1] If we have a network $x \wedge y$, but the opinion is about $A \wedge A$ (rather than $A \wedge B$), then our opinion about A is not necessarily well-defined (it could be x or y). Similarly, if we have $(x \cdot y) \cdot z$, and the first intermediate is B, the second intermediate is C, and the target is B, then our opinion about B is not necessarily well-defined (it could be x or $(x \cdot y) \cdot z$.

we assert that recommendations are shaped like interaction histories in Section 8.1. Finally, an aggregation of *composite trust opinions* cannot be analysed, as $x \wedge y$ is an opinion about a conjunct target, and $x' \vee y'$ is an opinion about a disjunct target, and the aggregate $x \wedge y + x' \vee y'$ must suppose both opinions are regarding the same target.

Taking the syntax and interpretation as formulated in Section 10.2, we naturally exclude trust networks where users appear more than once. Let $\mathcal{B}$ be the set of simple trust opinions. Recall the syntax, for simple trust opinions $x \in \mathcal{B}$:

$$\varphi ::= x | \varphi + \varphi | \varphi \cdot \varphi | \varphi \wedge \varphi | \varphi \vee \varphi | \overline{\varphi},$$

and the graphical and informal interpretations of the expressions. In the interpretations of the expressions, users are anonymous in the sense that the *subject* and the *target* are assigned arbitrary letters, and all intermediate users are assigned fresh letters. Therefore, none of the expressions can represent a network where a user appears at more than one place. (Note that opinions can appear in more than one place, such as $x \wedge x$, where both instances of x represent the same opinion about different users.) Hence, a model based on this syntax should allow us to apply the techniques from Chapter 7. However, expressions such as $x \cdot (y \cdot z)$, $x \cdot (y \wedge z)$ and $(x \wedge y) + z$ are allowed in this syntax. The first two expressions contain a recommendation which is not of the shape assumed in Chapter 8 – namely that recommendations are simple trust opinions, which $y \cdot z$ and $y \wedge z$ are not – thus not allowing us to apply the techniques from that chapter. The third expression aggregates different targets, which we cannot express using our random variables. We need to restrict the syntax.

The right-hand side of a trust chain $\varphi \cdot \psi$ represents the recommendation – ψ. The assumption in Chapter 8 is that recommendations are pairs of natural numbers which represent alleged interaction histories. Recall that simple trust opinions map one-to-one to pairs of natural numbers (s, f), via $\vartheta_{s,f}$ (recall that $\vartheta_{s,f}(x) = f_{\mathrm{B}}(x; s+1, f+1) = \mathrm{NF} \cdot x^s \cdot (1-x)^f$). The implication is that we need to enforce that ψ can only be a simple trust opinion. Hence, we can formulate trust chaining as a postfix operator. For simple trust opinions $x \in \mathcal{B}$, we define:

$$\varphi ::= x | \varphi + \varphi | \varphi \cdot x | \varphi \wedge \varphi | \varphi \vee \varphi | \overline{\varphi}.$$

However, this syntax still allows fusion of trust opinions regarding different users (e.g. $(x \wedge y) + z$).

To obtain the largest set of expressions, such that we can straightforwardly apply the techniques from Chapters 7 and 8, we need to reason about the anonymous users that form the targets. The expression $x + y$ is meaningful, only if x and y are trust opinions about the same target. If their targets are composite targets, say S and T, this is problematic. If $S = T$, then users appear more than once in a trust network. Therefore, we can assume without loss of generality that $S \neq T$. In the case $S \neq T$, x and y are trivially trust opinions about different targets (namely S and T), making $x + y$ is meaningless. In either case, trust aggregation over composite targets is meaningless. Therefore, we can distinguish trust opinions for which aggregation make sense, from those for which it does not. That leads us to the definition of the syntax of the Default Model:

Definition 10.1 (Syntax of the Default Model)**.** The set $\mathcal{B}$ is the set of simple trust opinions. For $x \in \mathcal{B}$, we define the language:

$$\varphi := x|\varphi + \varphi|\psi \cdot x$$

$$\psi := \varphi|\psi \wedge \psi|\psi \vee \psi|\overline{\psi}$$

The set $\mathcal{P}$ of trust opinions about simple targets contains exactly those expressions defined by φ. The set $\mathcal{D}$ of trust opinions contains exactly those expressions defined by ψ.

The set $\mathcal{D}$ contains all expressions in the syntax of the Default Model.

It is immediate that $\mathcal{B} \subset \mathcal{P} \subset \mathcal{D}$. Note that $\mathcal{P}$ includes a trust chain of a composite target and a simple target. Say a composite *recommender* makes a recommendation, that recommendation must still be a simple trust opinion. Therefore, the target of the *chained trust opinion* is a simple target. Thus we can place $x \cdot y$ (for $x \in \mathcal{D}$, $y \in \mathcal{B}$) in $\mathcal{P}$, and seeing we are looking for the most largest set on which our techniques apply, we must place $x \cdot y$ in $\mathcal{P}$.

10.1.2 Techniques

The semantics of an expression are intuitively straightforward. However, some technicalities need to be straightened out first. In particular, the random variables for interaction histories represent the entire *interaction history*, making it difficult to express trust aggregation of two simple trust opinions. Another technical issue, is the fact that the assumptions of the *Beta model* with composite trust and the Beta family with trust chains are different. We must integrate the assumptions in such a way that none of the theorems are invalidated. The last technicality, is the fact that in Theorem 8.1, the opinion on the recommender is assumed to be a simple trust opinion, whereas we syntactically allow arbitrary trust opinions.

The reason that a random variable for an interaction history, say O_C^A, represents the entire interaction between A and C is for convenience in reasoning about trust chaining. Having a single, complete interaction history as a random variable simplifies the notion of what constitutes a truthful recommendation and what constitutes a lie. The simplification of reasoning about trust chaining comes at the expense of the simplicity of trust aggregation. The problem is, therefore, that we do not have the right random variables to denote the trust aggregation of two simple trust opinions. We do not have two different random variables that both capture a fraction of the interaction history between A and C, we only have O_C^A. Although we could introduce new random variables to allow a direct formulation of trust aggregation of simple trust opinions, the deeper insight is that trust aggregation of simple trust opinions is not strictly necessary. The trust aggregation of all fractions of the interaction history between A and C is exactly the interaction history between A and C. Therefore, we can substitute any trust aggregation of simple trust opinions by a single simple trust opinion. Moreover, we can assume without loss of generality that expressions do not contain aggregation of simple trust opinions.

To allow the other operations, trust chaining and the logical trust operations, we need both the assumptions set out in Chapter 7 and those in Chapter 8. To

integrate these sets of assumptions, we must take care that all Dependencies $\mathrm{D1}_\ell$–$\mathrm{D6}_\ell$, Dependencies $\mathrm{D1}_S$–$\mathrm{D5}_S$, Independencies $\mathrm{I1}_\ell$–$\mathrm{I3}_\ell$ and Independencies $\mathrm{I1}_S$–$\mathrm{I4}_S$ follow from the assumptions we formulate below. We must also take care that we have the right random variables to deal with certain expressions. In particular, we must allow statements from composite users (e.g. $S_C^{A\wedge B}$), and therefore observation histories of composite users (e.g. $O_C^{A\wedge B}$). Hence, in addition to E_S and R_S, we also have O_C^S and S_C^S, for $C \in \mathbf{A}$ and $S \in \mathbf{T}$.[2]

It is convenient to define the collection of all random variables, given that we have $\mathbb{E}^+$ and $\mathbb{R}^+$, we furthermore define:

$$\begin{aligned}\mathbb{O}^+ &:= \{O_C^S : S \in \mathbf{T}, C \in \mathbf{A}\},\\ \mathbb{S}^+ &:= \{S_C^S : S \in \mathbf{T}, C \in \mathbf{A}\},\\ \mathbb{W}_\mathbb{D} &:= \mathbb{E}^+ \cup \mathbb{R}^+ \cup \mathbb{O}^+ \cup \mathbb{S}^+.\end{aligned}$$

Note that $\mathbb{W}^+ \subset \mathbb{W}_\mathbb{D}$ and $\mathbb{W}_\mathbb{S} \subset \mathbb{W}_\mathbb{D}$ follow immediately.

We can therefore formulate the dependencies and independencies:

$\mathrm{D1}_D$ R_C is uniformly distributed on $[0,1]$.

$\mathrm{D2}_D$ $P(E_T{=}\mathrm{s}|R_T{=}p) = p$.

$\mathrm{D3}_D$ $P(O_C^A{=}(x_s,x_f)|R_C{=}c) = \binom{x_s+x_f}{x_s} c^{x_s}(1-c)^{x_f}\lambda(x_s+x_f)$.

$\mathrm{D4}_D$ $E_{S\wedge T} = \mathrm{s}$ iff $E_S = \mathrm{s}$ and $E_T = \mathrm{s}$, for $\mathrm{act}(S) \cap \mathrm{act}(T) = \emptyset$.

$\mathrm{D5}_D$ $E_{S\vee T} = \mathrm{s}$ iff $E_S = \mathrm{s}$ or $E_T = \mathrm{s}$, for $\mathrm{act}(S) \cap \mathrm{act}(T) = \emptyset$.

$\mathrm{D6}_D$ $E_{\neg T} = \mathrm{s}$ iff $E_T = \mathrm{f}$.

$\mathrm{D7}_D$ There exist functions f, g, h, with $R_{S\wedge T} = f(R_S, R_T)$ and $R_{S\vee T} = g(R_S, R_T)$ when $\mathrm{act}(S) \cap \mathrm{act}(T) = \emptyset$, and $R_{\neg T} = h(R_T)$.

$\mathrm{D8}_D$ $P(S_C^B{=}(w_s,w_f)|E_B{=}\mathrm{s}, O_C^B{=}(w_s,w_n)) = 1$.

$\mathrm{D9}_D$ $P(S_C^B{=}(y_s,y_f)|E_B{=}\mathrm{f}, R_B{=}b, O_C^B{=}(w_s,w_f)) = \chi^B(b,w_s,w_f)(y_s,y_f)$.

$\mathrm{I1}_D$ For $W \in \mathbb{W}_\mathbb{D} \backslash \{O_C^T, S_C^T : T \in \mathbf{T}, \mathrm{act}(S) \cap \mathrm{act}(T) \neq \emptyset\}$, $O_C^S \perp\!\!\!\perp W|R_C$ holds.

$\mathrm{I2}_D$ For $W \in \mathbb{W}^+ \backslash \{R_S : B \in \mathrm{act}(S), R_S \in \mathbb{R}^+\}$ and $\{C, D_0, \ldots D_n\} = \mathbf{A}$, it holds that $R_C \perp\!\!\!\perp W|E_C, O_C^{D_0}, \ldots, O_C^{D_n}$.

$\mathrm{I3}_D$ For $W \in \mathbb{W}_\mathbb{S} \backslash \{E_S, S_D^S : D \in \mathbf{A}, S \in \mathbf{T}, B \in \mathrm{act}(S)\}$, $E_B \perp\!\!\!\perp W|R_B$ holds.

$\mathrm{I4}_D$ For $W \in \mathbb{W}_\mathbb{D} \backslash \{S_C^S\}$, it holds that $S_C^S \perp\!\!\!\perp W|E_S{=}\mathrm{f} \cap R_S \cap O_C^S$.

That the aforementioned assumptions imply the assumptions of the Beta model, the Beta model with logical trust operations and the Beta family with trust chaining is immediate. Moreover, if we restrict $\mathbb{W}_\mathbb{D}$ to $\mathbb{W}^+$ or $\mathbb{W}_\mathbb{S}$, these assumptions are identical to the assumptions of the Beta model with logical trust operations

[2] Recall that $\mathbf{T}$ is the closure of $\mathbf{A}$ over $\wedge$, $\vee$ and $\neg$.

and the Beta family with trust chains, respectively. Note that that statement is strictly stronger than the statement preceding it. The Default model is strictly stronger than the Beta model with logical trust operations and any element in the Beta family with trust chains:

Proposition 10.1. *The assumptions Dependencies $D1_D$-$D9_D$ and Independencies $I1_D$-$I4_D$ imply Dependencies $D1_\ell$-$D6_\ell$ and Independencies $I1_\ell$-$I3_\ell$, and the assumptions Dependencies $D1_D$-$D9_D$ and Independencies $I1_D$-$I4_D$ imply Dependencies $D1_S$-$D5_S$ and Independencies $I1_S$-$I4_S$.*

Proof. Each of Dependencies $\mathrm{D1}_\ell$-$\mathrm{D6}_\ell$, Dependencies $\mathrm{D1}_\ell$-$\mathrm{D3}_\ell$, Independencies $\mathrm{I1}_\ell$-$\mathrm{I3}_\ell$ and Independencies $\mathrm{I1}_S$-$\mathrm{I4}_S$ is trivially implied by the dependency or independency with the same index. Dependencies $\mathrm{D4}_S$ and $\mathrm{D5}_S$ are trivially implied by Dependencies $\mathrm{D8}_D$ and $\mathrm{D9}_D$. □

As a consequence of Proposition 10.1, all Theorems proven in Chapters 6, 7 and 8 remain valid in the Default Model.

To solve the last technicality before formulating the model, we must generalise Theorem 8.1. That theorem must hold for an arbitrary trust opinion of the subject, A, about the intermediate, B, and not just simple trust opinions. In particular, in $\mathbf{eq^2}$, the terms $(x_s{+}1)$ and $(x_f{+}1)$ represent the expected value of the simple trust opinion $\vartheta_{x_s,x_f}(b) = f_{\mathrm{B}}(b; x_s{+}1, x_f{+}1)$, and should be altered to reflect the expected value of an arbitrary trust opinion. Similarly, in $\mathbf{eq^5}$, the term $f_{\mathrm{B}}(b; x_s+1, x_f+2)$ must be updated. Hence, we can formulate the generalisation of Theorem 8.1 as follows:

Theorem 10.2. *Dependencies $D1_D$–$D9_D$ and Independencies $I1_D$–$I4_D$ are sufficient to derive the basic chained trust opinion of A about C with recommendation by B. Given that A's opinion about B is $f_{R_B}(b|\psi) = g(b)$: $f_{R_C}(c|\psi, S_C^B{=}(y_s, y_f)) =$*

$$\mathbf{eq^1}_{y_s,y_f}(c) \cdot \mathbf{eq^2} + \sum_{w \in O_C^B} (\mathbf{eq^1}_{w_s,w_f}(c) \cdot \mathbf{eq^3} \cdot (1 - \mathbf{eq^2})), \tag{10.1}$$

where,

$$\begin{aligned}
\mathbf{eq^1}_{\varphi_s,\varphi_f}(c) &= f_{\mathrm{B}}(c; \varphi_s + 1, \varphi_f + 1),\\
\mathbf{eq^2} &= \frac{\mathbf{eq^4} \cdot \int_0^1 b \cdot g(b)\ \mathrm{d}b}{\mathbf{eq^4} \cdot \int_0^1 b \cdot g(b)\ \mathrm{d}b + \sum_{w' \in O_C^B} \mathbf{eq^5}_{w'} \cdot (1 - \int_0^1 b \cdot g(b)\ \mathrm{d}b)},\\
\mathbf{eq^3} &= \frac{\mathbf{eq^5}_{w_s,w_f}}{\sum_{w' \in O_C^B} \mathbf{eq^5}_{w'_s,w'_f}},\\
\mathbf{eq^4} &= \lambda(y_s + y_f) \cdot \binom{y_s + y_f}{y_s} \cdot \frac{y_s! y_f!}{(y_s + y_f + 1)!}\\
\mathbf{eq^5}_{\varphi_s,\varphi_f} &= \int_0^1 \chi^B(b, \varphi_s, \varphi_f)(y_s, y_f) \cdot \frac{g(b) \cdot (1-b)}{1 - \mathbf{E}(g)}\ \mathrm{d}b\\
&\quad \cdot \lambda(\varphi_s + \varphi_f) \cdot \binom{\varphi_s + \varphi_f}{\varphi_s} \cdot \frac{\varphi_s! \varphi_f!}{(\varphi_s + \varphi_f + 1)!}
\end{aligned}$$

Proof. See Section B.3 of Appendix B. □

10.1.3 Semantics

If we were to only assert Dependencies $\mathrm{D1}_D$-$\mathrm{D9}_D$ and Independencies $\mathrm{I1}_D$-$\mathrm{I4}_D$, then we would have a family of Beta models with logical trust operations and trust chaining parameterised in the lying strategies χ and the entanglement λ. However, in this chapter, we are looking for a specific model. Hence we need to select the parameters. Recall that we want to have a model in which lies represent no information, in which every recommendation is equally likely successful, and which is independent of the entanglement.

The latter requirement can be fulfilled by fulfilling Lemma 8.7. Lemma 8.7 states that a lying strategy exists, with the property that the entanglement does not affect the computation of a trust chain. The lemma is proven by providing a class of lying strategies with that property. All members χ of the class have $\chi(b, w_s, w_f)(y_s, y_f) = 0$ for $w_s{+}w_f \neq y_s{+}y_f$. In other words, fake recommendations must have the same size as the recommenders' interaction history. To ensure that our model is independent of the entanglement, it suffices to select any lying strategy with $\chi(b, w_s, w_f)(y_s, y_f) = 0$ for $w_s + w_f \neq y_s + y_f$.

The restriction on the lying strategies does not interfere with our former requirements – that lies represent no information or that recommendations are successes with fixed probability. We have not yet formally defined what we mean by 'no information', however, a similar discussion is found in Game 9.4 in Section 9.2. The subject is not interested in the interaction history between the recommender and the target in general, but about its implications regarding the *integrity* of the target. Under the restriction that fake recommendations have the same size as true interaction histories, the subject can trivially derive the true size of the interaction history. This, however, does not entail that the subject can derive anything about the integrity of the target. To ensure that no information is leaked in a lie (see e.g. Section 9.3), it suffices to ensure that there is no correlation between lies and the truth. Formally, that $\chi(b, w_s, w_f) = \chi(b, w'_s, w'_f)$ whenever $w_s + w_f = w'_s + w'_f$.

There are still several possible lying strategies that adhere to the two aforementioned requirements. The last requirement is that each recommendation is equally likely successful. That means that the probability that each recommendation by B is true with a probability equal to his integrity parameter b, since if it is not, $P(E_B{=}\mathrm{S}|R_B{=}b) \neq b$. Formally stated, $P(E_B{=}\mathrm{S}|R_B{=}b, S_C^B = y) = P(E_B{=}\mathrm{S}|R_B{=}b)$. This holds when $P(S_C^B = y|R_B{=}b, E_B{=}\mathrm{S}) = P(S_C^B = y|R_B{=}b, E_B{=}\mathrm{F})$, which in turn holds if $P(S_C^B{=}(y_s, y_f)) = P(O_C^B{=}(w_s, w_f))$. Since that relation trivially holds when $E_B{=}\mathrm{S}$, we merely need to select χ correctly for $E_B{=}\mathrm{F}$. Earlier requirements on χ dictate that $y_s + y_f = w_s + w_f$, and that $\chi(b, w_s, w_f) = \chi(b, w'_s, w'_f)$, meaning that $\chi(b, w_s, w_f) \sim \frac{O_C^B}{\lambda(y_s, y_f)}$. Running ahead on the results of Proposition 10.3, we see that this means $\chi(b, w_s, w_f)(y_s, y_f) = \frac{1}{w_s + w_f + 1}$, when $w_s + w_f = y_s + y_f$ (and 0 otherwise).

Proposition 10.3. *The probability of finding an interaction history y_s, y_f is $P(O_C^B = (y_s, y_f)) = \frac{\lambda(y_s + y_f)}{y_s + y_f + 1}$.*

Proof. Apply the law of total probability, to get $\int_0^1 P(O_C^B = (n, m)|R_C = c) \cdot f_{R_C}(c)\ \mathrm{d}c$. Apply Dependencies $\mathrm{D1}_D$ and $\mathrm{D3}_D$, to get $\binom{y_s + y_f}{y_s} \cdot \lambda(y_s + y_f) \cdot \int_0^1 c^{y_s}(1 -$

$c)^{y_f}\ \mathrm{d}c$, which simplifies to $\lambda(y_s+y_f)\cdot\frac{(y_s+y_f)!}{y_s!\cdot y_f!}\cdot\frac{y_s!\cdot y_f!}{(y_s+y_f+1)!}=\frac{\lambda(y_s+y_f)}{y_s+y_f+1}$. □

Now, thanks to Theorem 8.9, we know that we can study the trust chain in isolation. Thus, $f_{R_C}(c|O_B^A=x,S_C^B=y,\varphi)\propto f_{R_C}(c|O_B^A=x,S_C^B=y)\cdot f_{R_C}(c|\varphi)$, meaning that we can assume without loss of generality, that the subjects opinion on the recommender and the recommendation are the only relevant terms. That means that we do not need to redo our above analysis when adding, say, a condition $O_C^A=(z_s,z_f)$.

Now, we have sufficient formal machinery to define the semantics of the expressions in the Default Model.

Definition 10.2 (Default Model)**.** The Default Model is defined by Dependencies $\mathrm{D1}_D$–$\mathrm{D9}_D$ and Independencies $\mathrm{I1}_D$–$\mathrm{I4}_D$, as well as arbitrary λ, and $\chi(b,w_s,w_f)=\frac{1}{w_s+w_f+1}$ for arbitrary $b\in[0,1]$ and $(w_s,w_f)\in\mathbb{N}\times\mathbb{N}$.

Interestingly, the Default Model is not just an arbitrary model that adheres to our three requirements, it is the only model to do so.

Theorem 10.4. *The Default Model is the only model that is independent of the entanglement, where lies leak no information and where all possible recommendations by a recommender are equally likely to be true.*

Proof. Assume that there exists a (w_s,w_f) and (y_s,y_f), such that $\chi(b,w_s,w_f)(y_s,y_f)>0$, for $w_s+w_f\neq y_s+y_f$. Now, since the theorem must hold for all entanglements λ, it must hold for the point distribution $\lambda(w_s+w_f)=1$. By assumption, there exists $y_s+y_f\neq w_s+w_f$, such that $\chi(b,w_s,w_f)(y_s,y_f)>0$, meaning that $P(S_C^B{=}(y_s,y_f)|E_B{=}\mathrm{F})>0$, but $P(S_C^B{=}(y_s,y_f)|E_B{=}\mathrm{S})=0$. Therefore, not all possible recommendations by a recommender are equally likely to be true.

Unless λ is the point distribution $\lambda(0)=1$ (i.e. it is impossible that recommenders interacted with the target), there exist at least two different possible interaction histories, since, with non-zero probability, $\lambda(k)>0$ for $k>0$, and there are at least $k+1$ recommendations of size k. Assume now, that $\chi(b,w_s,w_f)\neq\chi(b,w_s',w_f')$, for $w_s+w_f=w_s'+w_f'$. Then, by assumption, there exists (y_s,w_f), such that $\chi(b,w_s,w_f)(y_s,y_f)\neq\chi(b,w_s',w_f')(y_s,y_f)$, therefore, when the recommendation (y_s,y_f) is given, this must leak information regarding the true opinion.

Finally, assume that $\chi(b,w_s,w_f)(y_s,y_f)\neq\frac{1}{w_s+w_f+1}$ (for all (w_s,w_f) with $w_s+w_f=y_s+y_f$). Then $P(S_C^B{=}(y_s,y_f)|E_B{=}\mathrm{S})=P(O_C^B{=}(y_s,y_f)|\ {=})\frac{\lambda(y_s+y_f)}{y_s+y_f+1}$ (via Proposition 10.3). However, $P(S_C^B{=}(y_s,y_f)|E_B{=}\mathrm{F})=\int_0^1\sum_{0\leq k\leq y_s+y_f}\chi(b,k,y_s+y_f-k)(y_s,y_f)\cdot f_{R_B}(b)\cdot P(O_C^B{=}(k,y_s+y_f-k))\ \mathrm{d}b=\sum_{0\leq k\leq y_s+y_f}\chi(b,k,y_s+y_f-k)(y_s,y_f)\cdot\frac{\lambda(y_s+y_f)}{y_s+y_f+1}$, since all $\chi(b,w_s,w_f)$, we can reformulate χ, and k no longer appears: $\chi(b,k,y_s+y_f-k)(y_s,y_f)\cdot\lambda(y_s+y_f)\neq\frac{\lambda(y_s+y_f)}{y_s+y_f+1}$ by assumption. Hence, not all possible recommendations by a recommender are equally likely to be true.

The only remaining choice for χ is $\chi(b,w_s,w_f)(y_s,y_f)=\frac{1}{w_s+w_f+1}$ for all (w_s,w_f) with $w_s+w_f=y_s+y_f$, which is the choice of the Default Model. □

10.1.4 Analysis

The purpose of this section is twofold, the first is to provide insight into the Default Model, and by proxy, the other models, and the second is to illustrate the techniques from Chapter 9 In this section, we make three short analyses. First, we define an example of a trust chain where, intuitively, *endogenous filtering* seems required. This example provides an intuition behind Theorem 8.9. Second, we provide an alternative definition of the Default Model, using game theory. Finally, we analyse the information leakage of recommendations in the Default Model.

Example Illustrating Modularity In Section 8.3, we discuss the modularity of trust chaining. Modularity of trust chaining precludes the correctness of endogenous filtering. The intuition behind endogenous filtering, is that if a recommendation is unlikely given your own opinion, it should be weighted less in aggregation. If the subject's opinion about a target is diametrically opposed to a received recommendation regarding that target, the subject suspects the recommendation less likely to be true. Yet, the subject can compute the trust chain without his own opinion (i.e. modularly), and then aggregate the result with his own opinion without correction. We provide an example that should provide an intuitive insight into this result.

Example 10.1. The user A is fairly positive about user C, seeing that A observed 5 successes and only 1 failure, thus the simple trust opinion of A about C is $\vartheta_{5,1}$. User B claims to have observed only 0 success and 6 failures. Although it is not impossible for B to actually have observed this, (while A has opinion $\vartheta_{5,1}$), it is unlikely. The opinion of A about B has an expected value of $1/3$ (the actual opinion is irrelevant, perhaps it equals $\vartheta_{2,7}$). User A constructs a trust chain in the Default Model. The resulting chained trust opinion is $f(x) = 1/3 \cdot \vartheta_{0,6}(x) + 2/3$ (Theorem 10.2). The trust aggregation of these two opinions reflects the trust opinion of A about C using all available information. Due to Lemma 6.6, the aggregation $g(x)$ is proportional to $f(x) \cdot \vartheta_{5,1}(x)$, thus $g(x) = f(x) \cdot \vartheta_{5,1}(x) \cdot \mathrm{NF}_1{}^3 = (1/3 \cdot \vartheta_{0,6}(x) \cdot \vartheta_{5,1}(x) + 2/3 \cdot \vartheta_{5,1}(x)) \cdot \mathrm{NF}_1$. Now, for x where $\vartheta_{0,6}(x)$ is large, $\vartheta_{5,1}(x)$ is near zero, and vice versa (see Figure 10.1). This effect is a direct consequence of the fact that $\vartheta_{0,6}$ states something radically different from $\vartheta_{5,1}$. Since either $\vartheta_{0,6}(x)$ or $\vartheta_{5,1}(x)$ is very small, for all values of x, $h(x) = \vartheta_{0,6}(x) \cdot \vartheta_{5,1}$ is small for all x (see Figure 10.1). Therefore $2/3 \cdot \vartheta_{5,1}(x) \gg 1/3 \cdot \vartheta_{0,6} \cdot \vartheta_{5,1}(x)$ and $g(x) = (1/3 \cdot \vartheta_{0,6}(x) \cdot \vartheta_{5,1}(x) + 2/3 \cdot \vartheta_{5,1}(x)) \cdot \mathrm{NF}_1 \approx 2/3 \cdot \vartheta_{5,1}(x) \cdot \mathrm{NF}_2 = g(x)$.

The example is depicted graphically in Figure 10.1.

In the example $g \approx \vartheta_{5,1}$, meaning that the trust opinion of A about C based on A's interaction history and B's recommendation is similar to the trust opinion of A about C based only on A's interaction history. The reason for this approximate equality, is the fact that $\vartheta_{5,1}(x) \cdot \vartheta_{0,6}(x) \approx 0$. This, in turn, occurred due to the high level of disagreement between $\vartheta_{5,1}$ and $\vartheta_{0,6}$. Our intuition that the recommendation should have less impact when the content of the recommendation ($\vartheta_{0,6}$) conflicts with the subjects knowledge ($\vartheta_{5,1}$) is naturally satisfied, without requiring endogenous filtering.

[3]The normalisation factor is a unique scalar that turns a function into a distribution. Hence, we need not specify the normalisation factor explicity.

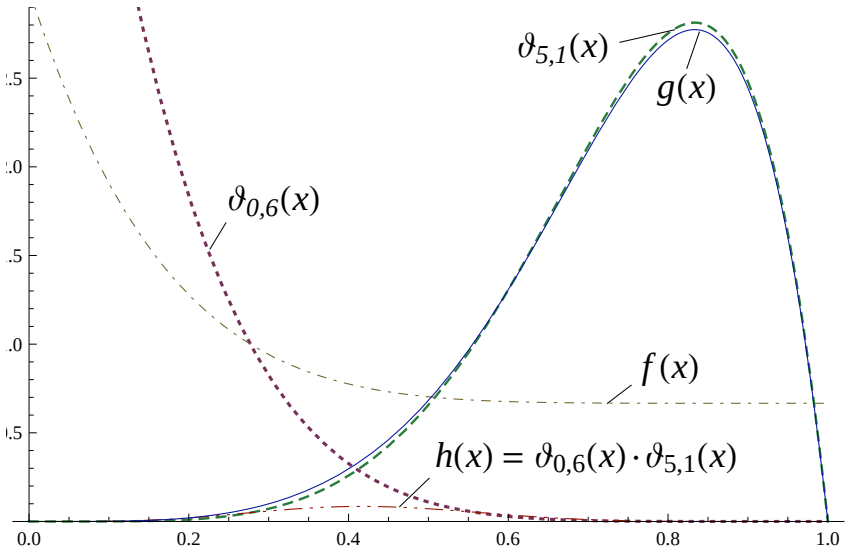

Figure 10.1: The graphs of the functions mentioned in Example 10.1.

Game Theoretic Definition of the Default Model In this section, we have informally specified the Default Model, and used the specification to formally define the Default Model by setting χ. In Section 9.2, we asserted that game theory may be used in defining trust models. To illustrate this assertion, we define a game for which (under Dependencies $\mathrm{D1}_D$-$\mathrm{D9}_D$ and Independencies $\mathrm{I1}_D$-$\mathrm{I4}_D$) the Default Model defines the optimal strategy for the recommender.

Game 10.1. For a recommender B, with integrity $R_B = b$, there is a probability b that B wants to leak no information about the target. Perhaps because B is colluding with a fraction b of the targets. Therefore, B does not want the subject, A, to gain information about which targets B is lying about, since that would allow A to gain information of whom B is colluding with.

We can define the information about the target in different ways, most definitions would work. Here, we simply take the information entropy of R_C, where C is the target. Hence, we might use $h(R_C|S_C^B)$ as our measure (see Definition 9.2), which we want to maximise. However, that measure does not include the information about whether a statement is true. We can define the information gain about whether B is lying provided his statement as $H(E_B|S_C^B)$. We need to combine the two terms, $h(R_C|S_C^B)$ and $H(E_B|S_C^B)$ into a single measure. Whatever the combined measure f is, we require it to be strictly monotonically increasing in both $h(R_C|S_C^B)$ and $H(E_B|S_C^B)$. The most obvious choice for f is a sum (or a weighted sum), but many other choices are possible, such as $f(x, y) > f(x', y')$ iff $x > x'$ or both $x = x'$ and $y > y'$. Different choices for f define the value of one type of information, relative to the other type of information. However, since we already know there is an optimal strategy that leaks no information about either, we know that the exact choice of f is irrelevant.

Under Dependencies $\mathrm{D1}_D$-$\mathrm{D9}_D$ and Independencies $\mathrm{I1}_D$-$\mathrm{I4}_D$, the Default Model is defined as the model in which all recommenders are rational users that do not know the entanglement λ, who try to maximise some utility function f, that is strictly monotonically increasing in both $h(R_C|S_C^B)$ and $H(E_B|S_C^B)$. The proof of

this statement is a minor modification of Theorem 10.4, in which we merely need to add the notion that the minimal information leakage is 0, i.e. that $h(R_C|S_C^B)$ and $H(E_B|S_C^B)$ are maximal when equal to $h(R_C)$ and $H(E_B)$, respectively.

Analysis of Information Leakage In the Default Model, information leakage is minimised in the cases where the recommender is lying. However, seeing that the recommender may not be lying, we should not expect the total information leakage to be minimised. First, we compute the information leakage of a recommendation, then we show, by example, that lying strategies with a lower total information leakage exist.

The information leakage of a recommendation is defined $h(R_C|S_C^B) - h(R_C)$, via Definition 9.3 Applying Definitions 9.1 and 9.2, that equates to $\sum_{y_s,y_f\in\mathbb{N}} P(S_C^B = (y_s,y_f)) \cdot \int_0^1 f_{R_C}(c|S_C^B{=}(y_s,y_f)) \cdot \log(\frac{1}{f_{R_C}(c|S_C^B=(y_s,y_f))})\ \mathrm{d}c - \int_0^1 f_{R_C}(c) \cdot \log(\frac{1}{f_{R_C}(c)})\ \mathrm{d}c$. We can simplify the term, and distinguish true and false recommendations: $\sum_{y_s,y_f\in\mathbb{N}} \frac{\lambda(y_s+y_f)}{y_s+y_f+1} \cdot \int_0^1 (f_{R_C}(c, E_B{=}\mathrm{S}|S_C^B{=}(y_s,y_f)) + f_{R_C}(c, E_B{=}\mathrm{F}|S_C^B{=}(y_s,y_f))) \cdot \log(\frac{1}{f_{R_C}(c,E_B=\mathrm{S}|S_C^B=(y_s,y_f)) + f_{R_C}(c,E_B=\mathrm{F}|S_C^B=(y_s,y_f))})\,\mathrm{d}c - 0$. And since every statement is equally likely true: $f_{R_C}(c, E_B{=}u|S_C^B{=}(y_s,y_f)) = f_{R_C}(c|S_C^B{=}(y_s,y_f), E_B{=}u) \cdot P(E_B{=}u|S_C^B{=}(y_s,y_f)) = f_{R_C}(c|S_C^B{=}(y_s,y_f), E_B{=}u) \cdot P(E_B{=}u)$. The term $f_{R_C}(c|S_C^B{=}(y_s,y_f), E_B{=}\mathrm{S}) = f_{R_C}(c|O_C^B{=}(y_s,y_f)) = \vartheta_{y_s,y_f}(c)$. In case of failure $f_{R_C}(c|S_C^B{=}(y_s,y_f), E_B{=}\mathrm{F}) = 1$, the uniform distribution. Letting $p = P(E_B{=}u)$ and $f(c; y_s, y_f) = f_{R_C}(c|O_C^B{=}(y_s,y_f))$, the information leakage is $\sum_{y_s,y_f\in\mathbb{N}} \frac{\lambda(y_s+y_f)}{y_s+y_f+1} \cdot \int_0^1 (p \cdot f(c; y_s, y_f) + (1-p)) \cdot \log(\frac{1}{p\cdot f(c;y_s,y_f)+(1-p)})\,\mathrm{d}c$.

It is interesting to note that it immediately follows that the Default Model is not optimal in the zero-sum recommendation game. Take any lying strategy which is identical to the lying strategy of the Default Model, with only the probability of $\chi(b, w_s, w_f)(w_s, w_f)$ lowered with δ (and therefore $\chi(b, w_s, w_f)(y_s, y_f)$ raised with $\frac{\delta}{y_s+y_f}$ for $(w_s, w_f) \neq (y_s, y_f)$), for any $0 < \delta \leq b$. Without formally analysing the exact difference, observe that we simply offset a fraction of the times we are forced to tell the truth, by reducing the probability of stating the truth when we may lie. Furthermore, note that picking the largest such δ provides an optimal strategy when $b \leq \frac{1}{w_s+w_f+1}$, coinciding with the results of Game 9.4.

10.2 Representation of the Default Model

Each *trust opinion* in the Default Model is represented by a probability density function ranging over $[0, 1]$. However, a trust opinion may have different representations, and as a result two identical elements may not be recognised as being identical. Theorem 7.4 and Corollary 7.6 result in a complicated computation, and Theorems 8.1 and 10.2 are even more complicated. One can imagine that two different nestings of operations may result in the same probability density function, but in different formulations thereof. It is undesirable to have identical trust opinions without a quick way of realising that they are identical.

Furthermore, some formulations of trust opinions may be more insightful than others. For example, $\int_0^x 12y^2 \,\mathrm{d}y$ happens to be equal to $4x^3$. Under most circumstances, the latter formulation is far more informative, as it tells us immediately that we are dealing with a trust opinion with 3 successes and 0 failures, i.e. $\mathrm{NF} \cdot x^3 \cdot (1-x)^0$.

The first representation we discuss deals with representing the trust opinion (the probability density function) in a standardised and clear manner, that allows us to recognise important properties and a general shape quickly. The representation is based on the notion that every trust opinion can be represented as the summation of a particular simpler class of formulae. We refer to this representation as the summation representation. We also present a second representation with a wholly different footing. For the second representation, we trade precision for efficiency and clarity, by intermediately approximating all integrations using the midpoint approximation. In other words, every opinion is stored as a collection of n points. Hence, this representation is called the midpoint representation.

10.2.1 The Summation Representation

In this section, we aim to represent trust opinions as a computation which is equivalent to the original trust opinion. The computation should consist of a summation of subexpressions of similar shapes. Summations have the desirable property that their properties (such as expected value or variance) are easy to compute, given the properties of its constituents. Complex summations can be analysed or approximated using algebraic techniques, in particular using the notion that summands with near-zero values hardly have any impact.

The exact shape of the summands is defined later. Informally, each summand represents (nested) conjunctions and disjunctions of simple trust opinions. That means that summands do not contain aggregation, chaining or negation. We show how to eliminate these operations constructively.

To simplify notation, we introduce two functions $\prod$ and $\coprod$, which we refer to as product and coproduct, since the former mimics the product random variable $R_{S \wedge T}$ and the latter mimics its converse $R_{S \vee T}$.

Definition 10.3 (Product and coproduct)**.** Let $\prod$ and $\coprod$ be typed $([0,1] \to \mathbb{R}) \times ([0,1] \to \mathbb{R}) \to [0,1] \to \mathbb{R}$. Then, for distributions $f, g : [0,1] \to \mathbb{R}$, we define:

$$\prod(f,g)(x) = \int_x^1 \frac{1}{y} \cdot f(\frac{x}{y}) \cdot g(y) \;\mathrm{d}y,$$

and

$$\coprod(f,g)(x) = \int_{1-x}^1 \frac{1}{y} \cdot f(1 - \frac{1-x}{y}) \cdot g(1-y) \;\mathrm{d}y.$$

Note that the product and coproduct are chosen such that if $g(x) = f_{R_S}(x|\varphi)$ and $h(x) = f_{R_T}(x|\varphi)$, then $\prod(g,h)(x) = f_{R_{S \wedge T}}(x|\varphi)$ (see Theorem 7.4) and $\coprod(g,h)(x) = f_{R_{S \vee T}}(x|\varphi)$ (see Corollary 7.6). The inverse of $f(x)$ is $f(1-x)$, which we denote $\hookleftarrow(f)(x) = f(1-x)$.[4] From Proposition 7.3, De Morgan, it

[4]The symbol $\hookleftarrow$ was selected to reflect a graphical intuition about negation. The graph of $\hookleftarrow(f)$ is the mirror image over the axis $x = 0.5$; as if reading the graph right-to-left.

follows that $\hookleftarrow(\prod(f,g)) = \coprod(\hookleftarrow(f), \hookleftarrow(g))$ and vice versa (since $\overline{x \wedge y} = \overline{x} \vee \overline{y}$).

We can use this notation to define the format of the summation representation. That any trust opinion fits the format will be proven afterwards.

Definition 10.4 (Summation representation)**.** Let B be the set of *simple trust opinions* $\mathcal{B}$. Let P be the smallest superset of B closed over $\prod$ and $\coprod$. We refer to the elements of P as summands. A trust opinion $f : [0,1] \to \mathbb{R}$ is in the summation representation, when it is shaped $f(x) = \sum_i a_i \cdot p_i(x)$, for $a_i \in \mathbb{Q}^+$ and $p_i \in P$. We may refer to a_i as the weight of the summand p_i. If all summands of f are elements of B, then we say f is in simple summation representation, i.e. $f(x) = \sum_i a_b \cdot b_i(x)$, for $a_i \in \mathbb{Q}^+$ and $b_i \in B$.

The set B contains exactly all simple trust opinions, and the set P contains exactly all (nested) applications of product and coproducts of the simple trust opinions. Every summand – element of P – represents a *composite trust opinion*.

Before we prove that all trust opinions fit in the summation representation, we prove that the expected value of a trust opinion in the summation representation can be computed efficiently.

Proposition 10.5. *Given a trust opinion f in the summation representation (with n summands), we can compute the expected value $\mathbf{E}(f)$ using only addition, subtraction, multiplication and division, in $O(n)$ time.*

Proof. The expected value of a simple trust opinion $\vartheta_{s,f}$ is $\frac{s+1}{s+f+2}$, via Proposition 6.4. The expected value of $\prod(f,g)$ is $\mathbf{E}(f) \cdot \mathbf{E}(g)$, and $\mathbf{E}(\coprod(f,g)) = \mathbf{E}(f) + \mathbf{E}(g) - \mathbf{E}(f) \cdot \mathbf{E}(g)$, via Corollary 7.8. Expected values are linear, meaning that $\mathbf{E}(a \cdot f) = a \cdot \mathbf{E}(f)$, for scalar a, and $\mathbf{E}(f + g) = \mathbf{E}(f) + \mathbf{E}(g)$. Hence, we can use the first notion to obtain the expected value of all simple trust opinions, the second notion to obtain the expected value of the summands, and the latter notions to obtain the expected value of f. □

In special cases the variance of a trust opinion in the summation representation can be computed as efficiently as the expected value. These special cases arise when the covariance between every pair of summands is zero. The trust aggregation of trust chains, however, typically yields summands that have a non-zero covariance. The variance of a trust opinion can also be computed incrementally, like the expected value, but for each pair of summands that correlate, the covariance needs to be computed. Therefore, computing the variance has a worst-case performance of $O(n^2)$ steps, where n is the number of summands.

To prove that every term in the Default Model can be represented in the summation representation, we provide a theorem with a constructive proof. This implies that our proof not only proves that there exists a computation in the summation representation for each expression in the Default model, but moreover provides a method of constructing the computation.

Theorem 10.6. *For any expression φ in the Default Model, the trust opinion f that is the semantics of φ can be formulated in the summation representation. Moreover, if the main operator of φ is not a logical trust operation, f can be formulated in the simple summation representation.*

Proof. We can prove this by structural induction over φ. If φ is a simple trust opinion ϑ, then $f(x) = 1 \cdot \vartheta(x)$. The induction hypothesis is straightforward, namely that the theorem holds for expressions φ and ψ:

$\varphi + \psi$ Then $f(x) = g(x) \cdot h(x) \cdot \alpha$, with $\alpha = \frac{1}{\int_0^1 g(y) \cdot h(y) \ \mathrm{d}y}$. By induction hypothesis, g and h are in simple summation representation. That means that $f(x) = \alpha \cdot \sum_{0 \leq i \leq n} a_i b_i(x) \cdot \sum_{0 \leq j \leq m} a'_j b'_j(x)$, by distributivity $f(x) = \sum_{0 \leq i \leq n} \sum_{0 \leq j \leq m} \alpha a_i a_j b_i(x) b'_j(x)$. By Lemma 6.6, there exists a simple trust opinion $c_{i,j}(x) \propto b_i(x) b'_j(x)$. Hence f fits in the simple summation representation, if both α and all $\frac{c_{i,j}(x)}{b_i(x) b'_j(x)}$ are rational. The latter holds, since via Definition 6.7, they equal $\frac{\mathrm{B}(s+1,f+1) \cdot \mathrm{B}(s'+1,f'+1)}{\mathrm{B}(s+s'+1,f+f'+1)}$, for natural s, f, s', f' such that $b_i(x) \propto x^s \cdot (1-x)^f$ and $b_j(x) \propto x^{s'} \cdot (1-x)^{f'}$, and $\mathrm{B}(n, m)$ is integer for integer n and m. That α is rational follows by pigeon hole principle, since α times a sum of products of rational numbers must equal one.

$\varphi + \psi$ Then $f(x) = \mathbf{E}(g) \cdot h(x) + (1 - \mathbf{E}(g))$, where $h(x)$ is a simple trust opinion. Due to Proposition 10.5, we can compute $a = \mathbf{E}(g)$ and $a' = 1 - a$. Therefore, $f(x) = a \cdot h(x) + a' \cdot 1$. Since both $h(x)$ and 1 are simple trust opinions, f fits in the simple summation representation.

$\overline{\varphi}$ Then $f(x) = g(1-x)$, where, by induction hypothesis, g is in the summation representation, i.e. $g(x) = \sum_i a_i p_i(x)$. Informally, we push the negation to smaller subexpressions, until the simple trust opinions whose negation is another simple trust opinion. Observe, $g(1-x) = \sum_i a_i p_i(1-x)$, where $p_i(1-x) = \hookleftarrow(p_i)(x)$. We apply structural induction to show that $\hookleftarrow(p_i)$ can be expressed as a summand. If p_i is a simple trust opinion, then $\hookleftarrow(p_i)$ is also a simple trust opinion, since $(\lambda x.x^S(1-x)^F)(1-x) = (1-x)^S x^F$. If $p_i = \prod(h, h')$, then $\hookleftarrow(p_i) = \coprod(\hookleftarrow(h), \hookleftarrow(h'))$. If $p_i = \coprod(h, h')$, then $\hookleftarrow(p_i) = \prod(\hookleftarrow(h), \hookleftarrow(h'))$.

$\varphi \wedge \psi$ Then $f(x) = \int_x^1 \frac{1}{y} \cdot g(\frac{x}{y}) \cdot h(y) \ \mathrm{d}y$, for g and h in the summation representation. Therefore, $f(x) = \int_x^1 \frac{1}{y} \cdot (\sum_i a_i p_i(x/y)) \cdot (\sum_j a'_j p'_j(y)) \ \mathrm{d}y$, and by distributivity $f(x) = \sum_i \sum_j a_i \cdot a_j \cdot \int_x^1 \frac{1}{y} \cdot p_i(x/y) \cdot p'_i(y) \ \mathrm{d}y$. By definition of the product, $f(x) = \sum_i \sum_j a_i \cdot a_j \cdot \prod(p_i, p_j)$, thus f is in the summation representation.

$\varphi \vee \psi$ For identical reasons as the previous case, $f(x) = \sum_i \sum_j a_i \cdot a_j \cdot \coprod(p_i, p_j)$.

□

The summation representation has the advantage that a trust opinion can be understood reasonably well. Since each summand is a probability distribution, each summand has equal surface area. That implies that the impact of a summand is determined by its weight; the associated scalar. When approximating or estimating properties of the graph, knowledge of which summands are relevant can increase the quality of the approximation or estimate. Furthermore, any linear operation can be applied to large formula in the summation representation relatively easily. Another advantage is that trust opinions in this representation can be transformed into trust opinions in the midpoint representation, which, in turn, can be used to plot these trust opinions.

10.2.2 The Midpoint Representation

Where the summation representation of a trust opinion is equal to the trust opinion itself, the midpoint representation is merely an approximation. The midpoint representation is the foundation of the *Canephora* tool (Section 8.2.1). The main advantages of the midpoint representation are practical, rather than theoretical.

The foundation of the midpoint representation is a standard technique in numerical analysis. A way to approximate a definite integral, is via the midpoint rule (also called rectangle method). The midpoint rule states that, in the interval $[0, 1]$, $\int_0^1 f(x)\ \mathrm{d}x \approx \frac{1}{N}\sum_{0\leq i\leq N-1} f(\frac{2i+1}{2N})$, for sufficiently large N. Take, for example, $N = 100$, then the graph $f(x)$ is cut into 100 rectangles, each with width $1/100$, the first with height $f(0.005)$, the second $f(0.015)$, until finally $f(0.995)$.

The midpoint representation is ideal for graphing purposes, as it is trivially sampled. If you want a graph based on 1000 samples, simply use the midpoint representation with $N = 1000$. Other types of analysis, like expected values or variances, can be computed rapidly, and more importantly in constant time (relative to the complexity of the trust network).

As noted earlier, a trust opinion can be symbolically computed using the summation representation, and then converted to the midpoint representation to be graphed. However, when only useful in that sense, the midpoint representation would not be a proper representation, as we would not perform trust operations over trust opinions in the midpoint representation. The midpoint representation does have a valid reason of being preferred over the summation representation under some circumstances. The number of summands in a trust opinion in the summation representation is unbounded. That means that there is no fixed, finite representation of trust opinions. In the midpoint representation, the size of the representation of a trust opinion is fixed, namely N values.

The *trust aggregation*, *trust chaining* and *trust negation* operators can trivially be defined over trust opinions in the midpoint representation. They are pairwise multiplication, weighted pairwise summation, and reversing the order of the values, respectively. These three operations have the desirable property that they do not introduce additional error. The other two operations – *trust conjunction* and *trust disjunction* – are less straightforward. Note that since trust negation introduces no errors, it suffices to treat only one of the two operators, via De Morgan.

We define trust conjunction of f and g in the midpoint representation, in the most straightforward fashion:

$$h(x) = \sum_{x\cdot N\leq i\leq N} \frac{N}{i-0.5} \cdot f(\frac{N\cdot x}{i-0.5}) \cdot g(\frac{i-0.5}{N}).$$

Note that this computation introduces an additional error, since we may poll, for example $f(0.4902)$, then $f(0.4998)$, then $f(0.5101)$, overrepresenting the value $f(0.4995)$ and not representing the value $f(0.5005)$.

In the Default Model, it is completely feasible to let N exceed $10,000$, since trust aggregation, trust chaining and trust negation are linear in N and trust conjunction (and thus trust disjunction) is quadratic in N. The midpoint representation can, therefore, be used for practical applications with bounded resources.

10.2.3 Illustrative Algorithms

In this section, we introduce a toy trust system to illustrate how a trust model like the Default Model can assist in defining a trust system. As mentioned in Chapter 1 (e.g. Figure 1.2), *trust systems* our not our object of focus. Therefore, we introduce a simple trust system, which is sufficiently expressive to incorporate a collection of algorithms and data structures based on the Default Model and its representations.

A possible practical scenario that would exploit the full capabilities of the trust model, is cloud computing. In cloud computing, the result of a computation may depend on several users, in several combinations, as discussed in Section 7.1. A client uses the cloud to solve NP-hard problems, of which answers are easy to verify. In a sufficiently large cloud, people may cooperate not only in executing each other's tasks, but also in establishing trust in new users. Formulated in our terminology, recommendations may play a large role in clouds, especially if typical tasks are sensitive, since we would not trust a stranger with a sensitive task. We are, therefore, interested in a user operating in his role as a subject in the cloud.

We assert that the trust system comes in the shape of a software package that users can install on their computer. A subject, such as Alice, can enter the outcomes of interactions and the recommendations into the trust system. In many settings, this step can be automated. If Alice wants to know whether she can trust Charlie, she can query the system. The system can provide an expected value, variance, a graph and more data about the integrity of Charlie, based on the trust opinion it has stored. The system can also be queried to trust composite targets, such as the trust that either Charlie or Debbie performs the computations correctly.

The trust system keeps track of all trust opinions about simple targets (i.e. trust opinions in $\mathcal{P}$). To that end, we introduce a data structure based directly on the simple summation representation (Definition 10.4), which can represent all trust opinions about simple targets, is $\sum_i a_i b_i(x)$, for simple trust opinions $b \in \mathcal{B}$ and $a_i \in \mathbb{Q}$. Simple trust opinions $\vartheta_{s,f} \in \mathcal{B}$ can be represented by a pair of natural numbers (s, f). Therefore, each simple summand can be represented as struct $\mathcal{S}$ consisting of three properties: $weight : \mathbb{Q}, success : \mathbb{N}, failure : \mathbb{N}$. Now, we can represent a trust opinion about simple targets as an array of such structs, which yields the type $\mathcal{T} = \mathbb{N} \mapsto \mathcal{S}$. Having defined the type $\mathcal{T}$, of which the inhabitants represent trust opinions about simple targets, we can define how the system stores all the trust opinions.

To store all trust opinions about simple targets (with $ID : \mathbb{N}$), it suffices to have a collection $C : \mathbb{N} \mapsto \mathcal{T}$, which maps a user identification (as a natural number) to a trust opinion denoted in type $\mathcal{T}$. The collection C is global, since it represents the persistent memory of the trust system. That means that C can be updated with new data. Let's go through a possible run of the system:

Example 10.2. Alice is considering providing a task to Charlie. The trust system checks whether it has a trust opinion about Charlie, and if not creates the uniform distribution as opinion. The trust system can quickly compute the expected value (Algorithm 1) and variance (quadratic time) of the opinion. The trust system can also sample 1000 points (Algorithm 2), and draw a graph through these points quickly. Alice receives these data points from the system. She can also perform

global : The collection $C : \mathbb{N} \mapsto \mathcal{T}$ of trust opinions.
input : A user $ID : \mathbb{N}$.
output: The expected value of the trust opinion represented by $C[ID]$.
function `GetEVSimple(`ID`)`
  $EV := 0$
  foreach S *in* $C[ID]$ **do**
    $EV := EV + S.weight \cdot (S.success + 1)/(S.success + S.failure + 2)$
  end
  return EV

Algorithm 1: Retrieve expected value of a simple trust opinion.

arbitrary queries on the trust opinion on Charlie, such as request the probability that the integrity of Charlie is at least 0.6.

If Alice is not convinced that she trusts Charlie, she may ask Bob for a recommendation. Upon receiving Bob's recommendation, the trust system computes the trust opinion f based solely on the recommendation (Algorithm 3). Then the trust system can update Alice's trust opinion about Charlie with f (Algorithm 4). Similarly, if Alice is convinced that she trusts Charlie, she interacts with Charlie. After the interaction, she knows whether it is a success or not. In either case, the trust system can update Alice's trust opinion about Charlie with the result (Algorithm 4).

In the example, we have used four algorithms. Algorithm 1 is a direct implementation of Proposition 6.4, combined with the notion that expected values are linear. Algorithm 2 exploits the definition of the beta distribution (Definition 6.7) and the definition of the simple summation representation (Definition 10.4). Algorithm 3 is an implementation of the second inductive step described in the proof of Theorem 10.6, thus exploiting the properties of the summation representation. Similarly, Algorithm 4 implements the first inductive step described in the proof of Theorem 10.6. In short, the first four algorithms are direct implementations of techniques we have already specified.

Note that Algorithm 3 actually requires an object typed as a propositional tree. A propositional tree is of recursive type $\mathcal{T}$ introduced to effectively describe composite targets. A simple target is a leaf. Leafs are structs containing $kind : \mathcal{E}, ID : \mathbb{N}$, where $\mathcal{E} = \{leaf, negation, and, or\}$. For a leaf l, $l.kind = leaf$. Composite targets are nodes. Nodes are structs containing $kind : \mathcal{E}, left : \mathcal{T}, right : \mathcal{T}$. If the target is a negation, then its node n has $n.kind = negation$ and $n.left = l$, where l is the node representing its subtarget ($n.right$ need not be defined). If the target is a conjunction (disjunction), then its node n has $n.kind = and$ ($n.kind = or$), and $n.left = l$ and $n.right = r$, where l and r are nodes representing its subtargets.

Example 10.3. The system cannot just query trust opinions, and update trust opinions, it can also compute composite trust opinions. Alice has a task t which can be split in two parts, t_1 and t_2, which she wants to run on three users, Bob, Charlie and Debbie, such that the expected value of success exceeds a threshold p. Alice can send the task t to Bob, Charlie and Debbie and efficiently verify the results. Alice will have the answer iff Bob, Charlie or Debbie succeeds; $B \vee C \vee D$. Of course,

global : The collection $C : \mathbb{N} \mapsto \mathcal{T}$ of trust opinions.
input : A user $ID : \mathbb{N}$, an integrity parameter $p : \mathbb{Q}$.
output: The expected value of the trust opinion represented by $C[ID]$.
function `GetValue(`ID`,`p`)`
  $r := 0$
  foreach S *in* $C[ID]$ **do**
    $r := r + S.weight \cdot p^{S.success} \cdot (1-p)^{S.failure} \cdot \frac{(s+f)!}{s! \cdot f!}$
  end
  return r

Algorithm 2: Retrieve probability that a user has a certain integrity.

input : A propositional tree P, a recommendation $(s, f) : (\mathbb{N} \times \mathbb{N})$.
output: The trust opinion represented by the recommendation (s, f) made by the (composite) target represented by T.
function `GetChainedOpinion(`P`,`s`,`f`)`
  R is a new array of size 2
  $R[0].weight :=$ `GetEVComposite(`P`)`
  $R[0].success, R[0].failure := s, f$
  $R[1].weight := 1 - R[0].weight$
  $R[1].success, R[0].failure := 0, 0$
  return r

Algorithm 3: Retrieve the chained trust opinion based on a recommendation.

global : The collection $C : \mathbb{N} \mapsto \mathcal{T}$ of trust opinions.
input : A user $ID : \mathbb{N}$, a trust opinion $L : \mathcal{T}$.
output: The element in $C[ID]$ became the aggregate of the original $C[ID]$ and L.
function `UpdateOpinion(`ID`,` L`)`
  R is a new array of size $size(C[ID]) \cdot size(L)$
  $cf := 0$
  foreach $0 \leq i < size(C[ID])$ **do**
    foreach $0 \leq j < size(L)$ **do**
      $w, s, f := C[ID][i].weight, C[ID][i].success, C[ID][i].failure$
      $w', s', f' := L[j].weight, L[j].success, L[j].failure$
      $R[i \cdot size(L) + j].weight := w \cdot w' \cdot \frac{s! \cdot f! \cdot s'! \cdot f'! \cdot (s+s'+f+f'+1)!}{(s+f)! \cdot (s'+f')! \cdot (s+s')! \cdot (f+f')!}$
      $R[i \cdot size(L) + j].success := s + s'$
      $R[i \cdot size(L) + j].failure := f + f'$
      $cf := cf + w \cdot w'$
    end
  end
  foreach S *in* R **do** $S.weight := S.weight/cf$
  $C[ID] := R$

Algorithm 4: Update a trust opinion about a simple target by aggregating it with an opinion based on new data.

input : A propositional tree $P : \mathcal{P}$ representing a composite target.
output: The expected value of the trust opinion about P.
function `GetEVComposite`(P)
  if $P.kind = leaf$ **then return** `GetEVSimple`$(P.ID)$
  if $P.kind = negation$ **then return** $1 -$ `GetEVComposite`$(P.left)$
  if $P.kind = and$ **then**
    $leftEV, rightEV :=$ `GetEVComposite`$(P.left) +$ `GetEVComposite`$(P.right)$
    return $leftEV \cdot rightEV$
  end
  if $P.kind = or$ **then**
    $leftEV, rightEV :=$ `GetEVComposite`$(P.left) +$ `GetEVComposite`$(P.right)$
    return $leftEV + rightEV - leftEV \cdot rightEV$
  end

Algorithm 5: Retrieve expected value of a composite trust opinion.

$B \vee C \vee D$ gives the greatest expected value of success, but it wastes resources. Alternatively we could send t to Bob, and a copy of t_1 to Charlie, and a copy of t_2 to Debbie. Alice will have the answer iff Bob succeeds or Charlie and Debbie succeed; $B \vee (C \wedge D)$. Perhaps it suffices to send t_1 to Bob, and let t_2 be run on both Charlie and Debbie. Alice will have the answer iff Bob succeeds and Charlie or Debbie succeeds; $B \wedge (C \vee D)$. The system can efficiently ($O(n)$, for n users) compute the expected value of each of the different composite trust opinions (Algorithm 5). This allows Alice to distribute her task with a sufficiently high probability of success, while minimising latency or minimising redundant executions. If Alice wants to query the trust system for additional properties about her trust opinion, using the midpoint rule, her trust opinion can be approximated effectively.

The additional algorithm used in this example, is Algorithm 5. The algorithm exploits Corollary 7.8 and Algorithm 1 to recursively compute the expected value of a (composite) trust opinion.

We have created a toy trust system for cloud computing to showcase a collection algorithms based on our theory. The collection is by no means complete. For example, there is no algorithm constructing a trust opinion in the midpoint representation, even tough such an algorithm would follow immediately from the theory. (For N midpoints, simply call `"GetValue"`N times.) Nor do we specify a the full functionality of a trust system. We have not specified recommendation gathering, assisted (or automated) decision making, etc. The specification of a full trust system based on the Default Model would be an interesting problem, but falls outside of the scope of this thesis as specified in Chapter 1.

10.3 Axioms and the Default Model

Recall that *fusion*, *dilution*, *AND*, *OR* and *inverse* were introduced as a symbolic notation to express the operations trust aggregation, trust chaining, trust conjunction, trust disjunction and trust negation. In this section, we verify the soundness of the symbolic operations with respect to the probabilistic operations, using the natural mapping.

The axiomatisations in Chapter 5 use a different signature than that outlined in Section 10.1.1. First, we need to restrict the signature, and thus the theories. Then, we can look at the soundness and completeness of the axiomatisation.

As mentioned in Section 5.3, we did not expect even the strong axiomatisation to be complete. We did expect both axiomatisations to be sound, and we will prove that they are both sound with respect to the Default Model. For both axiomatisations, we will provide a reason and a motive why they are not complete.

By restricting the signature to

$$\varphi := x|\varphi + \varphi|\psi \cdot x$$

$$\psi := \varphi|\psi \wedge \psi|\psi \vee \psi|\overline{\psi},$$

for $x \in \mathcal{B}$, axioms **C5**, **C11** and **C12** fall outside of the language. Soundness for the other axioms of **ATC** can be proven straightforwardly.

10.3.1 Soundness

As we have discussed the rationale behind the axioms of **ATC** in Section 5.2, we immediately provide the proof of soundness:

Theorem 10.7. *The axiomatisation* **ATC** *is sound with respect to the Default Model. For all terms* $x, y \in \mathcal{D}$*:* $\mathbf{ATC} \vdash x = y \Rightarrow \mathcal{DM} \models x = y$.

Proof. Soundness of each axiom from **ATC**, except **C5**, follows straightforwardly. Note that υ corresponds to $f_\upsilon(c) = 1$, $\mathbf{1}$ to $f_{\mathbf{1}}(c) = 2c$ and $\mathbf{0}$ to $f_{\mathbf{0}}(c) = 2(1-c)$. In the proof, we denote the trust opinion corresponding to a variable x as f_x.

Axioms **C1**, **C2** and **C3** follow from the identity of multiplication ($f_\upsilon(c) = 1$), commutativity and associativity of multiplication. Axiom **C4** states that $\mathbf{E}(f_x) \cdot 1 + (1 - \mathbf{E}(f_x)) \cdot 1 = 1$, which trivially holds. Axiom **C6**, and **C7** follow immediately from Proposition 4.19. Axiom **C8** states that $\leftarrow (\leftarrow (f_x)) = f_x$, which follows from $f_x(1 - (1 - c)) = f_x(c)$. Similarly, **C9** follows from $f_\upsilon(1 - c) = 1 = f_\upsilon(c)$, and **C10** from $\leftarrow (f_{\mathbf{1}}) = (1 - c) = f_{\mathbf{0}}$. Axioms **C11** and **C12** correspond to terms outside the syntax of the Default Model, however, their statements are sound. For **C11**, observe that $\leftarrow (f_x \cdot f_y) = \leftarrow (f_x) \cdot \leftarrow (f_y)$, via $(\lambda d. f_x(d) \cdot f_y(d))(1 - c) = f_x(1 - c) \cdot f_y(1 - c)$. Similarly, for **C12**, $\leftarrow (\mathbf{E}(f_x) \cdot f_y + (1 - \mathbf{E}(f_x))) = \mathbf{E}(f_x) \cdot \leftarrow (f_y) + (1 - \mathbf{E}(f_x))$. And axiom **C13** follows immediately from Proposition 7.3.

Finally, we need to prove that **C5** is sound within the language of the Default Model. We will prove this by showing that if $\mathbf{ATC} \vdash x = y$ applies **C5**, but $x, y \in \mathcal{D}$, then another derivation without **C5** exists, which is sound as per above. Observe that the only derivations in which **C5** can be applied are those of shape

$x \cdot (y \cdot z)$. The only two axioms that can interfere with this shape are axioms **C4** and **C12**. We can exclude the latter trivially. Thus **C4** remains. However, it can only be applied to create an opinion of shape $x \cdot (y \cdot z)$, if the original opinion is shaped $x \cdot \upsilon$. In that case the entire expression equals υ, thus making the application of **C5** superfluous, since $x \cdot (y \cdot \upsilon) = \upsilon = y \cdot (x \cdot \upsilon)$ via **C4**. □

We need to provide a model for the additional operators $\mathbf{E}(\_)$ and $\mathbf{W}(\_)$. For $\mathbf{E}(\_)$ the obvious choice is the expected value $\mathbf{E}(\_)$. Whether a suitable choice for $\mathbf{W}(\_)$ exists is an open question for now. Without such a choice, we cannot prove soundness of axioms containing this operator. Therefore, we weaken **EVW** to $\mathbf{EVW}^-$, by excluding all axioms that contain the $\mathbf{W}(\_)$ operator.

As we have discussed the rationale behind the axioms of **EVW** in Section 5.3, we immediately provide the proof of soundness of the axiomatisation $\mathbf{EVW}^-$:

Theorem 10.8. *The axiomatisation* $\mathbf{EVW}^-$ *is sound with respect to the Default Model. For all terms* $x, y \in \mathcal{D}$*:* $\mathbf{EVW}^- \vdash x = y \Rightarrow \mathcal{DM} \models x = y$.

Proof. Axioms **K1**, **K2** and **K3** correspond trivially to $\int_0^1 c \cdot f \ \mathrm{d}c$, for $f(c) = 1$, $f(c) = c$ and $f(c) = 1 - c$, respectively. Axiom **K4** corresponds to $\mathbf{E}(\mathbf{E}(f_x) \cdot f_y + (1 - \mathbf{E}(f_x)) \cdot 1) = \int_0^1 c \cdot \mathbf{E}(f_x) \cdot f_y(c) + c \cdot (1 - \mathbf{E}(f_x)) \ \mathrm{d}c = \mathbf{E}(f_x) \cdot \int_0^1 c \cdot f_y(c) \ \mathrm{d}c + (1 - \mathbf{E}(f_x)) \cdot \int_0^1 c \ \mathrm{d}c = \mathbf{E}(f_x) \cdot \mathbf{E}(f_y) + \mathbf{E}(1 - f_x) \cdot \mathbf{E}(\upsilon)$, assuming soundness of axiom **K6**. Axiom **K5** follows immediately from Corollary 7.8. Axiom **K6** corresponds to $\mathbf{E}(\hookrightarrow (f_x)) = 1 - \mathbf{E}(f_x)$, which follows from $\int_0^1 c \cdot f_x(1 - c) \ \mathrm{d}c = \int_0^1 f_x(c) - c \cdot f_x(c) \ \mathrm{d}c = 1 - \int_0^1 c \cdot f_x(c) \ \mathrm{d}c$, via integration by substitution. Finally, axiom **K16** holds, since if $\mathbf{E}(f_x) = \mathbf{E}(f_y)$, then $\mathbf{E}(f_x) \cdot f_z + (1 - \mathbf{E}(f_x)) = \mathbf{E}(f_y) \cdot f_z + (1 - \mathbf{E}(f_x))$ □

10.3.2 Incompleteness

The axiomatisation **ATC** is obviously lacking in axioms regarding trust chaining. For example, for every trust chain $x \cdot z$, there exists an equal trust chain $y \cdot z$, such that $x \neq y$. We prove this statement in Proposition 10.10. However, none of the axioms in **ATC** can possibly help us prove any instance of $x \cdot z = y \cdot z$, unless $\mathcal{DM} \models x = y$.

The above argument depends heavily on the following proposition:

Proposition 10.9. *In the Default Model, for each chained trust opinion there exists another identical chained trust opinion, based on the same recommendation but different opinions on the recommender.*

Proof. Recall that if $\mathbf{E}(f_x) = \mathbf{E}(f_y)$, then $\mathbf{E}(f_x) \cdot f_z + (1 - \mathbf{E}(f_x)) = \mathbf{E}(f_y) \cdot f_z + (1 - \mathbf{E}(f_x))$. □

The proof of Proposition 10.10 relies on the assumptions on the lying strategy. We cannot generalise Proposition 10.10 to all lying strategies.

The axiomatisation **EVW** mostly deals with axioms regarding expected value and weight. There is, however, one axiom (**K16**) that deals with trust opinions themselves. That axiom deals exactly with the scenario described above. Recall that **K16** states $\mathbf{E}(x) = \mathbf{E}(y) \Rightarrow x \cdot z = y \cdot z$.

Hence, in order to show that **EVW** is not complete, we need a more involved counterexample. Notice that neither **ATC** nor **EVW** deal with distributivity, with the exception of negation and the de Morgan law. Therefore, it is not surprising that our counterexample is based on nestings of different operators:

Proposition 10.10. *In the Default Model, there exist two equal trust opinions x and y, such that* $\mathbf{EVW} \nvdash x = y$.

Proof. Let $x = x' \cdot \mathbf{1} + x' \cdot \mathbf{0}$ and $y = y' \cdot (\mathbf{1} + \mathbf{0})$, for x', y' such that $\mathbf{E}(x') \cdot \mathbf{E}(x') = \mathbf{E}(y')$. Let f and g be the trust opinions corresponding to x' and y'.

Observe that $(\mathbf{E}(f) \cdot 2c + (1 - \mathbf{E}(f))) \cdot (\mathbf{E}(f) \cdot 2(1-c) + (1 - \mathbf{E}(f))) = \mathbf{E}(f) \cdot \mathbf{E}(f) \cdot 2c \cdot 2(1-c) + \mathbf{E}(f) \cdot (1 - \mathbf{E}(f)) \cdot 2c + \mathbf{E}(f) \cdot (1 - \mathbf{E}(f)) \cdot 2(1-c) + (1 - \mathbf{E}(f)) \cdot (1 - \mathbf{E}(f))$. Now, since $2c + 2(1-c) = 2$, this equals $\mathbf{E}(f) \cdot \mathbf{E}(f) \cdot 2c \cdot 2(1-c) + 1 - (\mathbf{E}(f) \cdot \mathbf{E}(f))$. Now, since $\mathbf{E}(f) \cdot \mathbf{E}(f) = \mathbf{E}(g)$, we can conclude that it equals $\mathbf{E}(g) \cdot 2c \cdot 2(1-c) + 1 - \mathbf{E}(g)$. Now, it suffices to note that $4 \cdot c \cdot (1 - c) = 2c \cdot 2(1 - c)$. Hence we prove that $\mathcal{DM} \models x = y$.

Seeing that there are no distributivity laws between fusion and dilution, $\mathbf{EVW} \nvdash x = y$. □

10.4 Conclusion

The results in this chapter imply that a formal correctness trust model that captures all relevant operations (trust aggregation, trust chaining and the logical trust operations) exists both in theory and in practice.

The results from Section 10.1 imply that a formal correctness trust model for all relevant operations exists in theory. We merge the assumptions of the simpler models, and generalise the results where necessary. We studied the restrictions and the powers of the resulting model.

In Section 10.2, we show that the theoretical model translates to a practical model. We tame the equations from Section 10.1, by forcing a uniform representation upon them. The representation is based on a summation of product distributions of beta distributions. One advantage is that many properties (notably the expected value) behave nicely over summations and product distributions. Another advantage is that is becomes relatively easy to compare two trust opinions in the summation representation. We provide an alternative representation, which trades precision for efficiency and clarity. Even the intermediate trust opinions are immediately understandable, and relations in trust networks can be visualised and understood readily.

Finally, we connect Part I and Part II, in Section 10.3. Effectively, the deep connection between the axiomatic method and probabilistic method shows that our probabilistic understanding of the building blocks matches our big-picture understanding of the operations that we study. We prove, in particular, that our axiomatisation is sound, however it is lacking in detail. This result is not surprising, as it shows that our big-picture understanding of the operators is correct but not complete. An interesting exercise for the future would be to formulate a complete axiomatisation of the Default Model.

Part III

Concluding remarks

11

Conclusion and Future Work

11.1 Conclusion

Recall the informal formulation of our research question: *How can we correctly combine trust opinions of users on a system where users interact sparsely and with an explicit goal?*. In the course of the thesis we have formalised our assumptions about "a system where users interact sparsely and with an explicit goal", in the form of the Beta paradigm. We have also looked at several ways of combining such trust opinions in the Beta paradigm, namely trust aggregation, trust chaining, trust conjunction, trust disjunction and trust negation. We applied two different approaches – the axiomatic and the probabilistic approach – that share some characteristics. Both approaches were formal, top-down and general.

A common pitfall in abstract and general views, is that an analysis or model may be too generic, vague or tautological, or that arguments descend in sophistry. To avoid this pitfall, we base our assumptions on a trust model used in practice, Subjective Logic – although we only look at a fraction of the model. The same foundation is also used by other trust models, such as TRAVOS and CertainLogic, further solidifying the notion that the Beta paradigm is useful in practice. In each section, therefore, we mention the relationship between the results in the thesis and Subjective Logic, and TRAVOS and CertainLogic, if applicable.

We made a general distinction between cognitive models and correctness models, where in the former, trust opinions relate to perceived integrity, and in the latter, trust opinions relate to actual integrity. We looked at correctness trust models, since they are more suitable for a formal approach. There exist several formal approaches, in the thesis, we applied an axiomatic approach and a probabilistic approach.

11.1.1 Axiomatic Approach

In the axiomatic approach, we asserted that trust opinions have a measure of trustworthiness and a measure of uncertainty. We formulated axioms (self-evident truths) about trust opinions formed using the trust operations. An example of such a self-evident truth, is associativity, the notion that the order in which opinions are aggregated is irrelevant. The advantage of the axiomatic approach is that properties of trust can be studied without restricting to a single model, and conversely, that properties of a particular model can be studied outside of the context of the model. The disadvantage is that there may be statements that are true, but are neither self-evident (i.e. axioms) nor follow from self-evident truths (i.e.

theorems). In particular, this may be the case if our notion of trust opinions is too coarse. We applied the axiomatic approach in Part I, in two different ways.

In Chapter 4, we applied the axiomatic approach to derive a complete finite axiomatisation of a fraction of Subjective Logic. We provided increasingly more sophisticated axiomatisations of Subjective Logic. We analysed the issues and benefits of each axiomatisation. Furthermore, we looked at potential variations for debatable axioms. In order to formulate the finite, complete axiomatisation of the fraction of Subjective Logic (**SL**), we generalised the axiomatisation of the arithmetic mean to deal with tuples, rather than numbers in $\mathbf{AV}^{\kappa}$.

We loosened the ties between the axiomatisation and Subjective Logic in Chapter 5. There, we discarded the axioms from the axiomatisation of Subjective Logic that are not self-evident, or that lead to problematic conclusions. The axiomatisation **ATC** admits Subjective Logic as a model, but rejects some aspects of Subjective Logic as self-evident, allowing variations. Finally, we added stronger axioms, directly regarding the degree of trust and the degree of uncertainty, to obtain **EVW**. Subjective Logic is not a model of **EVW**. The axioms from **ATC** and **EVW** can be seen as criteria for trust models.

11.1.2 Probabilistic Approach

The other method is the probabilistic approach, applied in Part II. In the probabilistic approach we encode the assumptions of the Beta paradigm in relations between random variables and we assign a probabilistic semantics to the trust operations.

Based on the principles of the Beta model, we studied the logical trust operations in Chapter 7. We added new random variables for expressing the operations, and introduced appropriate relations over these random variables. These relations are added to the Beta model's principles, to get the principles of the Beta model with logical trust operations. Together, these principles allowed us to provide a computational definition of the logical trust operations.Furthermore, we provided general properties of the operations in the context of the Beta paradigm, such as the fact that associativity, commutativity and De Morgan hold for conjunction and disjunction and that double negation holds for negation. Another property we show is that the result of a trust conjunction (a composite trust opinion) is generally not a beta distribution. The implication is that in any correctness trust model in the Beta paradigm, the representation of a composite must differ from the representation of a simple trust opinion.

In Chapter 8, we provided a collection of conservative extensions of the Beta model, namely the Beta family with trust chaining. The semantics of trust chaining are translated into additional random variables and relationships thereupon, which are added to the principles of the Beta model. There are two parameters that remain in the computation for trust chaining, the entanglement of the system and the lying strategy of the recommender. Given these two parameters, the chained trust opinions can immediately be computed (even in practice, via the Canephora tool). Any model, therefore, must make assumptions about the lying strategy, which we propagate developers of new trust models with chaining to do. We prove that

there are instances of lying strategies where the entanglement cancels out. Another property of models in the Beta family with trust chaining is modularity. Modularity is the notion that trust chains can be computed and aggregated separately from the subject's own opinion or other recommendations. The notion of modularity is the opposite of the notion of endogenous filtering. Trust chaining, like the logical trust operations, also does not generally yield beta distributions. Again, the implication is that correctness trust models in the Beta paradigm cannot use the same representation for simple trust opinions as for chained trust opinions.

We have analysed the lying strategies using game theory and information theory. In Chapter 9, we define a series of increasingly complex games. In the first game, options are limited and recommendations have a simple representation. We provide the Nash equilibrium of that game. The fourth game represents the relation between the subject and the worst-case recommender. In that game, we can only provide a small class of optimal strategies for the recommender. We saw that, in these games, the measure of information matters. We applied information theory to study the notion of information. Although the question which information is relevant to the subject is subjective, some measures exhibit better properties than others.

All the results from Part II can be combined into a single model. We have chosen a model in the Beta family with trust chaining, and combined it with the Beta model with the logical trust operations, in Chapter 10. We provided a possible representation scheme, which by necessity has the set of beta distributions as a strict subset. This model is related to the final axiomatisation, **EVW**, found in Part I.

The formal analyses throughout the thesis allow us to draw general conclusions regarding trust models. We formally verified some aspects of Subjective Logic, TRAVOS and CertainLogic, such as the treatment of trust aggregation, the assumption of modularity of trust chaining and adherence to axioms found in **ATC**. On the other hand, we show that these models have operations closed over beta distributions, that are designed to implement trust chaining, trust conjunction and trust disjunction, whereas we show that such operations cannot be closed over beta distributions. Hence, Subjective Logic, TRAVOS and CertainTrust are not valid correctness models in the Beta paradigm, despite the fact that they derive their validity of trust aggregation from the Beta paradigm.

11.2 Future Work

There remain several research directions to explore. One direction is to broaden the results, by looking at additional trust operations. Another direction is to generalise some of the assumptions of the Beta paradigm. The last direction is to explore trust chaining in more detail.

There are several operations that we have not considered. Additional operations can be taken from the more expressive Subjective Logic. Examples of interesting operations to add are deduction and abduction, which deal with interdependent trust opinions. Another type of additional operation is to include composite trust opinions based on arbitrary truth tables, e.g. an interaction that succeeds, if and

only if two out of three users succeeds. The SLVisualiser can also be extended with some, or all new operations.

The assumptions of the Beta paradigm can be generalised, and in fact, many formal generalisations of the Beta model are proposed in the literature. The assumption that the prior distribution is uniform can be lifted, after the base-rate in Subjective Logic. The assumption that interactions have only two outcomes can be lifted, after multinomial distributions in Subjective Logic, or other models based on Dirichlet distributions. We can generalise the assumption that targets are static, after the original Beta reputation system. Alternatively, we can generalise the assumption that targets are stateless, after the HMM-based model. Each of the generalisations of the assumptions only formally work for trust aggregation (i.e. in the Beta model, not its extensions). It is interesting to see the impact of the generalisations on the other operations. Our probabilistic approach provides a way of doing so formally.

We believe that exploring the details and implications of trust chaining in detail may prove to be the most interesting direction. The implication of Corollaries 8.5 and 8.6 – that setting the lying strategy (and entanglement) is both necessary and sufficient to select a Beta model with trust chaining – is that research on trust chaining should focus on the lying strategy. In practice, this means that research can focus explicitly on the behaviour of recommenders, and researchers no longer need to find a way to encode this in a computation for trust chaining. To give some concrete examples of research that is now more accessible: Worst-case recommenders, recommenders that want to advertise or smear, and recommenders attacking a system.

The information games we study in Chapter 9 study the worst-case recommender. Future research could find a Nash equilibrium for these games. Doing so would imply that instantiating the lying strategy with the optimal strategy results in a definition of trust chaining in which the subject cannot effectively be deceived. The reason for this is obvious, the Nash equilibrium guarantees the subject to have a non-negative information gain, and any deviation from the optimal strategy by the recommender, leads to more information for the subject.

In reality, recommenders have goals. In particular, these goals include advertising certain targets, and smearing other targets. The concept of implicit collusion may apply here, since the goals of the recommender (e.g. to advertise) is not completely opposite of that of the subject (to gain information). The subject and recommender can implicitly agree that the recommender only mildly skews the recommendation, and in exchange the subject accepts the recommendation as not far from the truth. It may be possible to derive the perfect advertisement strategy, and by virtue, the optimal way for the subject to deal with advertisements.

Finally, recommenders can be more sophisticated attackers that cooperate, rather than operate individually. A prominent example would be a setting where recommenders are being recommended (for example in the Web of Trust). In such a setting, attackers recommenders may skew their opinions favourably towards each other, and unfavourably towards people that have bad opinions on them. The fixed computation based on the lying strategy may prove to be helpful in analysing and categorising such cooperative strategies.

Rather than focussing on the lying strategy by itself, we can focus on the integration

of the lying strategy into practical computations. The Canephora tool calculates trust chains by themselves, but could be integrated with the SLVisualiser tool, to support arbitrarily nested formulae. Furthermore, the Canephora tool has not been optimised fully, for certain classes of entanglements and lying strategies, there may be significant speed-ups. Ultimately, Canephora may be able to function as a trust system, actually computing trust opinions in an online system, based on the extended Beta models.

A

Omitted Proofs of Part I

A.1 Omitted Properties and Proofs

Some completeness proofs in Part I relied on some technical Lemmas and Properties. In this appendix, we introduce and prove the Lemmas that we relied upon in Part I.

Proposition A.1. *If* $\mathbf{FDN} + \mathbf{AV}^3 \vdash x = y$*, then* x *is a term in* $\Sigma^{\Box}_{EXP+AVs^3}$ *if, and only if,* y *is a term in* $\Sigma^{\Box}_{EXP+AVs^3}$*.*

Proof. We can do structural induction with a strengthened induction hypothesis. If $\mathbf{FDN}s + \mathbf{AV}s^3 \vdash \kappa x = y$, then x in $\Sigma^{\Box}_{EXP+AVs^3}$ iff y in $\Sigma^{\Box}_{EXP+AVs^3}$, and if $D(x)$ (or $D(y)$) then x in $\Sigma^{\Box}_{EXP+AVs^3}$ (or y in $\Sigma^{\Box}_{EXP+AVs^3}$, respectively). We look at all the axioms in $\mathbf{FDN} + \mathbf{AV}^3$. We may ignore the axioms in **AV**, recalling the proof of Proposition 4.11. Axioms **D3**, **D4**, **D5** and **D6** do not break the strengthening of the hypothesis. In fact, the proof for all axioms except **D10** and **D11** is immediate. In the case of these axioms, we can use the strengthening of the hypothesis, and conclude that both sides of the equality are in $\Sigma^{\Box}_{EXP+AVs^3}$. □

Proposition A.2. *If* $\mathbf{SL} \vdash x = y$*, then* x *is a term in* $\Sigma^{\Box}_{SLs}$ *if, and only if,* y *is a term in* $\Sigma^{\Box}_{SLs}$*.*

Proof. Analogous to Proposition A.1. □

Corollary A.3. *If* $D(x)$ *then* x *in* $\Sigma^{\Box}_{EXP+AVs^3}$*.*

Proof. Immediate from the proof of Proposition A.1. □

Proposition A.4. *If* x *is a term in* $\Sigma_{EXP+AVs^3}$ *and* $y = \pm(y_0, \ldots, y_n)$ *is the unique normal form (Definition 4.6) of* x*, then there exists a* $y_i = \upsilon$*.*

Proof. It is an invariant over the axioms of $\mathbf{FDN}s + \mathbf{AV}s^3$. □

Corollary A.5. *Terms in* $\Sigma_{EXP+AVs^3}$ *are non-dogmatic.*

Lemma A.6. *If* $\pm(x)$*,* $\pm(y)$*,* $\pm(z)$ *are unique normal forms, and at least* $\pm(x)$ *or* $\pm(y)$ *is non-dogmatic, and* $\mathbf{FDN} + \mathbf{AV}^3 \vdash \pm(x) + \pm(y) = \pm(z)$*, then* $\mathbf{FDN}s + \mathbf{AV}s^3 \vdash \pm(x) + \pm(y) = \pm(z)$*.*

Proof. As $\pm(x)$ and $\pm(y)$ are in the unique normal form, take $x = x_1, \ldots, x_n$ and $y = y_1, \ldots, y_m$. If we apply **D7**, on $\pm(x) + \pm(y)$, the result is in $\Sigma^{\Box}_{EXP+AVs^3}$,

but not in $\Sigma_{EXP+AVs^3}$, let alone in the unique normal form. Recall Proposition A.1, disallowing destructive interference of counting or dogmatic operators to $\Sigma^{\square}_{EXP+AVs^3}$ terms. Hence, we need to apply **D9** and **D8** zero or more times to the result, then apply **D10** or **D11**. We can apply induction to the number of υ's in x (or y symmetrically). We assumed that there is at least one υ in x (or y). The base case of our induction is therefore that $\mathbf{FDN}s + \mathbf{AV}s^3 \vdash \pm(x) + \pm(y) = \pm(y)$, for dogmatic y, and $x = u, x'$ for dogmatic x', which holds. The induction step is that $\mathbf{FDN}s + \mathbf{AV}s^3 \vdash \pm((\beta)_l, (\delta)_m, (\upsilon)_{n+1}) + \pm((\beta)_{l'}, (\delta)_{m'}, (\upsilon)_{n'}) = \pm((\beta)_{l\cdot n'+l'\cdot n+l}, (\delta)_{m\cdot n'+m'\cdot n+m'}, (\upsilon)_{n\cdot n'+n'})$, which holds. □

Lemma A.7. *If $\pm(x)$, $\pm(y)$, $\pm(z)$ are unique normal forms, and $\mathbf{FDN} + \mathbf{AV}^3 \vdash \pm(x) \cdot \pm(y) = \pm(z)$, then $\mathbf{FDN}s + \mathbf{AV}s^3 \vdash \pm(x) \cdot \pm(y) = \pm(z)$.*

Proof. As $\pm(x)$ and $\pm(y)$ are in the unique normal form, take $x = x_1, \ldots, x_n$ and $y = y_1, \ldots, y_m$. If we apply **D12**, on $\pm(x) \cdot \pm(y)$, the result is in $\Sigma^{\square}_{EXP+AVs^3}$, but not in $\Sigma_{EXP+AVs^3}$, let alone in the unique normal form. Recall Proposition A.1, disallowing destructive interference of counting or dogmatic operators to $\Sigma^{\square}_{EXP+AVs^3}$ terms. We need to apply **D13** and **D14** zero or more times to the result, then apply **D15**, **D16** or **D17**. We can apply induction to the number of atoms in x and y. The base case is that x and y are atoms, of which there are nine cases, all are immediate instances of $\mathbf{D11}_\infty$. The inductive case splits x or y, reducing the number of atoms in each case.

We need to prove the case of splitting x and the case of splitting y. If we split x, then $\mathbf{FDN}s + \mathbf{AV}s^3 \vdash \pm((\beta)_{l+l'}, (\delta)_{m+m'}, (\upsilon)_{n+n'}) \cdot \pm((\beta)_{l''}, (\delta)_{m''}, (\upsilon)_{n''}) = \pm((\beta)_{l\cdot l''+l'\cdot l''}, (\delta)_{l\cdot m''+l'\cdot m''}, (\upsilon)_{l\cdot n''+(m+n)\cdot(l''+m''+n'')+l'\cdot n''+(m'+n')\cdot(l''+m''+n'')})$, which by basic calculus is the required result. If we split y, then $\mathbf{FDN}s + \mathbf{AV}s^3 \vdash \pm((\beta)_l, (\delta)_m, (\upsilon)_n) \cdot \pm((\beta)_{l'+l''}, (\delta)_{m'+m''}, (\upsilon)_{n'+n''}) = \pm((\beta)_{l\cdot l'+l\cdot l''}, (\delta)_{l\cdot m'+l\cdot m''}, (\upsilon)_{l\cdot n''+(m+n)\cdot(l''+m''+n'')+l'\cdot n''+(m'+n')\cdot(l''+m''+n'')})$, which again by basic calculus is the required result. □

Lemma A.8. *If $\pm(x)$ and $\pm(y)$ are unique normal forms, and $\mathbf{FDN} + \mathbf{AV}^3 \vdash \overline{\pm(x)} = \pm(y)$, then $\mathbf{FDN}s + \mathbf{AV}s^3 \vdash \overline{\pm(x)} = \pm(y)$.*

Proof. As $\pm(x)$ and $\pm(y)$ are in the unique normal form, take $x = x_1, \ldots, x_n$ and $y = y_1, \ldots, y_m$. If we apply **D18**, on $\overline{\pm(x)}$, the result is in $\Sigma^{\square}_{EXP+AVs^3}$, but not in $\Sigma_{EXP+AVs^3}$, let alone in the unique normal form. Recall Proposition A.1, disallowing destructive interference of counting or dogmatic operators to $\Sigma^{\square}_{EXP+AVs^3}$ terms. We need to apply **D19** zero or more times to the result, then apply **D20**, **D21** or **D22**. We can apply induction to the number of atoms in x. The base case is that x is an atom, in which case $\mathbf{D17}_\infty$ can immediately be applied. The inductive case splits x, reducing the number of atoms in each case. If we split x, then $\mathbf{FDN}s + \mathbf{AV}s^3 \vdash \overline{\pm((\beta)_{l+l'}, (\delta)_{m+m'}, (\upsilon)_{n+n'})} = \pm((\beta)_{m+m'}, (\delta)_{l+l'}, (\upsilon)_{n+n'})$, satisfying the requirement. □

Lemma A.9. *If $\pm(x)$, $\pm(y)$, $\pm(z)$ are unique normal forms, and $\mathbf{FDN} + \mathbf{AV}^3 \vdash \pm(x) \wedge \pm(y) = \pm(z)$, then $\mathbf{FDN}s + \mathbf{AV}s^3 \vdash \pm(x) \wedge \pm(y) = \pm(z)$.*

Proof. As $\pm(x)$ and $\pm(y)$ are in the unique normal form, take $x = x_1, \ldots, x_n$ and $y = y_1, \ldots, y_m$. If we apply **D23**on $\pm(x) \cdot \pm(y)$, the result is in $\Sigma^{\square}_{EXP+AVs^3}$,

but not in $\Sigma_{EXP+AVs^3}$, let alone in the unique normal form. Recall Proposition A.2, disallowing destructive interference of counting or dogmatic operators to $\Sigma^{\square}_{EXP+AVs^3}$ terms. Using induction, it is immediate that $((\beta)_l, (\delta)_m, (\upsilon)_n) \boxtimes y = (y)_l, (P_\delta(y))_m, (P_\upsilon(y))_n$ is the only term without the $\_ \boxtimes \_$ operator. By induction, $P_\delta((\beta)_{l'}, (\delta)_{m'}, (\upsilon)_{n'}) = (\delta)_{l'+m'+n'}$ is the only term not containing $P_\delta(\_)$. Furthermore, $P_\upsilon((\beta)_{l'}, (\delta)_{m'}, (\upsilon)_{n'}) = (\upsilon)_{l'+n'}, (\delta)_{m'}$ is the only term not containing $P_\upsilon(\_)$. Combining the two equalities to get the only equal statement without $\_ \boxtimes \_$, $P_\delta(\_)$ and $P_\upsilon(\_)$ we get $((\beta)_l, (\delta)_m, (\upsilon)_n) \boxtimes ((\beta)_{l'}, (\delta)_{m'}, (\upsilon)_{n'}) = (\beta)_{l \cdot l'}, (\delta)_{l \cdot m'}, (\upsilon)_{l \cdot n'}, (\delta)_{m \cdot (l'+m'+n')}, (\upsilon)_{n \cdot (l'+n')}, (\delta)_{n \cdot m'} = (\beta)_{l \cdot l'}, (\delta)_{m \cdot (l'+m'+n') + m' \cdot (l+n)}, (\upsilon)_{n \cdot l' + l \cdot n' + n \cdot n'}$.

□

B

Omitted Proofs of Part II

B.1 Omitted Proof of Trust Chaining Formula

In this section of the appendix, we derive several auxiliary equations which we then use to detail the steps omitted in the proof of Theorem 8.1. For a more concise formulation, we denote an interaction history (φ_s, φ_f) simply as φ.

Proposition B.1. *In the Beta model it holds that*

$$f_{R_B}(b|O_B^A{=}(x_s, x_f), E_B{=}\mathrm{F}) = f_{\mathrm{B}}(b; x_s + 1, x_f + 2.$$

Proof. Apply Bayes' theorem to the left hand side to get:

$$\frac{P(O_B^A{=}(x_s, x_f), E_B{=}\mathrm{F}|R_B{=}b) \cdot P(R_B{=}b)}{P(O_B^A{=}(x_s, x_f), E_B{=}\mathrm{F})}.$$

Which, using $P(R_B{=}b) = 1$ and Independency $\mathrm{I1}_S$ or Independency $\mathrm{I3}_S$ is proportional to:

$$P(O_B^A{=}(x_s, x_f)|R_B{=}b) \cdot P(E_B{=}\mathrm{F}|R_B{=}b).$$

Therefore, we have a term proportional to $b^{x_s} \cdot (1-b)^{x_f} \cdot (1-b) = b^{x_s} \cdot (1-b)^{x_f+1}$, which, due to Theorem 6.5 is proportional to $f_{\mathrm{B}}(b; x_s + 1, x_f + 2)$. If two functions are proportional, they represent the same distribution, and since both are distributions, they must be equal. □

Proposition B.2. *For discrete random variables A, B and C, if $P(A{=}a|B{=}b) = 1$ and $P(A{=}a|C{=}c) \neq 0$ then $A{=}a \perp\!\!\!\perp C{=}c|B{=}b$.*

Proof. Observe $P(A{=}a, C{=}c|B{=}b) = P(A{=}a|B{=}b, C{=}c) \cdot P(C{=}c|B{=}b) = 1 \cdot P(C{=}c|B{=}b) = P(A{=}a|B{=}b) \cdot P(C{=}c|B{=}b)$. □

Corollary B.3. *For all $W \in \mathbb{W}_{\mathbb{S}}$ it holds that:*

$$S_C^B \perp\!\!\!\perp W|E_B{=}\mathrm{S} \cap O_C^B.$$

Proof. This follows immediately from the dependencies and Proposition B.2 □

Proposition B.4. *Provided positive probabilities for X, Y, Z and W. If $X \perp\!\!\!\perp Y|Z,W$ and $X \perp\!\!\!\perp Z|Y,W$ then $X \perp\!\!\!\perp Y,Z|W$.*

Proof. Implication is known as intersection, e.g. in [Bou92]. □

Lemma B.5. *Let φ be a collection of events, not containing any of O_C^A, S_C^A (for any B), E_C and R_C. Let $\overrightarrow{O} = O_C^{B_0}, \dots O_C^{B_n}$, for $\{C, B_0, \dots B_n\} = \mathbf{A}$.Let $\psi = \psi_0, \psi_1$ be such that $\psi = E_C, \overrightarrow{O}$ Then, for any $C \in \mathbf{A}$, $f_{R_C}(c|\varphi, \psi_0) = f_{R_C}(c|\psi_0)$.*

Proof. Note that $\varphi \perp\!\!\!\perp \psi|R_C$ (Independencies $\mathrm{I1}_D$ and $\mathrm{I3}_D$) and $\varphi \perp\!\!\!\perp R_C|\psi$ (Independency $\mathrm{I2}_D$), and thus, via Proposition B.4, $\varphi \perp\!\!\!\perp \psi, R_C$. Apply the law of total probability on ψ_1 to the left hand side.

$$\sum_{\psi_1} f_{R_C}(c, \psi_1|\varphi, \psi_0).$$

Due to $\varphi \perp\!\!\!\perp \psi_0, \psi_1, R_C$, this simplifies to:

$$\sum_{\psi_1} f_{R_C}(c, \psi_1|\psi_0).$$

Reverse the law of total probability to obtain the right hand side. □

Before we perform the main equation, we introduce a collection of simpler equations, that we can apply to simplify the main analysis. These equations are numbered **A1** through **A7**.

Auxiliary equation A1

$$f(y) = \sum_w \left(f(w) \cdot \begin{cases} 1 & \text{if } w{=}y \\ 0 & \text{if } w \neq y \end{cases} \right).$$

Auxiliary equation A2

$$
\begin{aligned}
&P(O_C^B{=}w|O_B^A{=}x, S_C^B{=}y, E_B{=}\text{S})\\
&\{\text{Bayes' theorem.}\}\\
=&\frac{\begin{array}{c}P(S_C^B{=}y|O_B^A{=}x, O_C^B{=}w, E_B{=}\text{S})\\ \cdot P(O_C^B{=}w|O_B^A{=}x, E_B{=}\text{S})\end{array}}{\sum\limits_{w'\in O_B^C}\Big(\begin{array}{c}P(S_C^B{=}y|O_B^A{=}x, O_C^B{=}w', E_B{=}\text{S})\\ \cdot P(O_C^B{=}w'|O_B^A{=}x, E_B{=}\text{S})\Big)\end{array}}\\
&\{\text{Corollary B.3.}\}\\
=&\frac{P(S_C^B{=}y|O_C^B{=}w, E_B{=}\text{S})\cdot P(O_C^B{=}w|O_B^A{=}x, E_B{=}\text{S})}{\sum_{w'\in O_B^C} P(S_C^B{=}y|O_C^B{=}w', E_B{=}\text{S})\cdot P(O_C^B{=}w'|O_B^A{=}x, E_B{=}\text{S})}\\
&\{\text{Apply Dependency D4}_S\text{ to the first factor in denominator, evaluating}\\
&\text{to 1 iff } w'=y \text{ and } \mathbf{A1} \text{ setting } w'=y.\}\\
=&P(S_C^B{=}y|O_C^B{=}w, E_B{=}\text{S})\cdot\frac{P(O_C^B{=}w|O_B^A{=}x, E_B{=}\text{S})}{P(O_C^B{=}y|O_B^A{=}x, E_B{=}\text{S})}\\
&\{\text{If } w=y \text{ then both terms equal one,}\\
&\text{otherwise, the fist term equals zero, via Dependency D4}_S.\}\\
=&\begin{cases}1 & \text{if } w{=}y\\ 0 & \text{if } w\neq y\end{cases}.
\end{aligned}
$$

Auxiliary equation A4(φ)

$$
\begin{aligned}
&P(O_C^B{=}\varphi|O_B^A{=}x, E_B{=}u)\\
&\{\text{Law of total probability on } R_C.\}\\
=&\int_0^1 P(O_C^B{=}\varphi|O_B^A{=}x, E_B{=}u, R_C{=}c)\\
&\qquad\cdot f_{R_C}(c|O_B^A{=}x, E_B{=}u)\,\mathrm{d}c\\
&\{\text{Independency I1}_S\text{ on } O_B^A{=}x \text{ and } E_B{=}u \text{ and Lemma B.5}\}\\
&\text{on } O_B^A=x \text{ and } E_B=u.\}\\
=&\int_0^1 P(O_C^B{=}\varphi|R_C{=}c)\cdot f_{R_C}(c)\,\mathrm{d}c\\
&\{\text{Apply Dependency D3}_S\text{ and Dependency D1}_S.\}\\
=&\int_0^1 \binom{\varphi_s+\varphi_f}{\varphi_s} c^{\varphi_s}(1-c)^{\varphi_f}\cdot\lambda(\varphi_s+\varphi_f)\cdot 1\,\mathrm{d}c\\
&\{\text{Calculus.}\}\\
=&\lambda(\varphi_s+\varphi_f)\cdot\binom{\varphi_s+\varphi_f}{\varphi_s}\cdot\frac{\varphi_s!\varphi_f!}{(\varphi_s+\varphi_f+1)!}.
\end{aligned}
$$

Auxiliary equation A5

$$
\begin{aligned}
&P(E_B = \mathrm{s}|O_B^A = x)\\
&\{\text{Law of total probability over } R_B.\}\\
=&\int_0^1 P(E_B = \mathrm{s}|O_B^A = x, R_B = b) \cdot f_{R_B}(b|O_B^A = x)\,\mathrm{d}b\\
&\{\text{Independency I3}_S \text{ on } O_B^A.\}\\
=&\int_0^1 P(E_B = \mathrm{s}|R_B = b) \cdot f_{R_B}(b|O_B^A = x)\,\mathrm{d}b\\
&\{\text{Apply Dependency D2}_S \text{ and Theorem 6.5.}\}\\
=&\int_0^1 b \cdot f_{\mathrm{B}}(b; x_s + 1, x_f + 1)\,\mathrm{d}b\\
&\{\text{Expected value of a beta distribution, Proposition 6.4.}\}\\
=&\frac{x_s + 1}{x_s + x_f + 2}.
\end{aligned}
$$

Auxiliary equation A6

$$
\begin{aligned}
&P(S_C^B{=}y|O_B^A{=}x, E_B{=}\mathrm{s})\\
&\{\text{Law of total probability over } O_C^B.\}\\
=&\sum_{w' \in O_C^B} \Big(P(S_C^B{=}y|O_B^A{=}x, E_B{=}\mathrm{s}, O_C^B{=}w')\\
&\qquad \cdot P(O_C^B{=}w'|O_B^A{=}x, E_B{=}\mathrm{s})\Big)\\
&\{\text{Apply Corollary B.3.}\}\\
=&\sum_{w' \in O_C^B} \Big(P(S_C^B{=}y|E_B{=}\mathrm{s}, O_C^B{=}w') \cdot P(O_C^B{=}w'|O_B^A{=}x, E_B{=}\mathrm{s})\Big)\\
&\{\text{Apply Dependency D4}_S, \text{ and } \mathbf{A1}.\}\\
=&P(O_C^B{=}y|O_B^A{=}x, E_B{=}\mathrm{s})\\
&\{\text{Apply } \mathbf{A4}(y).\}\\
=&\lambda(y_s + y_f) \cdot \binom{y_s + y_f}{y_s} \cdot \frac{y_s! y_f!}{(y_s + y_f + 1)!}.
\end{aligned}
$$

Auxiliary equation A7

$$
\begin{aligned}
&P(S_C^B{=}y|O_B^A{=}x, O_C^B{=}\varphi, E_B{=}\text{F}) \\
&\{\text{Law of total probability on } R_B.\} \\
=&\int_0^1 P(S_C^B{=}y|O_B^A{=}x, O_C^B{=}\varphi, E_B{=}\text{F}, R_B{=}b) \\
&\qquad \cdot f_{R_B}(b|O_B^A{=}x, O_C^B{=}\varphi, E_B{=}\text{F})\,\mathrm{d}b \\
&\{\text{Independency I4}_S \text{ on } O_B^A = x \text{ and Lemma B.5, to cancel } O_C^B{=}\varphi.\} \\
=&\int_0^1 P(S_C^B{=}y|O_C^B{=}\varphi, E_B{=}\text{F}, R_B{=}b) \cdot f_{R_B}(b|O_B^A{=}x, E_B{=}\text{F})\,\mathrm{d}b \\
&\{\text{Apply Dependency D5}_S \text{ and Lemma B.1.}\} \\
=&\int_0^1 \chi^B(b, \varphi_s, \varphi_f)(y_s, y_f) \cdot f_{\text{B}}(b; x_s + 1, x_f + 2)\ \mathrm{d}b.
\end{aligned}
$$

The semantics of equations eq^1–eq^5

With help of the auxiliary equations we can now prove Theorem 8.1.

$$
\begin{aligned}
\mathbf{eq}^{\mathbf{1}}_{\varphi}(c) &= P(R_C{=}c|O_B^A{=}x, S_C^B{=}y, E_B{=}u, O_C^B{=}w), \\
\mathbf{eq}^{\mathbf{2}} &= P(E_B{=}\text{S}|O_B^A{=}x, S_C^B{=}y), \\
\mathbf{eq}^{\mathbf{3}} &= P(O_C^B{=}w|O_B^A{=}x, S_C^B{=}y, E_B{=}\text{F}), \\
\mathbf{eq}^{\mathbf{4}} &= P(S_C^B{=}y|O_B^A{=}x, E_B{=}\text{S}), \\
\mathbf{eq}^{\mathbf{5}}_{\varphi} &= P(S_C^B{=}y, O_C^B{=}\varphi|O_B^A{=}x, E_B{=}\text{F}).
\end{aligned}
$$

Main equation

Using $\mathbf{eq^1}$, $\mathbf{eq^2}$, and $\mathbf{eq^3}$, we can formulate Formula (8.1):

$$
\begin{aligned}
& P(R_C{=}c|O_B^A{=}x, S_C^B{=}y) \\
& \{\text{Law of total probability on } E_B.\} \\
=& P(R_C{=}c|O_B^A{=}x, S_C^B{=}y, E_B{=}\text{S}) \cdot P(E_B{=}\text{S}|O_B^A{=}x, S_C^B{=}y) \\
& + P(R_C{=}c|O_B^A{=}x, S_C^B{=}y, E_B{=}\text{F}) \cdot P(E_B{=}\text{F}|O_B^A{=}x, S_C^B{=}y) \\
& \{\text{Law of total probability on } O_C^B.\} \\
=& \sum_{w \in O_C^B} \Big(P(R_C{=}c|O_B^A{=}x, S_C^B{=}y, E_B{=}\text{S}, O_C^B{=}w) \\
& \qquad \cdot\ P(O_C^B{=}w|O_B^A{=}x, S_C^B{=}y, E_B{=}\text{S}) \Big) \\
& \qquad \cdot P(E_B{=}\text{S}|O_B^A{=}x, S_C^B{=}y) \\
& + \sum_{w \in O_C^B} \Big(P(R_C{=}c|O_B^A{=}x, S_C^B{=}y, E_B{=}\text{F}, O_C^B{=}w) \\
& \qquad \cdot\ P(O_C^B{=}w|O_B^A{=}x, S_C^B{=}y, E_B{=}\text{F}) \Big) \\
& \qquad \cdot P(E_B{=}\text{F}|O_B^A{=}x, S_C^B{=}y) \\
& \{\text{Apply } \mathbf{A2} \text{ and } \mathbf{A1}.\} \\
=& P(R_C{=}c|O_B^A{=}x, S_C^B{=}y, E_B{=}\text{S}, O_C^B{=}y) \\
& \qquad \cdot P(E_B{=}\text{S}|O_B^A{=}x, S_C^B{=}y) \\
& + \sum_{w \in O_C^B} \Big(P(R_C{=}c|O_B^A{=}x, S_C^B{=}y, E_B{=}\text{F}, O_C^B{=}w) \\
& \qquad \cdot\ P(O_C^B{=}w|O_B^A{=}x, S_C^B{=}y, E_B{=}\text{F}) \Big) \\
& \qquad \cdot P(E_B{=}\text{F}|O_B^A{=}x, S_C^B{=}y) \\
& \{\text{Apply } \mathbf{eq^1}, \mathbf{eq^2} \text{ and } \mathbf{eq^3}.\} \\
=& \mathbf{eq^1}_{y_s,y_f}(c) \cdot \mathbf{eq^2} + \sum_{w \in O_C^B} (\mathbf{eq^1}_{w_s,w_f}(c) \cdot \mathbf{eq^3} \cdot (1 - \mathbf{eq^2})).
\end{aligned}
$$

Equation for $\mathbf{eq^1}_\varphi(c)$

Now we derive the correctness of $\mathbf{eq^1}$:

$$
\begin{aligned}
& P(R_C{=}c|O_B^A{=}x, S_C^B{=}y, E_B{=}u, O_C^B{=}\varphi) \\
& \{\text{Lemma B.5 on } O_B^A{=}x,\ S_C^B{=}y,\ E_B{=}u.\} \\
=\ & P(R_C{=}c|O_C^B{=}\varphi) \\
& \{\text{Let } \varphi = (\varphi_s, \varphi_f) \text{ and apply Theorem 6.5.}\} \\
=\ & f_{\text{B}}(c; \varphi_s + 1, \varphi_f + 1).
\end{aligned}
$$

Equation for $\mathbf{eq^2}$

Now we derive the correctness of $\mathbf{eq^2}$ using **A5b**, $\mathbf{eq^4}$ and $\mathbf{eq^5}$:

$$
\begin{aligned}
& P(E_B{=}\text{S}|O_B^A{=}x, S_C^B{=}y) \\
& \{\text{Bayes' theorem.}\} \\
=& \frac{P(S_C^B{=}y|O_B^A{=}x, E_B{=}\text{S}) \cdot P(E_B{=}\text{S}|O_B^A{=}x)}{P(S_C^B{=}y|O_B^A{=}x, E_B{=}\text{S}) \cdot P(E_B{=}\text{S}|O_B^A{=}x) \\ & + P(S_C^B{=}y|O_B^A{=}x, E_B{=}\text{F}) \cdot P(E_B{=}\text{F}|O_B^A{=}x)} \\
& \{\text{Apply } \mathbf{A5} \text{ and cancel denominators.}\} \\
=& \frac{P(S_C^B{=}y|O_B^A{=}x, E_B{=}\text{S}) \cdot (x_s+1)}{P(S_C^B{=}y|O_B^A{=}x, E_B{=}\text{S}) \cdot (x_s+1) + P(S_C^B{=}y|O_B^A{=}x, E_B{=}\text{F}) \cdot (x_f+1)} \\
& \{\text{Law of total probability over } O_C^B.\} \\
=& \frac{P(S_C^B{=}y|O_B^A{=}x, E_B{=}\text{S}) \cdot (x_s+1)}{P(S_C^B{=}y|O_B^A{=}x, E_B{=}\text{S}) \cdot (x_s+1) \\ & + \sum_{w' \in O_C^B} P(S_C^B{=}y, O_C^B{=}w'|O_B^A{=}x, E_B{=}\text{F}) \cdot (x_f+1)} \\
=& \frac{\mathbf{eq^4} \cdot (x_s+1)}{\mathbf{eq^4} \cdot (x_s+1) + \sum_{w' \in O_C^B} \mathbf{eq^5_{w'}} \cdot (x_f+1)}.
\end{aligned}
$$

Equation for $\mathbf{eq^3}$

Now we derive the correctness of $\mathbf{eq^3}$ using $\mathbf{eq^5}$:

$$
\begin{aligned}
& P(O_C^B{=}w|O_B^A{=}x, S_C^B{=}y, E_B{=}\text{F}) \\
& \{\text{Bayes' theorem.}\} \\
=& \frac{P(S_C^B{=}y, O_C^B{=}w|O_B^A{=}x, E_B{=}\text{F})}{P(S_C^B{=}y|O_B^A{=}x, E_B{=}\text{F})} \\
& \{\text{Law of total probability over } O_C^B.\} \\
=& \frac{P(S_C^B{=}y, O_C^B{=}w|O_B^A{=}x, E_B{=}\text{F})}{\sum_{w' \in O_C^B} P(S_C^B{=}y, O_C^B{=}w'|O_B^A{=}x, E_B{=}\text{F})} \\
& \{\text{Apply equation } \mathbf{eq^5}.\} \\
=& \frac{\mathbf{eq^5_w}}{\sum_{w' \in O_C^B} \mathbf{eq^5_{w'}}}.
\end{aligned}
$$

Equation for $\mathbf{eq^4}$

Now we derive the correctness of $\mathbf{eq^4}$ using **A6**:

$$
\begin{aligned}
& P(S_C^B{=}y|O_B^A{=}x, E_B{=}\mathrm{S}) \\
& \{\text{Apply } \mathbf{A6}.\} \\
=& \lambda(y_s+y_f)\cdot\binom{y_s+y_f}{y_s}\cdot\frac{y_s!y_f!}{(y_s+y_f+1)!}.
\end{aligned}
$$

Equation for $\mathbf{eq^5_\varphi}$

Now we derive the correctness of $\mathbf{eq^5}$ using **A4** and **A7**:

$$
\begin{aligned}
& P(S_C^B{=}y, O_C^B{=}\varphi|O_B^A{=}x, E_B{=}\mathrm{F}) \\
& \{\text{Conjunction.}\} \\
=& P(S_C^B{=}y|O_B^A{=}x, O_C^B{=}\varphi, E_B{=}\mathrm{F})\cdot P(O_C^B{=}\varphi|O_B^A{=}x, E_B{=}\mathrm{F}) \\
& \{\text{Apply } \mathbf{A4} \text{ and } \mathbf{A7}.\} \\
=& \int_0^1 \chi^B(b,\varphi_s,\varphi_f)(y_s,y_f)\cdot f_{\mathrm{B}}(b;x_s+1,x_f+2)\ \mathrm{d}b\cdot \\
& \qquad \lambda(\varphi_s+\varphi_f)\cdot\binom{\varphi_s+\varphi_f}{\varphi_s}\cdot\frac{\varphi_s!\varphi_f!}{(\varphi_s+\varphi_f+1)!}.
\end{aligned}
$$

B.2 Omitted Proof for Modularity

In this appendix, we derive several auxiliary equations which we then use to detail the steps omitted in the proof of Theorem 8.9. The main equation proving Theorem 8.9 can be found at the end.

Corollary B.6. *For all $W \in \mathbb{W}_{\mathbb{S}}$ it holds that:*

$$S_C^B \perp\!\!\!\perp W | R_B{=}b \cap E_B{=}u \cap O_C^B.$$

Proof. This follows immediately from Corollary B.3 and Independency $I4_S$, □

Auxiliary equation B1

$$\begin{aligned}
&f_{R_D}(d|R_C{=}c, R_B{=}b, E_B{=}u, O_C^B{=}w, \psi, \bigcap_D O_D^{E_i} = e_i, E_D = u) \\
&\{\text{Independency I2.}\} \\
=&f_{R_D}(d|R_C{=}c, \psi, \bigcap_D O_D^{E_i} = e_i, E_D = u)
\end{aligned}$$

Auxiliary equation B2

$$\begin{aligned}
&P(E_D = u|R_C{=}c, R_B{=}b, E_B{=}u, O_C^B{=}w, \psi, \bigcap_D O_D^{E_i} = e_i, R_D{=}d) \\
&\{\text{Independency } I3_S.\} \\
=&P(E_D = u|R_C{=}c, \psi, \bigcap_D O_D^{E_i} = e_i, R_D{=}d)
\end{aligned}$$

Auxiliary equation B3

$$\begin{aligned}
&P(E_D = u|R_C{=}c, R_B{=}b, E_B{=}u, O_C^B{=}w, \psi, \bigcap_D O_D^{E_i} = e_i, R_D{=}d) \\
&\{\text{Independency } I1_S.\} \\
=&P(E_D = u|R_C{=}c, \psi, \bigcap_D O_D^{E_i} = e_i, R_D{=}d)
\end{aligned}$$

Auxiliary equation B4

$$
\begin{aligned}
& f_{R_D}(d|R_C{=}c, R_B{=}b, E_B{=}u, O_C^B{=}w, \psi) \\
& \{\text{Law of total probability.}\} \\
=& \sum f_{R_D}(d, \bigcap_D O_D^{E_i} = e_i, E_D = u | R_C{=}c, R_B{=}b, E_B{=}u, O_C^B{=}w, \psi) \\
& \{\text{Intersection of } \mathbf{B1}, \mathbf{B2}, \text{ and } \mathbf{B3}.\} \\
=& \sum f_{R_D}(d, \bigcap_D O_D^{E_i} = e_i, E_D = u | R_C{=}c, \psi) \\
& \{\text{Law of total probability.}\} \\
=& f_{R_D}(d|R_C{=}c, \psi)
\end{aligned}
$$

Auxiliary equation B5

$$
\begin{aligned}
& P(O_D^A{=}x'|R_C{=}c, R_B{=}b, E_B{=}u, O_C^B{=}w, \psi) \\
& \{\text{Law of total probability.}\} \\
=& \int_0^1 P(O_D^A{=}x'|R_C{=}c, R_B{=}b, E_B{=}u, O_C^B{=}w, \psi, R_D{=}d) \\
& \qquad \cdot f_{R_D}(d|R_C{=}c, R_B{=}b, E_B{=}u, O_C^B{=}w, \psi) \\
& \{\text{Independency I1}_S.\} \\
=& \int_0^1 P(O_D^A{=}x'|R_C{=}c, \psi, R_D{=}d) \cdot f_{R_D}(d|R_C{=}c, R_B{=}b, E_B{=}u, O_C^B{=}w, \psi) \\
& \{\text{Apply } \mathbf{B4}.\} \\
=& \int_0^1 P(O_D^A{=}x'|R_C{=}c, \psi, R_D{=}d) \cdot f_{R_D}(d|R_C{=}c, \psi) \\
& \{\text{Law of total probability.}\} \\
=& P(O_D^A{=}x'|R_C{=}c, \psi)
\end{aligned}
$$

Auxiliary equation B6

$$
\begin{aligned}
& f_{R_D}(d, E_D{=}u'|R_C{=}c, R_B{=}b, E_B{=}u, O_C^B{=}w, \psi) \\
& \{\text{Law of total probability.}\} \\
=& \sum f_{R_D}(d, \bigcap_D O_D^{E_i} = e_i, E_D = u | R_C{=}c, R_B{=}b, E_B{=}u, O_C^B{=}w, \psi) \\
& \{\text{Intersection of } \mathbf{B1}, \mathbf{B2}, \text{ and } \mathbf{B3}.\} \\
=& \sum f_{R_D}(d, \bigcap_D O_D^{E_i} = e_i, E_D = u | R_C{=}c, \psi) \\
& \{\text{Law of total probability.}\} \\
=& f_{R_D}(d, E_D{=}u'|R_C{=}c, \psi)
\end{aligned}
$$

Auxiliary equation B7

$$
\begin{aligned}
&f_{R_D}(d, E_D{=}u', O_C^D{=}w' | R_C{=}c, R_B{=}b, E_B{=}u, O_C^B{=}w, \psi) \\
&\quad \{\text{Apply Independency I1}_S.\} \\
=&f_{R_D}(d, E_D{=}u' | R_C{=}c, R_B{=}b, E_B{=}u, O_C^B{=}w, \psi) \cdot P(O_C^D{=}w' | R_C{=}c, \psi) \\
&\quad \{\text{Apply } \mathbf{B6}.\} \\
=&f_{R_D}(d, E_D{=}u' | R_C{=}c, \psi) \cdot P(O_C^D{=}w' | R_C{=}c, \psi) \\
&\quad \{\text{Apply Independency I1}_S.\} \\
=&f_{R_D}(d, E_D{=}u', O_C^D{=}w' | R_C{=}c, \psi)
\end{aligned}
$$

Auxiliary equation B8

$$
\begin{aligned}
&P(S_C^D{=}x' | R_C{=}c, R_B{=}b, E_B{=}u, O_C^B{=}w, \psi) \\
&\{\text{Law of total probability.}\} \\
=&\sum_{u' \in \{\text{S,F}\}} \sum_{w' \in \mathbb{N}\cdot\mathbb{N}} \int_0^1 P(S_C^D{=}x' | R_C{=}c, R_B{=}b, E_B{=}u, O_C^B{=}w, \psi, R_D{=}d, E_D{=}u', O_C^D{=}w') \\
&\qquad \cdot f_{R_D}(d, E_D{=}u', O_C^D{=}w' | R_C{=}c, R_B{=}b, E_B{=}u, O_C^B{=}w, \psi) \\
&\{\text{Apply Corollary B.6.}\} \\
=&\sum_{u' \in \{\text{S,F}\}} \sum_{w' \in \mathbb{N}\cdot\mathbb{N}} \int_0^1 P(S_C^D{=}x' | R_C{=}c, \psi, R_D{=}d, E_D{=}u', O_C^D{=}w') \\
&\qquad \cdot f_{R_D}(d, E_D{=}u', O_C^D{=}w' | R_C{=}c, R_B{=}b, E_B{=}u, O_C^B{=}w, \psi) \\
&\{\text{Apply } \mathbf{B7}.\} \\
=&\sum_{u' \in \{\text{S,F}\}} \sum_{w' \in \mathbb{N}\cdot\mathbb{N}} \int_0^1 P(S_C^D{=}x' | R_C{=}c, \psi, R_D{=}d, E_D{=}u', O_C^D{=}w') \\
&\qquad \cdot f_{R_D}(d, E_D{=}u', O_C^D{=}w' | R_C{=}c, \psi) \\
&\{\text{Law of total probability.}\} \\
=&P(S_C^D{=}x' | R_C{=}c, \psi)
\end{aligned}
$$

Auxiliary equation B9

$$
\begin{aligned}
&P(\varphi | R_C{=}c, R_B{=}b, E_B{=}u, O_C^B{=}w) \\
&\quad \{\text{Repeated application of Proposition B.4, using } \mathbf{B5} \text{ and } \mathbf{B8}, \text{ for all } D.\} \\
=&\ P(\varphi | R_C{=}c).
\end{aligned}
$$

Auxiliary equation B10

$$
\begin{aligned}
& P(S_C^B = y, \varphi | R_C{=}c, R_B{=}b) \\
& \{\text{Law of total probability.}\} \\
= & \sum_{u \in \{\text{S},\text{F}\}} \sum_{w \in \mathbb{N}\cdot\mathbb{N}} P(S_C^B = y, \varphi | R_C{=}c, R_B{=}b, E_B{=}u, O_C^B{=}w) \\
& \qquad \cdot P(E_B{=}u, O_C^B{=}w | R_C{=}c, R_B{=}b) \\
& \{\text{Apply Corollary B.6.}\} \\
= & \sum_{u \in \{\text{S},\text{F}\}} \sum_{w \in \mathbb{N}\cdot\mathbb{N}} P(S_C^B = y | R_C{=}c, R_B{=}b, E_B{=}u, O_C^B{=}w) \\
& \qquad \cdot P(\varphi | R_C{=}c, R_B{=}b, E_B{=}u, O_C^B{=}w) \cdot P(E_B{=}u, O_C^B{=}w | R_C{=}c, R_B{=}b) \\
& \{\text{Auxiliary equation } \mathbf{B9}.\} \\
= & \sum_{u \in \{\text{S},\text{F}\}} \sum_{w \in \mathbb{N}\cdot\mathbb{N}} P(S_C^B = y | R_C{=}c, R_B{=}b, E_B{=}u, O_C^B{=}w) \cdot P(\varphi | R_C{=}c) \\
& \qquad \cdot P(E_B{=}u, O_C^B{=}w | R_C{=}c, R_B{=}b) \\
& \{\text{Law of total probability.}\} \\
= & P(S_C^B = y | R_C{=}c, R_B{=}b) \cdot P(\varphi | R_C{=}c)
\end{aligned}
$$

Auxiliary equation B11

$$
\begin{aligned}
& P(O_B^A{=}x, S_C^B{=}y, \varphi | R_C{=}c) \\
& \{\text{Law of total probability.}\} \\
= & \int_0^1 P(O_B^A{=}x, S_C^B{=}y, \varphi | R_C{=}c, R_B{=}b) \cdot f_{R_B}(b | R_C{=}c) \\
& \{\text{Similar to Proposition 8.8.}\} \\
= & \int_0^1 P(O_B^A{=}x | R_C{=}c, R_B{=}b) \cdot P(S_C^B{=}y, \varphi | R_C{=}c, R_B{=}b) \cdot f_{R_B}(b | R_C{=}c) \\
& \{\text{Auxiliary equation } \mathbf{B10}.\} \\
= & \int_0^1 P(O_B^A{=}x | R_C{=}c, R_B{=}b) \cdot P(S_C^B{=}y | R_C{=}c, R_B{=}b) \cdot P(\varphi | R_C{=}c) \cdot f_{R_B}(b | R_C{=}c) \\
& \{\text{Similar to Proposition 8.8.}\} \\
= & \int_0^1 P(O_B^A{=}x, S_C^B{=}y | R_C{=}c, R_B{=}b) \cdot P(\varphi | R_C{=}c) \cdot f_{R_B}(b | R_C{=}c) \\
& \{\text{Law of total probability.}\} \\
= & P(O_B^A{=}x, S_C^B{=}y | R_C{=}c) \cdot P(\varphi | R_C{=}c)
\end{aligned}
$$

Main equation

$$
\begin{aligned}
& f_{R_C}(c|O_B^A{=}(x_s,x_f),S_C^B{=}(y_s,y_f),\varphi) \\
& \{\text{Bayes theorem.}\} \\
=& \frac{P(O_B^A{=}(x_s,x_f),S_C^B{=}(y_s,y_f),\varphi|R_C{=}c)\cdot f_{R_C}(c)}{P(O_B^A{=}(x_s,x_f),S_C^B{=}(y_s,y_f),\varphi)} \\
& \{\text{Auxiliary equation } \mathbf{B11}.\} \\
=& \frac{P(O_B^A{=}(x_s,x_f),S_C^B{=}(y_s,y_f)|R_C{=}c)\cdot P(\varphi|R_C{=}c)\cdot f_{R_C}(c)}{P(O_B^A{=}(x_s,x_f),S_C^B{=}(y_s,y_f),\varphi)} \\
& \{\text{Change constant factor.}\} \\
\propto& \frac{P(O_B^A{=}(x_s,x_f),S_C^B{=}(y_s,y_f)|R_C{=}c)\cdot P(\varphi|R_C{=}c)\cdot f_{R_C}(c)}{P(O_B^A{=}(x_s,x_f),S_C^B{=}(y_s,y_f))\cdot P(\varphi)} \\
& \{\text{Apply Dependency D1.}\} \\
=& \frac{P(O_B^A{=}(x_s,x_f),S_C^B{=}(y_s,y_f)|R_C{=}c)\cdot f_{R_C}(c)\cdot P(\varphi|R_C{=}c)\cdot f_{R_C}(c)}{P(O_B^A{=}(x_s,x_f),S_C^B{=}(y_s,y_f))\cdot P(\varphi)} \\
& \{\text{Bayes theorem (2x).}\} \\
=& f_{R_C}(c|\varphi)\cdot f_{R_C}(c|S_C^B{=}(y_s,y_f),O_B^A{=}(x_s,x_f))
\end{aligned}
$$

B.3 Omitted Proof Generalised Trust Chaining

First we generalise Lemma B.1.

Proposition B.7. *Let $g(b) = f_{R_B}(b|\psi)$. In the Beta model it holds that*

$$f_{R_B}(b|\psi, E_B{=}\text{F}) = g(x) \cdot \frac{1-x}{1-\mathbf{E}(g)}.$$

Proof. Apply Bayes' theorem to the left hand side to get:

$$\frac{P(\psi, E_B{=}\text{F}|R_B{=}b) \cdot P(R_B{=}b)}{P(\psi, E_B{=}\text{F})}$$

Which, using $P(R_B{=}b) = 1$ and Independency I3_S is proportional to:

$$P(\psi|R_B{=}b) \cdot P(E_B{=}\text{F}|R_B{=}b)$$

Therefore, we have a term proportional to $g(b) \cdot (1-b)$. The distribution that it represents is $\frac{g(b)\cdot(1-b)}{\int_0^1 g(y)\cdot(1-y)\ \mathrm{d}y} = \frac{g(b)\cdot(1-b)}{1-\int_0^1 g(y)\cdot y\ \mathrm{d}y} = \frac{g(b)\cdot(1-b)}{1-\mathbf{E}(g)}$. □

Observe that in **A4**, we can substitute $O_B^A{=}x$ for an arbitrary term ψ, provided ψ does not contain O_C^A or S_C^A.

Similarly, in **A6**, we can substitute $O_B^A{=}x$ for an arbitrary term ψ.

Finally, we need to modify the remaining auxiliary equations that contain O_B^A, namely **A5**, **A7**. As a consequence, we need to update the equations dependent on these, namely $\mathbf{eq^2}$ and $\mathbf{eq^5}$.

Auxiliary equation A5b

$$\begin{aligned}
&P(E_B = \text{S}|\psi)\\
&\{\text{Law of total probability over } R_B.\}\\
=&\int_0^1 P(E_B = \text{S}|\psi, R_B = b) \cdot f_{R_B}(b|\psi)\,\mathrm{d}b\\
&\{\text{Independency } \text{I3}_S \text{ on } O_B^A.\}\\
=&\int_0^1 P(E_B = \text{S}|R_B = b) \cdot f_{R_B}(b|\psi)\,\mathrm{d}b\\
&\{\text{Apply Dependency } \text{D2}_S \text{ and the assumption } g(b) = f_{R_B}(b|\psi).\}\\
=&\int_0^1 b \cdot g(b)\,\mathrm{d}b\\
&\{\text{Definition expected value of } g.\}\\
=&\mathbf{E}(g).
\end{aligned}$$

Auxiliary equation A7b

$$
\begin{aligned}
&P(S_C^B{=}y|\psi, O_C^B{=}\varphi, E_B{=}\text{F}) \\
&\{\text{Law of total probability on } R_B.\} \\
=&\int_0^1 P(S_C^B{=}y|\psi, O_C^B{=}\varphi, E_B{=}\text{F}, R_B{=}b) \\
&\qquad \cdot f_{R_B}(b|\psi, O_C^B{=}\varphi, E_B{=}\text{F})\,\mathrm{d}b \\
&\{\text{Independency I4}_S \text{ on } \psi \text{ and Lemma B.5 on } O_C^B = \varphi.\} \\
=&\int_0^1 P(S_C^B{=}y|O_C^B{=}\varphi, E_B{=}\text{F}, R_B{=}b) \cdot f_{R_B}(b|\psi, E_B{=}\text{F})\,\mathrm{d}b \\
&\{\text{Apply Dependency D5}_S \text{ and Lemma B.7.}\} \\
=&\int_0^1 \chi^B(b, \varphi_s, \varphi_f)(y_s, y_f) \cdot g(x) \cdot \frac{1-x}{1-\mathbf{E}(g)}\ \mathrm{d}b.
\end{aligned}
$$

Equation for $\mathbf{eq^{2b}}$

Now we derive the correctness of $\mathbf{eq^{2b}}$ using **A5b**, $\mathbf{eq^4}$ and $\mathbf{eq^{5b}}$:

$$
\begin{aligned}
&P(E_B{=}\text{S}|\psi, S_C^B{=}y) \\
&\{\text{Bayes' theorem.}\} \\
=&\frac{P(S_C^B{=}y|\psi, E_B{=}\text{S}) \cdot P(E_B{=}\text{S}|\psi)}{P(S_C^B{=}y|\psi, E_B{=}\text{S}) \cdot P(E_B{=}\text{S}|\psi) + P(S_C^B{=}y|\psi, E_B{=}\text{F}) \cdot P(E_B{=}\text{F}|\psi)} \\
&\{\text{Apply } \mathbf{A5b}.\} \\
=&\frac{P(S_C^B{=}y|\psi, E_B{=}\text{S}) \cdot \mathbf{E}(g)}{P(S_C^B{=}y|\psi, E_B{=}\text{S}) \cdot \mathbf{E}(g) + P(S_C^B{=}y|\psi, E_B{=}\text{F}) \cdot (1-\mathbf{E}(g))} \\
&\{\text{Law of total probability over } O_C^B.\} \\
=&\frac{P(S_C^B{=}y|\psi, E_B{=}\text{S}) \cdot (x_s+1)}{P(S_C^B{=}y|\psi, E_B{=}\text{S}) \cdot (x_s+1) + \sum_{w' \in O_C^B} P(S_C^B{=}y, O_C^B{=}w'|\psi, E_B{=}\text{F}) \cdot (x_f+1)} \\
=&\frac{\mathbf{eq^4} \cdot (x_s+1)}{\mathbf{eq^4} \cdot (x_s+1) + \sum_{w' \in O_C^B} \mathbf{eq}^{\mathbf{5b}}_{w'} \cdot (x_f+1)}.
\end{aligned}
$$

Equation for $\mathbf{eq}^{\mathbf{5b}}_{\varphi}$

Now we derive the correctness of $\mathbf{eq^{5b}}$ using **A4** and **A7b**:

$$
\begin{aligned}
&P(S_C^B{=}y, O_C^B{=}\varphi|\psi, E_B{=}\text{F}) \\
&\{\text{Conjunction.}\} \\
=&P(S_C^B{=}y|\psi, O_C^B{=}\varphi, E_B{=}\text{F}) \cdot P(O_C^B{=}\varphi|\psi, E_B{=}\text{F}) \\
&\{\text{Apply } \mathbf{A4} \text{ and } \mathbf{A7b}.\} \\
=&\int_0^1 \chi^B(b, \varphi_s, \varphi_f)(y_s, y_f) \cdot g(x) \cdot \frac{1-x}{1-\mathbf{E}(g)}\ \mathrm{d}b \cdot \\
&\qquad \lambda(\varphi_s+\varphi_f) \cdot \binom{\varphi_s+\varphi_f}{\varphi_s} \cdot \frac{\varphi_s!\varphi_f!}{(\varphi_s+\varphi_f+1)!}.
\end{aligned}
$$

Bibliography

[AM09] Baptiste Alcalde and Sjouke Mauw. An algebra for trust dilution and trust fusion. In *Formal Aspects in Security and Trust*, pages 4–20. 2009.

[And04] David P. Anderson. Boinc: A system for public-resource computing and storage. In *Proceedings of the 5th IEEE/ACM International Workshop on Grid Computing*, GRID '04, pages 4–10. IEEE Computer Society, 2004.

[ARH00] Alfarez Abdul-Rahman and Stephen Hailes. Supporting trust in virtual communities. In *Proceedings of the 33rd Hawaii International Conference on System Sciences*, volume 6 of *HICSS '00*. IEEE Computer Society, 2000.

[AS64] Milton Abramowitz and Irene A. Stegun. *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables*. Applied mathematics series. Dover Publications, 1964.

[Bai86] Annette Baier. Trust and antitrust. *Ethics*, 96(2):231, 1986.

[BDM95] Joyce Berg, John Dickhaut, and Kevin McCabe. Trust, reciprocity, and social history. *Games and economic behavior*, 10(1):122–142, 1995.

[BFL96] Matt Blaze, Joan Feigenbaum, and Jack Lacy. Decentralized trust management. In *Proceedings of the 1996 IEEE Symposium on Security and Privacy*, pages 164–173. IEEE Computer Society, 1996.

[BH95] Jay B. Barney and Mark H. Hanson. Trustworthiness as a source of competitive advantage. *Long Range Planning*, 28:127–127(1), 1995.

[BHK98] John S. Breese, David Heckerman, and Carl Kadie. Empirical analysis of predictive algorithms for collaborative filtering. In *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence*, pages 43–52. Morgan Kaufmann, 1998.

[Bil95] Patrick Billingsley. *Probability and measure*. Wiley, 3 edition, 1995.

[BLB04] Sonja Buchegger and Jean-Yves Le Boudec. A robust reputation system for peer-to-peer and mobile ad-hoc networks. In *Proceedings of P2PEcon*, volume 2004, 2004.

[Bou92] Remco R. Bouckaert. Bayesian belief networks and conditional independencies. Technical Report RUU-CS-92-36, Utrecht University, The Netherlands, 1992.

[BP66] Leonard E. Baum and Ted Petrie. Statistical inference for probabilistic functions of finite state markov chains. *The annals of mathematical statistics*, 37(6):1554–1563, 1966.

[CF98] Cristiano Castelfranchi and Rino Falcone. Principles of trust for mas: Cognitive anatomy, social importance, and quantification. In *Proceedings of the International Conference on Multi Agent Systems*, pages 72–79. IEEE Computer Society, 1998.

[CH67] Thomas Cover and Peter Hart. Nearest neighbor pattern classification. *Information Theory, IEEE Transactions on*, 13(1):21–27, 1967.

[CH97] Bruce Christianson and William S. Harbison. Why isn't trust transitive? In *Security Protocols*, volume 1189 of *Lecture Notes in Computer Science*, pages 171–176. Springer Berlin Heidelberg, 1997.

[Dan03] Milan Daniel. Algebraic structures related to the consensus operator for combining of beliefs. In *Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, volume 2711 of *Lecture Notes in Computer Science*, pages 332–344. Springer Berlin / Heidelberg, 2003.

[Dem67] Arthur P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *The annals of mathematical statistics*, 38(2):325–339, 1967.

[Dem04] Robert Demolombe. Reasoning about trust: A formal logical framework. In *Trust Management*, volume 2995 of *Lecture Notes in Computer Science*, pages 291–303. Springer Berlin Heidelberg, 2004.

[DLL+10] James Davidson, Benjamin Liebald, Junning Liu, Palash Nandy, Taylor Van Vleet, Ullas Gargi, Sujoy Gupta, Yu He, Mike Lambert, Blake Livingston, and Dasarathi Sampath. The youtube video recommendation system. In *Proceedings of the fourth ACM conference on Recommender systems*, pages 293–296. ACM, 2010.

[Duf65] Richard J. Duffin. Topology of series-parallel networks. *Journal of Mathematical Analysis and Applications*, 10(2):303–318, 1965.

[Dwo06] Cynthia Dwork. Differential privacy. In *Automata, languages and programming*, pages 1–12. Springer, 2006.

[EFL+99] Carl Ellison, Bill Frantz, Butler Lampson, Ron Rivest, Brian Thomas, and Tatu Ylonen. Spki certificate theory. Technical report, IETF RFC 2693, September, 1999.

[Ell61] Daniel Ellsberg. Risk, ambiguity, and the savage axioms. *The Quarterly Journal of Economics*, 75:643–669, 1961.

[ElS11] Ehab ElSalamouny. *Probabilistic trust models in network security*. PhD thesis, University of Southampton, 2011.

[ESN10] Ehab ElSalamouny, Vladimiro Sassone, and Mogens Nielsen. Hmm-based trust model. In *Formal Aspects in Security and Trust*, pages 21–35. Springer, 2010.

[FC01] Rino Falcone and Cristiano Castelfranchi. Social trust: A cognitive approach. In *Trust and deception in virtual societies*, pages 55–90. Springer, 2001.

[Fie06] Stephen E. Fienberg. When did bayesian inference become "bayesian"? *Bayesian analysis*, 1(1):1–40, 2006.

[FZAB11] Carol J. Fung, Jie Zhang, Issam Aib, and Raouf Boutaba. Dirichlet-based trust management for effective collaborative intrusion detection networks. *Network and Service Management, IEEE Transactions on*, 8(2):79–91, 2011.

[Gam88] Diego Gambetta. Can we trust trust? In *Trust: Making and Breaking Cooperative Relations*, pages 213–237. Basil Blackwell, 1988.

[Gol05] Jennifer A. Golbeck. *Computing and applying trust in web-based social networks*. PhD thesis, University of Maryland at College Park, 2005.

[GS03] Tyrone Grandison and Morris Sloman. Specifying and analysing trust for internet applications. In *Towards the Knowledge Society*, pages 145–157. Springer, 2003.

[Gut07] Allan Gut. *Probability: A Graduate Course (Springer Texts in Statistics)*. Springer, 2007.

[Hae06] Rolf Haenni. Uncover dempster's rule where it is hidden. In *9th International Conference on Information Fusion*, pages 1–8. IEEE, 2006.

[HLHV10] Andreas Herzig, Emiliano Lorini, Jomi F. Hübner, and Laurent Vercouter. A logic of trust and reputation. *Logic Journal of IGPL*, 18(1):214–244, 2010.

[Jay57] Edwin T. Jaynes. Information theory and statistical mechanics. *Physical review*, 106(4):620, 1957.

[JDV03] Audun Jøsang, Milan Daniel, and Patrick Vannoorenberghe. Strategies for combining conflicting dogmatic beliefs. In *Proceedings of the 6th International Conference on Information Fusion*, pages 1133–1140, 2003.

[JH07] Audun Jøsang and Jochen Haller. Dirichlet reputation systems. In *The Second International Conference on Availability, Reliability and Security*, pages 112–119. IEEE Computer Society, 2007.

[JI02] Audun Jøsang and Roslan Ismail. The beta reputation system. In *Proceedings of the 15th Bled Electronic Commerce Conference*, volume 160, pages 324–337, 2002.

[JIB07] Audun Jøsang, Roslan Ismail, and Colin Boyd. A survey of trust and reputation systems for online service provision. *Decision support systems*, 43(2):618–644, 2007.

[JJ98] Audun Jøsang and Knapskog. Svein J. A metric for trusted systems. In *Proceedings of the 21st National Security Conference*, 1998.

[JKB95] Norman L. Johnson, Samuel Kotz, and Narayanaswamy Balakrishnan. *Continuous Univariate Distributions*, volume 2. Wiley, 2 edition, 1995.

[JKD05] Audun Jøsang, Claudia Keser, and Theo Dimitrakos. Can we manage trust? In *Proceedings of the Third International Conference on Trust Management (iTrust), Versailes*, pages 93–107. Springer-Verlag, 2005.

[JM05] Audun Jøsang and David McAnally. Multiplication and comultiplication of beliefs. *International Journal of Approximate Reasoning*, 38(1):19–51, 2005.

[JMP06] Audun Jøsang, Stephen Marsh, and Simon Pope. Exploring different types of trust propagation. In *Trust Management*, volume 3986 of *Lecture Notes in Computer Science*, pages 179–192. Springer Berlin Heidelberg, 2006.

[JOO10] Audun Jøsang, Stephen O'Hara, and Kristi O'Grady. Base rates for belief functions. In *Workshop on the Theory on Belief Functions*, 2010.

[Jøs97] Audun Jøsang. Artificial reasoning with subjective logic. In *Proceedings of the Second Australian Workshop on Commonsense Reasoning*, volume 48, 1997.

[Jøs01] Audun Jøsang. A logic for uncertain probabilities. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 9(03):279–311, 2001.

[Jøs02] Audun Jøsang. The consensus operator for combining beliefs. *Artificial Intelligence*, 141(1):157–170, 2002.

[JP05] Audun Jøsang and Simon Pope. Semantic constraints for trust transitivity. In *Proceedings of the 2nd Asia-Pacific conference on Conceptual modelling*, volume 43, pages 59–68. Australian Computer Society, 2005.

[Kal60] Rudolph E. Kalman. A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, 82(1):35–45, 1960.

[KBR05] Michael Kinateder, Ernesto Baschny, and Kurt Rothermel. Towards a generic trust model–comparison of various trust update algorithms. In *Trust Management*, pages 177–192. Springer, 2005.

[KG06] Yarden Katz and Jennifer Golbeck. Social network-based trust in prioritized default logic. In *Proceedings of the 21st national conference on Artificial intelligence*, volume 2 of *AAAI'06*, pages 1345–1350. AAAI Press, 2006.

[KGO10] Simon Kramer, Rajeev Goré, and Eiji Okamoto. Formal definitions and complexity results for trust relations and trust domains. In *Proceedings of the ESSLLI-affiliated Workshop on Logics in Security*, 2010.

[Kle50] Stephen C. Kleene. *Introduction to Metamathematics*. Princeton, 1950.

[KMRS12] Barbara Kordy, Sjouke Mauw, Saša Radomirović, and Patrick Schweitzer. Attack–defense trees. *Journal of Logic and Computation*, 2012. (to appear).

[KNS08] Karl Krukow, Mogens Nielsen, and Vladimiro Sassone. Trust models in ubiquitous computing. *Royal Society of London Philosophical Transactions Series A*, 366:3781–3793, 2008.

[Kol30] Andrey Kolmogorov. Sur la notion de la moyenne. *Proceedings Accademia Nazionale Lincei*, 12:388–391, 1930.

[KSGM03] Sepandar D. Kamvar, Mario T. Schlosser, and Hector Garcia-Molina. The eigentrust algorithm for reputation management in p2p networks. In *Proceedings of the 12th international conference on World Wide Web*, pages 640–651. ACM, 2003.

[LGTL85] Wen-Shing Lee, Doris L. Grosh, Frank A. Tillman, and Chang H. Lie. Fault tree analysis, methods, and applications: A review. *Reliability, IEEE Transactions on*, 34(3):194–203, 1985.

[LIB+07] Timothy E. Levin, Cynthia E. Irvine, Terry V. Benzel, Paul C. Clark, Thuy D. Nguyen, and Ganesha Bhaskara. Design principles and guidelines for security. Technical report, Monterey, California. Naval Postgraduate School, 2007.

[LSS10] Mass S. Lund, Bjørnar Solhaug, and Ketil Stølen. Evolution in relation to risk and trust management. *IEEE Computer*, 43(5):49–55, 2010.

[LW85] J. David Lewis and Andrew Weigert. Trust as a Social Reality. *Social Forces*, 63(4):967–985, 1985.

[LZL12] Yuan Liu, Jie Zhang, and Qin Li. Design of an incentive mechanism to promote honesty in e-marketplaces with limited inventory. In *Proceedings of the 14th Annual International Conference on Electronic Commerce*, pages 54–61. ACM, 2012.

[Mar94] Stephen P. Marsh. *Formalising trust as a computational concept*. PhD thesis, University of Stirling, 1994.

[MC96] D. Harrison McKnight and Norman L. Chervany. The meanings of trust. Technical report, University of Minnesota, 1996.

[MC01] D. Harrison McKnight and Norman L. Chervany. Trust and distrust definitions: One bite at a time. In *Trust in Cyber-societies*, pages 27–54. Springer, 2001.

[McE01] Robert J. McEliece. *Theory of Information and Coding*. Cambridge University Press, New York, NY, USA, 2nd edition, 2001.

[MFGL12] Francisco Moyano, Carmen Fernandez-Gago, and Javier Lopez. A conceptual framework for trust models. In *Trust, Privacy and Security in Digital Business*, volume 7449 of *Lecture Notes in Computer Science*, pages 93–104. Springer Berlin Heidelberg, 2012.

[MM02] Lik Mui and Mojdeh Mohtashemi. A computational model of trust and reputation. In *Proceedings of the 35th HICSS*, 2002.

[MRS03] Kevin A. McCabe, Mary L. Rigdon, and Vernon L. Smith. Positive reciprocity and intentions in trust games. *Journal of Economic Behavior & Organization*, 52(2):267 – 275, 2003.

[MS13a] Tim Muller and Patrick Schweitzer. A formal derivation of composite trust. In *Foundations and Practice of Security*, pages 132–148. Springer, 2013.

[MS13b] Tim Muller and Patrick Schweitzer. On beta models with trust chains. In *Proceedings of IFIP WG 11.11 International Conference on Trust Management*. Springer-Verlag Berlin Heidelberg, 2013.

[Mul11] Tim Muller. Semantics of trust. In *Formal Aspects of Security and Trust*, volume 6561 of *Lecture Notes in Computer Science*, pages 141–156. Springer Berlin / Heidelberg, 2011.

[OR94] Martin J. Osborne and Ariel Rubinstein. *A Course in Game Theory*, volume 1 of *MIT Press Books*. The MIT Press, 1994.

[Pea96] Karl Pearson. Mathematical contributions to the theory of evolution.–on a form of spurious correlation which may arise when indices are used in the measurement of organs. *Proceedings of the Royal Society of London*, 60(359-367):489–498, 1896.

[PTJL05] Jigar Patel, W. T. Luke Teacy, Nicholas R. Jennings, and Michael Luck. A probabilistic trust model for handling inaccurate reputation sources. In *Trust Management*, pages 193–209. Springer, 2005.

[Put93] Robert D. Putnam. The prosperous community: social capital and public life. *The American Prospect*, 13(1995):65–78, 1993.

[QH07] Daniele Quercia and Stephen Hailes. Mate: Mobility and adaptation with trust and expected-utility. *International Journal of Internet Technology and Secured Transactions*, 1(1), 2007.

[RHMV11] Sebastian Ries, Sheikh Mahbub Habib, Max Mühlhäuser, and Vijay Varadharajan. Certainlogic: A logic for modeling trust and uncertainty. In *Trust and Trustworthy Computing*, pages 254–261. Springer, 2011.

[Rie07] Sebastian Ries. Certain trust: a trust model for users and agents. In *Proceedings of the 2007 ACM symposium on Applied computing*, pages 1599–1604. ACM, 2007.

[RIS+94] Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, and John Riedl. Grouplens: an open architecture for collaborative filtering of netnews. In *Proceedings of the 1994 ACM conference on Computer supported cooperative work*, pages 175–186. ACM, 1994.

[SA06] James Salter and Nick Antonopoulos. Cinemascreen recommender agent: combining collaborative and content-based filtering. *Intelligent Systems, IEEE*, 21(1):35–41, 2006.

[SF02] Kari Sentz and Scott Ferson. Combination of evidence in dempster-shafer theory. Technical report, Sandia National Laboratories, 2002.

[SFE08] Eugen Staab, Volker Fusenig, and Thomas Engel. Towards trust-based acquisition of unverifiable information. In *Cooperative Information Agents XII*, volume 5180 of *LNCS*, pages 41–54. Springer Verlag, 2008.

[Sha76] Glenn Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.

[SKKR01] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th international conference on World Wide Web*, pages 285–295. ACM, 2001.

[Sob02] Joel Sobel. Can we trust social capital? *Journal of Economic Literature*, 40(1):139–154, March 2002.

[Spr79] Melvin D. Springer. *The algebra of random variables*. Wiley New York, 1979.

[SR11] Saurav Sahay and Ashwin Ram. Socio-semantic health information access. In *AAAI Spring Symposium: AI and Health Communication*. AAAI, 2011.

[SS01] Jordi Sabater and Carles Sierra. Regret: A reputation model for gregarious societies. In *Fourth workshop on deception fraud and trust in agent societies*, volume 70, 2001.

[Sta10] Eugen Staab. Reliable information acquisition in the presence of malicious sources. Technical report, University of Luxembourg, 2010.

[Sti86] Stephen M. Stigler. *The history of statistics: The measurement of uncertainty before 1900*. Harvard University Press, 1986.

[SYHL05] Yan Sun, Wei Yu, Zhu Han, and K. J. Ray Liu. Trust modeling and evaluation in ad hoc networks. In *Proceedings of the 2005 Global Telecommunications Conference*, volume 3, pages 1862–1867. IEEE, 2005.

[TPJL06] W. T. Luke Teacy, Jigar Patel, Nicholas Jennings, and Michael Luck. Travos: Trust and reputation in the context of inaccurate information sources. *Autonomous Agents and Multi-Agent Systems*, 12:183–198, 2006.

[WLS12] Xiaofeng Wang, Ling Liu, and Jinshu Su. Rlm: A general model for trust representation and aggregation. *Services Computing, IEEE Transactions on*, 5(1):131–143, 2012.

[Zad65] Lotfi A. Zadeh. Fuzzy sets. *Information and control*, 8(3):338–353, 1965.

[ZLTV10] Theodore Zahariadis, Helen C. Leligou, Panagiotis Trakadas, and Stamatis Voliotis. Trust management in wireless sensor networks. *European Transactions on Telecommunications*, 21(4):386–395, 2010.

[ZMLM11] Dell Zhang, Robert Mao, Haitao Li, and Joanne Mao. How to count thumb-ups and thumb-downs: user-rating based ranking of items from an axiomatic perspective. In *Advances in Information Retrieval Theory*, pages 238–249. Springer, 2011.

Glossary

AND ($x \wedge y$) See Section 10.2. 31, 41, 61, 68, 144
An abstract operator used in the axiomatic approach. The operator is an abstraction of trust conjunction and multiplication.

asymmetric interaction See page 1. 1
An interaction where one party (the target) determines whether the outcome of the interaction is beneficial (a success) or costly (a failure) to the other party (the subject). The subject decides whether or not to have this interaction. Usually simply referred to as "interaction".

axiom See Definition 3.3. 3, 8, 29
A self-evident truth.

axiomatic approach See Part I. 3, 8, 29
An approach in which correctness trust models are studied by providing and analysing axioms. Refers to the methodology applied in Part I.

Bayes' theorem See Theorem 6.2. 3, 78
An mathematical theorem in probability theory. The theorem links the probability of a set of outcomes under a hypothesis to the probability of a hypothesis given a set of outcomes. Bayes' theorem is relevant, if we see a target having a certain integrity as a hypothesis.

belief triple ((b, d, u)) See page 31. 21, 31, 32, 45, 55
A trust opinion in Subjective Logic.

Bernoulli distribution See e.g. [Bil95, Gut07]. 69, 123
A probability distribution with one parameter p and two possible outcomes, one outcome happens with probability p, the other with probability $1 - p$.

beta distribution ($f_{\mathrm{B}}(x; s, f)$) See Definition 6.7. 7, 15, 19, 75, 78, 94, 102, 110, 121, 126
A probability distribution with two parameters, which represent the number of failures and successes. Such beta distributions range over the integrity of the target.

Beta family with trust chaining See Definition 8.2. 97, 100, 125
A family of correctness trust models that allow trust aggregation and trust chaining.

Beta model See Definition 6.8. 5, 19, 81, 128
A correctness trust model that allows trust aggregation.

Beta model with logical trust operations See Definition 7.1. 85, 88, 125
A correctness trust model that allows trust aggregation and the logical trust operations.

Beta paradigm See page 7. 2, 7, 13, 97
The paradigm in which this thesis is set. The main principles are that we are interested in formally correct trust opinions, that trust is (asymmetric) interaction oriented, that outcomes are objectively good or bad, and that information is limited to observing outcomes and recommendations.

Canephora See Section 8.2.1. 104, 105, 139
A tool that computes trust chains given predefined or user-defined entanglements and lying strategies.

carrier set See Definition 3.2. 16, 38
Given some (binary) operations $\diamond$, whenever $a \diamond b = c$ and a, b are in the carrier set, c is also in the carrier set. For example, integers and real numbers are a carrier set for addition and subtraction, but natural numbers are not.

chained trust opinion See Definition 8.3. 101
A trust opinion constructed using only a trust chain that incorporates exactly one recommender.

chained trust opinion See page 97. 81, 97, 112, 128
A trust opinion constructed using trust chaining.

cognitive trust model See page 3. 3, 14
A trust model that attempts to accurately capture the trust opinions people hold.

complement ($\ominus(x)$) See Section 2.4.2. 21, 32
A computationally defined operation from Subjective Logic, designed to implement trust negation.

composite trust opinion See page 85. 81, 85, 127, 137
A trust opinion constructed using logical trust operations.

comultiplication ($x \oslash y$) See Section 2.4.2. 21, 32
A computationally defined operation from Subjective Logic, designed to implement trust disjunction.

consensus ($x \oplus y$) See Section 2.4.2. 7, 21, 31
A computationally defined operation from Subjective Logic, designed to implement trust aggregation.

correctness trust model See page 3. 3, 14, 75
A trust model where trust opinions are mathematical descriptions of probabilities regarding trustworthiness of users (targets).

dilution ($x \cdot y$) See Section 10.2. 7, 31, 41, 68, 144
An abstract operator used in the axiomatic approach. The operator is an abstraction of trust chaining and discounting.

discounting ($x \otimes y$) See Section 2.4.2. 7, 21, 31
A computationally defined operation from Subjective Logic, designed to implement trust chaining.

dogmatic belief See Section 4.1. 41, 42, 58
A dogmatic opinion in Subjective Logic.

dogmatic opinion See Section 4.1. 10, 41, 45
A trust opinion in which there is no uncertainty. Dogmatic opinions can only be constructed from non-dogmatic opinions in limit cases.

endogenous filtering See page 19. 11, 19, 108, 133
Weighing or selecting recommendations (about a target) based on the content of the recommendation, comparing it to data the subject already has on the target. If a recommendation differs significantly from the data that the subject has, the subject rejects or assigns little weight to a recommendation. Contrast with exogenous filtering.

entanglement (λ) See page 80. 11, 80, 102, 125
A distribution over natural numbers, that determines the probability that a subject and target share n interactions.

exogenous filtering See page 19. 11, 19
Weighing or selecting recommendations (about a target) based on the recommender. If a recommender is not sufficiently trusted, the subject rejects or assigns little weight to a recommendation. Contrast with endogenous filtering.

experiment See Definition 3.1. 35, 36, 45
The symbolic representation of an interaction.

failure (F) See page 1. 1, 20, 36, 79
The outcome of an asymmetric interaction can be objectively determined to be a success or a failure. It is a failure if the outcome is costly to the subject.

fusion ($x + y$) See Section 10.2. 7, 30, 41, 68, 144
An abstract operator used in the axiomatic approach. The operator is an abstraction of trust aggregation and consensus.

game-theoretical trust model See page 14. 14
A correctness trust model which applies game-theory, emphasises interpersonal dynamics, and typically assumes rationality.

integrity (R_C) See page 80. 6, 20, 79, 113, 131
The integrity of a user is the probability that the succeeds in an arbitrary interaction. The integrity is unknown to other users, who can only estimate its value.

interaction history (O_C^A) See page 79. 79, 89, 101, 120, 128
A pair of natural numbers that represent the number of successes and the number of failures between the subject and the target.

inverse ($\overline{x}$) See Section 10.2. 31, 41, 68, 144
An abstract operator used in the axiomatic approach. The operator is an abstraction of trust negation and complement.

isomorphic See page 8. 8, 21, 35, 60, 75, 94, 110
Two mathematical structures are isomorphic if there is a mapping between them that respects the operations. Thus, two trust models are isomorphic, if they are indistinguishable with regard to trust aggregation, trust chaining and the logical trust operations.

logical trust operations See Chapter 7. 2, 7, 15, 29, 76, 85, 98, 125
The three logical trust operations are trust conjunction, trust disjunction and trust negation.

lying strategy (χ^B) See page 99. 11, 99, 102, 104, 116, 125
A distribution of recommendations from which a recommender selects recommendations when he lies. The lying strategy may depend on the recommender's integrity and the recommender's actual opinion of target.

multiplication ($x \otimes y$) See Section 2.4.2. 21, 32
A computationally defined operation from Subjective Logic, designed to implement trust conjunction.

OR ($x \vee y$) See Section 10.2. 31, 41, 61, 68, 144
An abstract operator used in the axiomatic approach. The operator is an abstraction of trust disjunction and comultiplication.

probabilistic approach See Part I. 3, 9
An approach in which correctness trust models are studied by providing a probabilistic semantics to trust related concepts. Refers to the methodology applied in Part II.

recommendation (S_C^B) See Section 2.3. 1, 18, 34, 47, 66, 81, 97, 113, 126
A claim towards the subject, by a recommender, concerning a target. An honest recommendation reflects the trust opinion of the recommender about the target.

recommender See page 97. 6, 18, 61, 97, 110, 113, 128
A user that provides a recommendation.

recommender system See page 17. 6, 18
A trust system where recommendations are determined by taste, and outcomes of interactions are subjective.

reputation system See page 17. 6, 18
A trust system where recommendations are determined by fact, and outcomes of interactions are objective.

simple trust opinion ($\vartheta_{s,f}$) See page 81. 81, 85, 108, 126, 137
A trust opinion constructed using only direct interactions and trust aggregation.

SLVisualiser See Section 3.2.2. 37
A tool that visualises trust computations in Subjective Logic in real-time.

subject See page 1. 1, 13, 31, 78, 127
The subject is a user that forms a trust opinion about another user (the target), to determine whether to have an asymmetric interaction with that target.

Subjective Logic See Section 3.2.1. 8, 19, 21, 31, 41, 65
An expressive trust model with foundations in the Beta model. It contains consensus, discounting, multiplication, comultiplication and complement, which are designed to capture trust aggregation, chaining, conjunction, disjunction and negation, respectively.

success (s) See page 1. 1, 20, 36, 79
The outcome of an asymmetric interaction can be objectively determined to be a success or a failure. It is a success if the outcome is beneficial to the subject.

target ($S \in \mathbf{T}$) See page 1. 1, 13, 31, 78, 127
The target determines the outcome of an asymmetric interaction. The target is assumed to succeed with a certain probability, called his integrity.

trust aggregation See Definition 6.9. 2, 7, 21, 24, 29, 41, 65, 76, 97, 125, 139
Trust aggregation is a binary operation over trust opinions. Two (independent) trust opinions regarding a single target are transformed into a single trust opinion regarding that target.

trust chaining See Definition 8.1. 2, 7, 19, 29, 41, 65, 76, 97, 139
Trust chaining is a binary operation from a trust opinion and a recommendation to a recommendation. A trust opinion regarding a recommender and his recommendation regarding a target are transformed into a single trust opinion regarding that target.

trust conjunction See Chapter 7. 2, 31, 89, 139
Trust conjunction is a binary operation over trust opinions. Two (independent) trust opinions regarding a two targets are transformed into a single trust opinion regarding an imaginary target that succeeds if and only if both original targets succeed.

trust disjunction See Chapter 7. 2, 31, 89, 139
Trust disjunction is a binary operation over trust opinions. Two (independent) trust opinions regarding a two targets are transformed into a single trust opinion regarding an imaginary target that succeeds if and only if either original targets succeed.

trust model See page 3. 3, 31, 125
Trust models are used to verify or analyse trust systems. Trust models often either model the trust people have under given circumstances (cognitive models), or the trust people should have under given circumstances (correctness models).

trust negation See Chapter 7. 2, 31, 89, 139
Trust negation is a unary operation over trust opinions. A trust opinion regarding a target is transformed into a trust opinion regarding an imaginary target that succeeds if and only if the original targets fails.

trust opinion See Section 2.2. 1, 7, 14, 16, 29, 58, 68, 80, 126, 135
An estimate of the likelihood that an interaction will be a success and reflects confidence of that estimate. The trust operations range over trust opinions.

trust system See page 3. 3, 140
Trust systems are (online) systems that assist in decision making where trust is involved. In particular recommendation systems and reputation systems are trust systems.

user ($A \in \mathbf{A}$) See page 1. 1, 18, 32, 42, 65, 79, 85, 97, 116, 126
Users are entities that may be subjects, targets or recommenders.

Index of subjects