The Fallacy of Endogenous Discounting of Trust Recommendations

Tim Muller, Yang Liu, Jie Zhang
School of Computer Engineering, Nanyang Technological University
50, Nanyang Avenue, Singapore
{tmuller,yangliu,zhangj}@ntu.edu.sg

ABSTRACT

Recommendations are widely used in recommender systems, reputation systems, and trust-based security systems. Some existing reputation systems and trust-based security systems use the flawed notion of endogenous discounting. Endogenous discounting is the idea that claims deviating from prior expectations should be ignored or discounted, which introduces confirmation bias. To show the fallacy of endogenous discounting, we construct a semantic meta-model that captures the key notions surrounding recommendations. We prove that any model covered by the meta-model can be formulated in a divide-and-conquer fashion. We show how divide-and-conquer clashes with endogenous discounting. Lastly, we discuss the implications on existing work that applies endogenous discounting, and provide suggestions for future work.

Categories and Subject Descriptors

I.2.11 [**Distributed Artificial Intelligence**]: Multiagent Systems

Keywords

Endogenous Discounting; Recommendation; Trust System; Rating

1. INTRODUCTION

Recommendations are an increasingly common source of information for decision making over the Internet. There are different models for reasoning with different forms of recommendations. Some models are designed for a particular goal, e.g., for recommending movies [13], books (amazon.com), music (last.fm) or security. Some models apply in a particular setting, e.g., for peer to peer networks [1] or mobile ad-hoc networks [15]. Other models are generic trust models, e.g., Subjective Logic [10], TRAVOS [27], or HMM-based models [5, 7]. Whether recommendations are deceitful is an important question in all. Recommendations that are deceitful must be filtered out (qualitatively or quantitatively), and the remainder must be combined into an informed decision.

There are two ways to filter recommendations: *exogenous discounting* and *endogenous discounting*, as named in [30]. In exogenous discounting, recommendations from unreliable sources are filtered out[1]. In endogenous discounting, recommendations are filtered out when they deviate from other recommendations and/or first-hand experience. Examples of endogenous discounting are: "This seller did not scam me, therefore I can ignore the warnings from others" or "9 people loved the movie, and 1 person hated it; that single person's opinion can be ignored". These two are clear examples of fallacious reasoning, due to confirmation bias. In this work, we prove the fallacy of endogenous discounting, in generality.

We did not find examples of endogenous discounting in recommender systems in the literature – presumably, there, examples are too obviously fallacious. We did, however, find a significant number of examples in reputation systems and trust-based security systems (discussed in Section 5.4). Since the fallacy occurs in reputation systems and trust-based security systems, our formalism more closely matches their paradigm. Nevertheless, the formalism is sufficiently general to cover all collusion-free systems with recommendations. The impact of collusion is discussed in Section 5.2.

Our main contributions are three-fold:

1. We introduce a generic semantics for dealing with opinions and recommendations. The semantics is generalised over the possible behaviours of the agents.
2. We prove viability of modular/incremental/algebraic approaches in systems with recommendations – approaches where users maintain opinions about agents, which are updated when new information arrives.
3. We show that endogenous discounting should not be applied to evaluate recommendations. We discuss the impact of this fallacy.

Results 1 and 2 corroborate results found in parts of the literature: Result 1 supports the approaches based on Bayesian methods [7, 26], based on HMM models [5, 28], and based on model-based collaborative filtering [25]. Result 2 indicates that decentralised and dynamic approaches to forming opinions, e.g., [10, 26], are viable. Result 3 has a direct impact to a significant fraction of research on trust models that apply endogenous discounting. Endogenous discounting is applied in (recent) work on: reputation systems [3, 30, 32], security protocols – like secure routing [14], mobile ad-hoc networks [9], access control [8], specific domains – like cloud computing [6], peer to peer networks [31] – and general analysis of recommendations [23].

Organization. In Section 2, we introduce a simple trust model with endogenous filtering, to exemplify the notions used in this paper, as well as exemplify how endogenous filtering can introduce error. To generalise that, we need to introduce a general notion that covers the relevant models. The meta-model is introduced in Section 3 for this purpose. Using the meta-model, we prove decompositionality in Section 4. In Section 5, we complete the argument against

[1]E.g. in collaborative filtering [2], unreliable sources are people with different tastes; e.g. with referral trust [11], unreliable sources are typically people with hidden agendas.

Appears in: *Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2015), Bordini, Elkind, Weiss, Yolum (eds.), May, 4–8, 2015, Istanbul, Turkey.*

endogenous discounting. We discuss the impact of decompositionality on endogenous discounting in Section 5.1, the relation with collusion in Section 5.2, why endogenous discounting is used despite being fallacious in Section 5.3, the impact of our results to existing work in Section 5.4 and the implications for future research in Section 5.5.

2. CONFIRMATION BIAS

The confirmation bias is the fallacy where an agent accepts or rejects evidence, not on the basis of the quality of the evidence, but on whether or not it fits in the agent's view. Weighting evidence against counter-evidence and deciding the former outweighs the latter does not suffer from confirmation bias. However, discarding the counter-evidence is an instance of confirmation bias. When the counter-evidence is not considered, the resulting opinion will be overconfident. Moreover, in the case that the opinion is wrong, correcting it is difficult. In this section, we show these characteristics of the confirmation bias using a simplified model. The general idea that endogenous filtering is fallacious is discussed in the remainder of the paper.

In this work, we assume that trust models contain *interactions* between *users* and *targets*. The users cannot determine the *outcome* of the interaction, therefore, they first form an *opinion* about the target. In order to form an opinion, users use past interactions, and users may consult with *recommenders*, who provide *recommendations*. Agents in the system may be users, recommenders, targets, or possibly fulfill multiple roles. The interactions are, typically, only visible to the user and the target involved. In this section, we set up a simple trust model that incorporates and links these notions. The notions are generalised in Section 3.

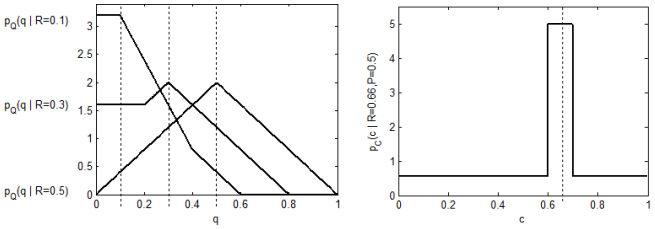
Ad-hoc Trust Model with Endogenous Discounting.

The *ad-hoc trust model* exemplifies a trust model, where we can exhibit our general results in a concrete manner. The details of the model are not important, but for concrete data, strong assumptions must be made. The assumptions are chosen for being both representative for reputation systems, yet easy to compute with.

The outcome of an interaction Q is denoted with a quality parameter $q \in (0, 1)$. A high *quality* interaction (e.g., stay at a nice hotel, or receive the timely delivery of intact goods) has a value q near 1. Targets with high *integrity* (e.g., good hotels or honest salesmen), denoted with R, tend to provide higher a quality interaction Q. We assume that given integrity R, the quality Q of an outcome is determined as $p_Q(q|R = r) = \max(2 - 4 \cdot |r - q|, 0) + \max(2 - 4 \cdot (r + q), 0) + \max(-6 + 4 \cdot (r + q), 0)$. We depict the distribution for $R = 0.1$, $R = 0.3$ and $R = 0.5$, in Figure 1(a). This particular distribution is chosen for its convenient computational properties, but the general argument holds for all distributions of Q given R, and even for subjective notions of quality. The crucial notion is that a target has a (random) integrity, which determines the distribution of the quality, where higher integrity targets tend to provide higher quality interactions.

In the ad-hoc trust model, we assert that a trustworthy recommender y provides a recommendation $C_y \in (0, 1)$ close to the quality parameter R of the target. We partition $(0, 1)$ into ten intervals of size 0.1, called bins. "Close to" simply means that C_y and R are in the same bin. We introduce the parameter P_y to capture the trustworthiness of recommender y; y is trustworthy with probability P_y. Both trustworthy and untrustworthy recommendations are uniformly distributed. The distribution of C_y is, therefore:

$$P_{C_y}(c_y|P_y{=}p_y, R{=}r) = \begin{cases} 10 \cdot p_y & \text{if } r \text{ and } c_y \text{ share a bin.} \\ \frac{1-p_y}{0.9} & \text{otherwise.} \end{cases}$$



(a) Distrib. $p_Q(q|R{=}0.1)$, $p_Q(q|R{=}0.3)$, $p_Q(q|R{=}0.5)$ (b) Dist. $p_{C_y}(c_y|R = 0.66)$

Figure 1: Distribution of the quality and the recommendations, given the integrity of the target.

The distribution of recommendations is depicted in Figure 1(b), with $P_y = 0.5$ and $R = 0.66$.

Remark 1. For simplicity, we let all integrity values be equiprobable a priori, i.e., $p_R(r)$ is uniformly distributed. Using Bayes' theorem, with $p_R(r) = 1$, we obtain $p_R(r|Q{=}q) = p_Q(q|R{=}r)$. Similarly, $p_R(r|C_y{=}c_y, P_y{=}p_y) = p_{C_y}(c_y|R{=}r, P_y{=}p_y)$ These equations do not hold in full generality, but are a direct consequence of the convenient choices of our distributions.

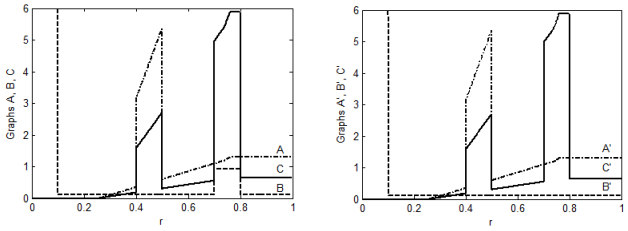
A user may form an opinion about a target x, based on an interaction with quality Q, and some recommendations C_{y_0}, C_{y_1}. We denote this as the opinion $\tau_x(Q{=}q, C_{y_0}{=}c_{y_0}, C_{y_1}{=}c_{y_1})$. This opinion may not correspond to single values, it simply succinctly denotes the perspective of a user in a system. For such an opinion about a target to be relevant, it must have implications on the integrity R of the target.

We implement endogenous discounting in the ad-hoc trust model using statistical hypothesis testing. The null hypothesis is that recommenders are trustworthy – the recommendation and the target's integrity share a bin. The alternative hypothesis is its negation. We let $\alpha = 0.05$, meaning that we reject recommendations c_y, if the probability of getting c_y from a trustworthy recommender y is smaller than 5%, given a distribution over R. Thus, given distribution f over R, we reject when the bin of c_y has probability mass under 0.05, or formally: $\int_{\frac{\lceil 10c \rceil}{10}}^{\frac{\lfloor 10c \rfloor}{10}} f(r)\mathbf{d}r \leq \alpha$. This function resembles the form of endogenous filtering used in TRAVOS [27]. Recommendations are provided to the system in chronological order, and hypothesis testing is performed in that order.

Experimental Analysis of the Ad-hoc Trust Model.

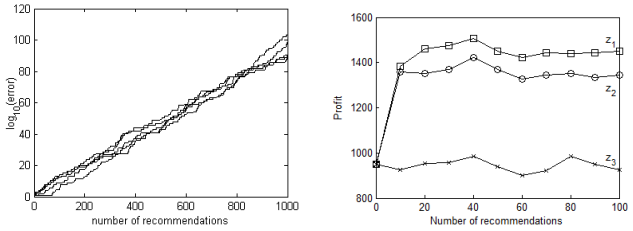
The purpose of the experimental analysis, is to provide a concrete interpretation of the general results that we prove in the paper. Therefore, we simplify the experimental analysis, to expose the core issues: decompositionality and the confirmation bias. In particular, this means that we discretise the random variables, and that we set all $P_{y_i} = 0.5$ as public knowledge. Setting P_{y_i} as a fixed value precludes exogenous discounting, since all recommenders are equally helpful, thus exposing endogenous discounting.

A target has integrity R with probability $p_R(r)$. In the ad-hoc trust model, only the integrity determines the quality of the interactions, so the integrity of the target is of great interest to the user. By obtaining evidence (recommendations or interactions), the probabilities change, since generally $p_R(r) \neq p_R(r|E)$, where E is some evidence. Decompositionality is the notion that we can derive $p_R(r|E_1, E_2)$ from $p_R(r|E_1)$ and $p_R(r|E_2)$. The computation that allows us to derive the former is called aggregation (Definition 3). In the ad-hoc trust model, aggregation is simply normalised multiplication. That means that $p_R(r|E_1, E_2)$ is proportional to $p_R(r|E_1) \cdot p_R(r|E_2)$.



(a) Graphs A, B, C (b) Graphs A', B' C'

Figure 2: Probability distributions of the integrity of the target, given different evidence.



(a) Error after n rec. (b) Profit of z_1, z_2 and z_3

Figure 3: Simulations demonstrating that endogenous discounting introduces an error, and decreases profit.

To show concretely what decompositionality means, consider the following simulation: We have two bodies of evidence, one body consists of an interaction with quality $q = 0.76$ and $C_{y_1} = 0.43$, thus $\tau_x(Q{=}0.76, C_{y_0}{=}0.43)$, and one body is recommendations $C_{y_1} = 0.04$, $C_{y_2} = 0.07$ and $C_{y_3} = 0.72$, thus $\tau_x(C_{y_1}{=}0.04, C_{y_2}{=}0.07, C_{y_3}{=}0.72)$. We run a Monte Carlo simulation, generating R, Q and $C_{y_0}, \ldots C_{y_3}$. In variation A, we drop all samples except those where $\tau_x(q{=}0.76, C_{y_0}{=}0.43)$ – thus where there is one interaction with quality 0.76 and only recommender y_0 provides a recommendation, namely 0.43. In variation B, we drop all samples except those where $\tau_x(C_{y_1}{=}0.04, C_{y_2}{=}0.07, C_{y_3}{=}0.72)$. In variation C both must hold, thus samples are kept only if $\tau_x(Q{=}0.76, C_{y_0}{=}0.43, C_{y_1}{=}0.04, C_{y_2}{=}0.07, C_{y_3}{=}0.72)$. Figure 2(a) shows the probability of the integrity, in A, B and C. Note that graph B exceeds the dimensions of the figure. Figure 2(a) furthermore reveals that C is proportional to the pointwise product of A and B, decompositionality holds. Thus, in reality, there exists a link between A&B and C. There is no reason why a trust model should not break this link.

Endogenous discounting, as proposed in the ad-hoc trust model, does break the link between A&B and C. We run the same Monte Carlo simulation as before, with one difference: If we are not confident in the trustworthiness of the recommender, then we reject the recommendation (thus ignore the effect of the recommendation on the samples). Figure 2(b) shows A', B', C', which correspond to A, B, C with endogenous discounting. Here B' is cut off too. In particular, observe that in graph B', the dashed block spanning from 0.7 to 0.8 disappeared, compared to B. This is because after receiving recommendations $C_{y_1} = 0.04$ and $C_{y_2} = 0.07$, the recommendation $C_{y_3} = 0.72$ is rejected (confidence 0.99). However, in C', $C_{y_1} = 0.04$ and $C_{y_2} = 0.07$ are rejected, and as a consequence $C_{y_3} = 0.72$ is accepted. It is immediate that A'&B' are no longer linked with C'. The ad-hoc trust model with endogenous discounting failed to uphold decompositionality.

The two preceding simulations make the main argument of this paper concrete: decompositionality is a natural property, but endogenous discounting breaks decompositionality. The next two simulations demonstrate that endogenous discounting can be harmful. Notice that the difference between B and B', is that B' puts more probability density on what it believes to be the likely integrity value, $c < 0.1$. This is caused by the confirmation bias, as B' rejects evidence ($C_{y_3}{=}0.72$) due to the fact that the evidence does not match expectations. First, we show that arbitrarily large errors can be introduced by the confirmation bias with a bounded probability. Second, we show that endogenous discounting actually leads to impaired decision making.

We take a simplistic view on error, namely the inverse of the probability density assigned to the integrity of the target, $\frac{1}{p_R(r|\ldots)}$. A distribution with high density on r gives error close to 0, and a distribution with low density on r gives increasingly high error. We initialise every simulation by selecting random r, and two recommendations C_{y_0}, C_{y_1}, such that C_{y_0} and C_{y_1} are in the same bin, but in a different bin from R. This happens with probability $\frac{(1-P_{y_0})\cdot(1-P_{y_1})}{9} = \frac{1}{36}$. The remainder of the simulation uses n randomly generated recommendations (using $p_{C_{y_i}}(c_{y_i}|R = r, P_{y_i} = p_{y_i})$). Figure 3(a) show the error as n increases. The reason the error increases rapidly, is because of an unfortunate false start, whereafter the truth is rejected for being unlikely, and further misleading evidence may be accepted. The error grows exponentially in with respect to n, and the probability of the false start remains a constant $\frac{1}{36}$.

To show that endogenous discounting negatively impacts decision making, consider a game between three users, z_1, z_2 and z_3. Users z_1, z_2 and z_3 are offered the same interactions, with a certain payoff in case of success, and cost in case of failure. If they believe the interaction is profitable, they interact, otherwise they do not. User z_1 does not apply endogenous discounting, z_2 does, and z_3 is the baseline, as he ignores all recommendations. We perform a Monte Carlo simulation of $10,000$ runs, in which we generate an integrity, and k recommendations according to the assumptions (e.g. a random $\tau_t(Q{=}q, C_{y_1}{=}c_{y_1}, \ldots, C_{y_k}{=}c_{y_k})$) and payoff and cost of interaction uniformly in $[0, 1]$ and $[-1, 0]$. The users z_1 and z_2 typically make the same decision, for $k = 10$, users z_1 and z_2 make different decisions in less than 5% - 10% of the potential interactions. This is despite the fact that user z_2 falls for the confirmation bias in 15% - 20% of the cases. Specifically, in 15% - 20% of the cases, the user z_2 rejects recommendations that were in the correct bin.

We see, in Figure 3(b), that if $k = 0$, the profit is identical for all users, as no user actually uses any recommendations. User z_3 ignores all further recommendations, meaning that any variation in the graph is pure chance. We see that z_1 and z_2 increase and rapidly stabilise, with z_1 stabilising consistently higher than z_2. The stabilisation is because the additional impact of extra recommendations quickly decreases. The absolute difference between z_1 and z_2 is fairly small. However, relative to the baseline, z_3, we see a difference that roughly corresponds with the 15% of the cases in which z_1 falls for the confirmation bias. This implies that falling for the confirmation bias has a real and detrimental effect on trust-based decision making. It is important to note that fallaciously using recommendations (z_2) is significantly better than not using recommendations at all (z_3).

Observe that the unbounded error from Figure 3(a) has a bounded effect on the profit, in Figure 3(b). The reason is that the maximal size of a mistake is bounded, and even an exponentially large error cannot increase the size of the mistake.

When we introduce the formalism in full generality, we refer to this concrete example. Whenever we refer to the running example, we refer to the ad-hoc trust model, and specifically to the experiments from Figure 2(a).

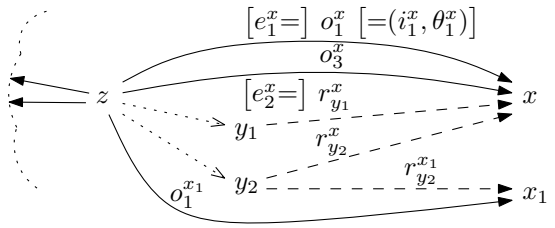


Figure 4: Example of a network from user z's perspective

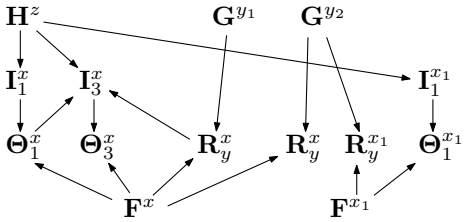


Figure 5: A Bayesian network depicting the relationships.

3. META-MODEL

In this section, we introduce the meta model, which aims to capture the key semantics of the recommendation related components and concepts. As it is a meta-model, we leave some details unspecified and use them to reason about abstract entities and relationships. Consequently, our meta-model requires only the bare assumptions, found often in reputation systems and trust-based security systems. Concretely, we reason about strategies and probabilities, without assigning strategies to agents, nor selecting prior distributions.

3.1 Setting

The main assumption of our meta-model, is that models deal with *opinions* based on evidence. Specifically, $\tau_x(e_1, \ldots, e_n)$ is the opinion about x based on evidence $e_1, \ldots, e_n$. These opinions have a semantics, which:

R1 makes predictions about the agent (or product) that the opinion applies to.

R2 can be updated when additional evidence arises.

A significant number of reputation systems, recommender systems and trust management systems aim to satisfy R1 and R2. Not all models, however, satisfy R1 and R2, e.g., models of human cognition may purposely implement the confirmation bias or other fallacies. In this section, we provide a meta-model that provides semantics that satisfy both R1 and R2.

Before introducing the meta model, we first formally define the notations for the concepts introduced in the previous section. Let Z be a set of users, Y be the set of recommenders and X be the set of targets. Interactions are a pair consisting of an initiation $i_j^x \in I$ and an outcome $\theta_j^x \in \Theta$ of one interaction (e.g., purchasing a laptop is an interaction, consisting of an initiation in the form of a prepayment of the price of a laptop, and an outcome in the form of a timely delivery of the offered laptop). Throughout the paper, we adopt the viewpoint of a user – therefore, it suffices to identify an interaction (or initiation or outcome) by the target x and an index j. We may also omit the target or the index, if they are not relevant.

We use O to denote the set of observations of the interactions between the users and targets $o_j^x \in O$ is short for the interaction between user j and target x. (One observation is about exactly one interaction.) We use R to denote the set of recommendations of recommenders provide. $r_y^x \in R$ represents the recommendation from recommender y about target x. Observations and recommendations collectedly form the evidence E, i.e., $E = O \cup R$. Users have opinions $\tau_x(e_1, e_2, \ldots)$ about x based on *evidence*, $e_j \in E$.

Figure 4 depicts an example of a network with user z, two recommenders y_1 and y_2, and two targets x and x_1. In the network the user has two pieces of direct evidence about x: $e_1^x{=}o_1^x{=}(i_1^x, \theta_1^x)$ and $e_3^x{=}o_3^x{=}(i_3^x, \theta_3^x)$ – we omit the equivalent notations of o_3^x in the figure, to avoid cluttering. The user, further, has two recommendations about x, from agents y_1 and y_2, namely $e_2^x{=}r_{y_1}^x$ and $e_4^x{=}r_{y_2}^x$. In order to establish the reliability of the recommenders, the user may look at other recommendations by the recommenders. In the example, recommender y_2 also makes a recommendation about x_1, $e_2^{x_1}{=}r_{y_2}^{x_1}$, and the user has direct evidence about x_1, namely $e_1^{x_1}{=}o_1^{x_1}$.

In our meta-model, we are not necessarily interested in the behaviour of the agents. We are interested in which factors may influence the decisions of the agents. Events unobserved by an agent, for example, cannot influence the behaviour of that agent. Here, we could use Bayesian network to visualise the influence. A Bayesian network is a graphical notation to show the conditional independence between random variables. Intuitively, an arrow can be interpreted as meaning that the source influences the sink.

Figure 5 represents the relations of the random variables corresponding to the entities in Figure 4. We use random variables $\mathbf{E}_i^x$, $\mathbf{O}_i^x$, $\mathbf{R}_y^x$, $\mathbf{I}_i^x$ and $\mathbf{\Theta}_i^x$ for evidence, observations, recommendations, initiations and outcomes, respectively. There are 5 additional random variables, $\mathbf{H}^z$, $\mathbf{G}^{y_1}$, $\mathbf{G}^{y_2}$, $\mathbf{F}^x$ and $\mathbf{F}^{x_1}$, representing the behaviour of z, y_1, y_2, x and x_1, that will be discussed later. This Bayesian network represents the essentials of a typical trust system, with two caveats: there is no collusion and no recommendations about recommendations. The relevance of these caveats is discussed after the formalism and results.

The meta-model presented in this work encodes the relations between the entities, and which events are visible to whom. If an agent performs an action under some circumstances, then the behaviour of that agent is more likely to be the sort of behaviour in which he performs that action under those circumstances with high probability – this is Bayes' theorem. However, the circumstances are typically only partially known, and subtle hidden connections may exist. Notions like cost, benefit, taste and discrimination must be gracefully handled by the meta-model, without explicitly modeling these. As a consequence, the precise formulation of the meta-model contains a fair amount of technicalities.

3.2 Formal Meta-model

In the meta-model, users, recommenders and targets have reactive strategies. A user strategy takes into account a collection of evidence, in order to decide whether to interact with a target. A recommender strategy takes into account the target of the recommendation, when recommending. A target strategy takes into account the interaction at hand, in order to decide the outcome of the interaction. Thus, a reactive strategy allows (the probability of) the outcome to depend, e.g., on the value of the interaction (as in the value imbalance attack [12]).

Remark 2. The reactive strategy is stateless, it cannot (naively) model, e.g., performing ten successes and only then a failure (as in the playbook attack [12]). Decompositionality also holds for a meta-model allowing the more technically involved stateful reactive strategies (similar to [16]). See our technical report [17].

We use some shorthand notations. A list $\langle a_1, \ldots, a_n \rangle$ may be shorthanded to $\bar{a}$. We define $\triangleleft(\Phi)$ to be the set of distributions

over the arbitrary set Φ. If Φ is discrete, $\triangleleft(\Phi) = (\Phi \to [0,1])$, where for all $f : \triangleleft(\Phi)$, $\sum_{\varphi\in\Phi} f(\varphi) = 1$, and if Φ is continuous, $\triangleleft(\Phi) = (\Phi \to \mathbb{R}^{\geq 0})$, where $\int_\Phi f(\varphi)\mathbf{d}\varphi = 1$. We use P and p for probability mass and probability density functions. For random variable $\mathbf{A}$ and outcome a, rather than writing $P(\mathbf{A}{=}a)$ or $p_{\mathbf{A}}(a)$, we write $P(a)$ or $p(a)$, whenever this introduces no ambiguity.

Rather than explicitly assigning strategies to agents, we merely define the shape of the strategies. A target finalises an interaction, based on the initiation of that interaction. A recommender provides recommendations based on the target that the recommendation is about (be it truthful or not). A user initiates interactions with targets, based on the evidence he has about the target. From game theory, we learn that agents may want to perform actions with probabilities between 0 and 1; so-called mixed moves. In our definition, we assert that a strategy is simply a function from the input, to a distribution over the output:

Definition 1. The *strategy of a target* is a function $f : I \to \triangleleft(\Theta)$. The *strategy of a recommender* is a function $g : F \to \triangleleft(R)$. The *strategy of a user* is a function $h : E \times \cdots \times E \to \triangleleft(I)$.

In the technical report [17], the strategy is extended to a form probabilistic automata, similar to hidden Markov models.

Running Example 1. In the ad-hoc trust model, it is assumed that every target x has an internal parameter r_x that determines the quality of its interactions. The outcomes are denoted with a quality, between 0 and 1, thus $\Theta = (0,1)$. The probability of the outcome being θ, is equal to $\min(\frac{2\theta}{r}, \frac{2(1-\theta)}{1-r})$. Thus, we can identify each r_x with exactly one strategy f^x_r, such that $f^x_r(i)(\theta) = \min(\frac{2\theta}{r}, \frac{2(1-\theta)}{1-r})$. Similarly, we can identify every p_y with exactly one strategy g^y_p, such that

$$g^y_p(r)(c) = \begin{cases} 10 \cdot p & \text{if } r \text{ and } c \text{ share a bin.} \\ \frac{1-p}{0.9} & \text{otherwise.} \end{cases}$$

In some systems, particularly reputation systems, recommenders cannot observe the target's strategy, but only perform interactions and observe these. Thus, we may propose a function from interactions to recommendations, rather than from a strategy to recommendations: $(\hat{g}^y \in \hat{G}) : \overline{O} \to \triangleleft(R)$. Each target strategy f^x (in conjunction with the recommender strategy) yields a distribution over the interactions ($P(\overline{o^x}|\hat{g}^y, f^x)$). Therefore, given the target's strategy f^x, for every $\hat{g}^y$, there exists an identical g^y, namely: $g^y(f^x)(r^x_y) = \sum_{\overline{o^x}\in\overline{O}} P(\overline{o^x}|\hat{g}^y, f^x) \cdot \hat{g}^y(\overline{o^x})(r^x_y)$. Hence, as long as the user has no direct information about the interactions between the recommender and the target, every $\hat{g} \in \hat{G}$ can be substituted by some $g \in G$. Thus, without loss of generality we can assert the shape of the input of the recommender strategy to be the as in Definition 1.

Running Example 2. We can alter the ad-hoc trust system, such that the recommender interacts with the target, and receives an interaction with quality q' based on r. The recommender then bases his recommendation on q' (rather than r), as many reputation systems assert. Specifically, the trustworthy recommenders provide a recommendation c in the same bin as q' (rather than r). Thus, as before $\hat{g}^y_p(q')(r) = \begin{cases} 10 \cdot p & \text{if } q' \text{ and } c \text{ share a bin.} \\ \frac{1-p}{0.9} & \text{otherwise.} \end{cases}$ Now, since the link between r and q' is known to be $P(q'|r)$ and q' is not visible to any of the users: $g^y_p(r)(r) = \int_0^1 P(q'|r) \cdot \hat{g}^y_p(q')(r)\mathbf{d}q'$.

We use the random variables $\mathbf{F}^x$, $\mathbf{G}^y$ and $\mathbf{H}^z$, for the strategies of the targets x, recommenders y and users z. As we assume agents do not collude, we assume mutual independence between $\overline{\mathbf{F}}, \overline{\mathbf{G}}, \overline{\mathbf{H}}$; any collection of strategies is independent of any collection of different strategies.

By the definition of the strategy, $\mathbf{I}^x = h(\overline{e^x})$, $\mathbf{\Theta}^x = f^x(i^x)$, and $\mathbf{R}^x_y = g^y(f^x)$. The strategy completely determines the actions of the agents. That means, in particular, that the probability that an agent performs an action a, under the condition that his strategy is φ and the input is b, equals $\varphi(b)(a)$ – as long as a does not appear in the condition. Formally, $P(a|\varphi, b, \overline{c}) = \varphi(b)(a)$, if $\overline{c}$ does not contain a. (Note that variables may be nested, e.g. $c_i = \psi(a)$.)

Based on these random variables, we can introduce the semantics of an opinion, called a valuation:

Definition 2. The *valuation* $[\![\tau_x(e_1, \ldots, e_n)]\!]$ is the semantics of the opinion $\tau_x(e_1, \ldots, e_n)$. The valuation $[\![\tau_x(e_1, \ldots, e_n)]\!]$ is the distribution $p(f^x|e_1, \ldots, e_n)$.

In the technical report [17], valuations are extended to stateful valuations, which are distributions over states that the target may be in.

Running Example 3. Opinion $\tau_x(Q{=}0.76, C_{y_0}{=}0.43)$ was introduced in Section 2. The opinion $\tau_x(Q{=}0.76, C_{y_0}{=}0.43)$ occurs when the user observes $Q = 0.76$ and recommendations $C_{y_0} = 0.43$ about the target x. Graph A, in Figure 2(a) depicts the true distribution of the integrity parameter, given $Q = 0.76$ and $C_{y_0} = 0.43$, namely $p_R(r|Q{=}0.76, C_{y_0}{=}0.43)$. Definition 2 states that graph A – $p_R(r|Q{=}0.76, C_{y_0}{=}0.43)$ – is the semantics of $\tau_x(Q{=}0.76, C_{y_0}{=}0.43)$.

We introduce aggregation to denote the required operation for updating opinions:

Definition 3. The *aggregation* of two opinions $\tau_x(\overline{e}), \tau_x(\overline{e'})$ is denoted $\tau_x(\overline{e}) \bowtie \tau_x(\overline{e'})$ and equals $\tau_x(\overline{e}, \overline{e'})$.

Definitions 1 and 2 are sufficient to prove that for direct observations, valuations are updateable:

PROPOSITION 1. *For arbitrary two observations $\overline{o^x}$ and $\overline{o'^x}$,* $[\![\tau_x(\overline{o^x}) \bowtie \tau_x(\overline{o'^x})]\!] \propto \frac{[\![\tau_x(\overline{o^x})]\!] \cdot [\![\tau_x(\overline{o'^x})]\!]}{[\![\tau_x(\emptyset)]\!]}$.

PROOF. We first apply induction over n to prove

$$p(f^x|o^x_1, \ldots, o^x_n) = p(f^x) \cdot \prod_{j=1}^n f^x(i^x_j)(\theta^x_j). \tag{1}$$

The base case trivially holds. The induction step is as follows:

$$\begin{aligned}
&p(f^x|o^x_1, \ldots, o^x_n) \\
&\propto P(\theta^x_n, i^x_n|f^x, o^x_1, \ldots, o^x_{n-1}) \cdot p(f^x|o^x_1, \ldots, o^x_{n-1}) \\
&= P(\theta^x_n|i^x_n, f^x, o^x_1, \ldots, o^x_{n-1}) \cdot P(i^x_n|f^x, o^x_1, \ldots, o^x_{n-1}) \\
&\qquad \cdot p(f^x|o^x_1, \ldots, o^x_{n-1}) \\
&= f^x(i^x_n)(\theta^x_n) \cdot h^z(o^x_1, \ldots, o^x_{n-1})(i^x_n) \cdot p(f^x|o^x_1, \ldots, o^x_{n-1}) \\
&\propto f^x(i^x_n)(\theta^x_n) \cdot p(f^x|o^x_1, \ldots, o^x_{n-1})
\end{aligned}$$

From equation (1), the proposition follows as:

$$\begin{aligned}
&[\![\tau_x(o^x_1, \ldots, o^x_k) \bowtie \tau_x(o^x_{k+1}, \ldots, o^x_n)]\!] \\
&= [\![\tau_x(o^x_1, \ldots, o^x_n)]\!] \\
&\propto p(f^x) \cdot \prod_{j=1}^n f^x(i^x_j)(\theta^x_j) \\
&= \frac{p(f^x) \cdot \prod_{j=1}^k f^x(i^x_j)(\theta^x_j) \cdot p(f^x) \cdot \prod_{j=k+1}^n f^x(i^x_j)(\theta^x_j)}{p(f^x)} \\
&= \frac{[\![\tau_x(o^x_1, \ldots, o^x_k)]\!] \cdot [\![\tau_x(o^x_{k+1}, \ldots, o^x_n)]\!]}{[\![\tau_x(\emptyset)]\!]} \qquad \square
\end{aligned}$$

This proposition is proven for stateful valuations in the technical report [17].

Proposition 1 proves that aggregation's semantics are the pointwise multiplication of the valuations, corrected for prior distributions. Bayesian updates are pointwise multiplications. Hence, the pointwise multiplication of two valuations contains both their Bayesian updates. However, the prior distribution is contained twice – it is pointwise squared. Hence, we need to divide by the prior, $p(f^x)$.

Running Example 4. Aggregation is the link between A&B and C (see Figure 2(a)). It states that $\tau_x(Q{=}0.76, C_{y_0}{=}0.43)$ aggregated with $\tau_x(C_{y_1}{=}0.04, C_{y_2}{=}0.07, C_{y_3}{=}0.72)$ is $\tau_x(Q{=}0.76, C_{y_0}{=}0.43, C_{y_1}{=}0.04, C_{y_2}{=}0.07, C_{y_3}{=}0.72)$. Proposition 1 is the formalisation of the notion that the link between A&B and C can be computed using only the values of $p_R(r|Q{=}0.76, C_{y_0}{=}0.43)$, $p_R(r|C_{y_1}{=}0.04, C_{y_2}{=}0.07, C_{y_3}{=}0.72)$ and $p_R(r|Q{=}0.76, C_{y_0}{=}0.43, C_{y_1}{=}0.04, C_{y_2}{=}0.07, C_{y_3}{=}0.72)$. The computation is a normalised pointwise multiplication, which is independent from the evidence. The same does not hold for A'&B' and C'.

The meta-model ensures that the semantics of every model it covers satisfies R1 and R2. Users can predict (R1) the probability that a target performs action θ upon our initiation i, when our opinion is $\tau_x(e_1, \ldots, e_n)$. First take the semantics: $[\![\tau_x(e_1, \ldots, e_n)]\!] = p(f^x|e_1, \ldots, e_n)$, then take the expected value of the probability that θ happens: $\int_0^1 p(f^x|e_1, \ldots, e_n) \cdot f^x(i)(\theta)\mathbf{d}f^x$. The latter formula provides the exact probability that θ happens, whenever the assumptions about the prior distribution are correct. For observations, the opinions can be updated in a straightforward manner, as Proposition 1 proves. In Section 4, we generalise Proposition 1 to Lemma 1, and ultimately Theorem 1, and show that R2 holds in full generality. Not only can the trust opinions be updated without reevaluating the evidence, the procedure for updating is aggregation as defined in Definition 3.

4. DECOMPOSITIONALITY

A scenario where an opinion is computed based on recommendations (and possibly direct observations) is called a *chain*. In a *basic chain*, all evidence the user has about the target, is one single recommendation. Since a basic chain is a chain, any system that computes chains can compute the basic chain. Endogenous discounting cannot be applied to the basic chain, since there is nothing to compare the recommendation to. We prove that any chain can be decomposed into basic chains with aggregation. Neither basic chains nor aggregation need endogenous discounting. Therefore, any chain can be computed (via decomposition) without endogenous discounting.

Decompositionality has been proven in a special case – the Beta model – in [18]. Our Lemma 1 and Theorem 1 are proper generalisations of the Modularity Proposition and Modularity Theorem.

Proposition 1 can be generalised to prove that the user's direct observations can be isolated from the recommendations. We refer to the ability to separate direct observations and recommendations as *weak decompositionality*.

LEMMA 1. *Let $\overline{e^{\overline{x}}}$ be evidence about targets $x_1, \ldots, x_\ell \neq x$, $\overline{o^x}$ be observations about the target x, and $r^x_{\overline{y}}$ be recommendations about x from recommenders $\overline{y} = y_1, \ldots, y_k$. Then:*
$[\![\tau_x(\overline{o^x}, r^x_{\overline{y}}, \overline{e^{\overline{x}}})]\!] = [\![\tau_x(\overline{o^x}, \overline{e^{\overline{x}}}) \bowtie \tau_x(r^x_{\overline{y}}, \overline{e^{\overline{x}}})]\!]$

PROOF. The induction in the proof of Proposition 1 can be straightforwardly adapted to prove:

$$p(f^x|o^x_1, \ldots, o^x_n, r^x_{\overline{y}}, \overline{e^{\overline{x}}}) = p(f^x|r^x_{\overline{y}}, \overline{e^{\overline{x}}}) \cdot \prod_{j=1}^n f^x(i^x_j)(\theta^x_j)$$

Note that $p(f^x|\overline{r^x_{\overline{y}}}, \overline{e^{\overline{x}}}) = p(f^x|\overline{r^x_{\overline{y}}})$, as follows:

$$\begin{aligned}
&p(f^x|\overline{r^x_{\overline{y}}}, \overline{e^{\overline{x}}}) \quad \propto \quad P(\overline{e^{\overline{x}}}, \overline{r^x_{\overline{y}}}|f^x) \cdot p(f^x) \\
&= \int_{\overline{F}}\int_{\overline{G}} P(\overline{e^{\overline{x}}}, \overline{r^x_{\overline{y}}}|f^x, \overline{f^{x'}}, \overline{g^y}) \cdot p(\overline{f^{x'}}, \overline{g^y}|\overline{f^x}) \mathbf{d}\overline{f^{x'}}\mathbf{d}\overline{g^y} \cdot p(f^x) \\
&= \int_{\overline{F}}\int_{\overline{G}} \left(\prod_i \left(\prod_j f^{x'_i}(i^{x'_i}_j)(\theta^{x'_i}_j)\right) \cdot \left(\prod_j g^{y_j}(f^{x'_i})(r^{x'_i}_j)\right)\right) \\
&\qquad \cdot \left(\prod_j f^x(i^x_j)(\theta^x_j)\right) \cdot \; p(\overline{f^{x'}}, \overline{g^y}) \mathbf{d}\overline{f^{x'}}\mathbf{d}\overline{g^y} \cdot p(f^x) \\
&\propto \left(\prod_j f^x(i^x_j)(\theta^x_j)\right) \cdot p(f^x) \quad \propto \quad p(f^x|\overline{r^x_{\overline{y}}})
\end{aligned}$$

Using equation (1), $\prod_{j=1}^n f^x(i^x_j)(\theta^x_j) = \frac{p(f^x|\overline{r^x_{\overline{y}}})}{p(f^x)}$, and thus $\prod_{j=1}^n f^x(i^x_j)(\theta^x_j) = \frac{p(f^x|\overline{r^x_{\overline{y}}}, \overline{e^{\overline{x}}})}{p(f^x)}$, proving the lemma. □

Weak decompositionality, Lemma 1, states that user's own observations can be ignored without loss of generality, with respect to analysing recommendations. Decompositionality requires that recommendations can furthermore be treated in isolation from each other.

THEOREM 1. *Let $\overline{e^{\overline{x}}}$ be evidence about targets $x_1, \ldots, x_\ell \neq x$, and $e^x_1, \ldots, e^x_n$ be evidence about target x. Then:*
$[\![\tau_x(e^x_1, \ldots, e^x_n, \overline{e^{\overline{x}}})]\!] = [\![\tau_x(e^x_1, \overline{e^{\overline{x}}}) \bowtie \ldots \bowtie \tau_x(e^x_n, \overline{e^{\overline{x}}})]\!]$

PROOF. Via Lemma 1, we assume w.l.o.g. $e^x_j = r^x_{y_j}$. Thus, it suffices to prove that $p(f^x|r^x_{y_1}, \ldots, r^x_{y_k}, \overline{e^{\overline{x}}}) = p(f^x|r^x_{y_1}, \overline{e^{\overline{x}}}) \bowtie \ldots \bowtie p(f^x|r^x_{y_k}, \overline{e^{\overline{x}}})$.

Suppose $\frac{p(f^x|e^{\overline{x}})}{p(f^x)} \propto 1$ and $p(\overline{g^y}|f^x, \overline{e^{\overline{x}}}) = p(\overline{g^y}|\overline{e^{\overline{x}}})$, then:

$$\begin{aligned}
&p(f^x|r^x_{y_1}, \ldots r^x_{y_n}, \overline{e^{\overline{x}}}) \\
&\propto P(r^x_{y_1}, \ldots r^x_{y_n}|f^x, \overline{e^{\overline{x}}}) \cdot p(f^x|\overline{e^{\overline{x}}}) \\
&= \int_{\overline{G}} P(r^x_{y_1}, \ldots, r^x_{y_n}|f^x, \overline{g^y}, \overline{e^{\overline{x}}}) \cdot p(\overline{g^y}|f^x, \overline{e^{\overline{x}}}) \mathbf{d}\overline{g^y} \cdot p(f^x|\overline{e^{\overline{x}}}) \\
&\propto \int_{\overline{G}} \left(\prod_j g^{y_j}(f^x)(r^x_{y_j})\right) \cdot p(g^{y_1}, \ldots, g^{y_k}|\overline{e^{\overline{x}}}) \cdot p(f^x|\overline{e^{\overline{x}}}) \mathbf{d}\overline{g^y} \\
&\propto \frac{\prod_j p(f^x|r^x_{y_j}, \overline{e^{\overline{x}}})}{p(f^x)^{n-1}}
\end{aligned}$$

In Lemma 1, we find $p(f^x|\overline{r^x_{\overline{y}}}, \overline{e^{\overline{x}}}) = p(f^x|\overline{r^x_{\overline{y}}})$. Letting $e^{\overline{x}}$ be empty, this is our first assertion. The second assertion holds too:

$$\begin{aligned}
&p(\overline{g^y}|f^x, \overline{e^{\overline{x}}}) \\
&\propto p(\overline{e^{\overline{x}}}|f^x, \overline{g^y}) \cdot p(\overline{g^y}) \\
&= \int_{\overline{F}} p(\overline{e^{\overline{x}}}|f^x, \overline{g^y}, \overline{f^{x'}}) \cdot p(\overline{f^{x'}}|f^x, \overline{g^y}) \mathbf{d}\overline{f^{x'}} \cdot p(\overline{g^y}) \\
&= \int_{\overline{F}} p(\overline{e^{\overline{x}}}|\overline{g^y}, \overline{f^{x'}}) \cdot p(\overline{f^{x'}}|\overline{g^y}) \mathbf{d}\overline{f^{x'}} \cdot p(\overline{g^y}) \\
&\propto p(\overline{g^y}|\overline{e^{\overline{x}}}) \quad \square
\end{aligned}$$

5. FALLACY OF ENDOGENOUS DISCOUNTING

We have introduced a meta-model, and proven a divide-and-conquer approach to evaluating recommendations. In this section, we show that endogenous discounting is indeed fallacious, then relate these theoretical results to existing work and extract useful data for future work.

5.1 Endogenous Discounting is Fallacious

To show that endogenous discounting is fallacious, we need to show that:

H1 Our meta-models cover the relevant models.

H2 Decompositionality is not an artifact from our choice of formalism.

H3 Decompositionality precludes endogenous discounting.

Our meta-models are a semantic model. Thus, for claim H1, it suffices to argue that the meaning of an opinion in a trust model[2] coincides with our semantics. Some trust models explicitly reason about strategies and/or probability, and are trivially covered by our meta-models. Examples under the meta model are Subjective Logic [10], TRAVOS [27], HABIT [26], BLADE [21], PRep [7], and recommender systems using model-based collaborative filtering [25] or memory-based collaborative filtering [33], whereas HMM models [5, 28] fall under the stateful extension of the meta-model.

Typically, probabilistic trust and reputation systems ascribe a fixed quality or integrity to agents, and construct a distribution over that parameter. Take the set of strategies f_p, for $p \in [0, 1]$, such that $f_p(i)(\theta{=}1) = p$, for arbitrary initiations and $\Theta = \{0, 1\}$. The strategy f_p corresponds to the behaviour of an agent with integrity p; in both interpretations, the agent succeeds with probability p. BLADE and HABIT require more parameters, but are still captured by the meta-model. The HMM models are instances of the stateful meta-model, where the distribution over the initial parameterisations can be transformed to the distribution over the root nodes in the valuation.

In the probabilistic recommender systems, users, recommenders and targets have hidden profiles, features or classifications. The outcomes of interactions are assumed to be determined by these profiles, in a relatively straightforward process. The strategy merely emulates this process. (Thus, our strategies may simply be a modelling tool, as, e.g., movies cannot be ascribed to a strategy.)

For the remaining (non-probabilistic) models, as long as they aim to be predictive (R1) and updatable (R2), our results are useful. The model aims to be able to combine opinions based on different recommendations (R2). Since the result should be accurate (R1), its semantics must resemble the semantics of the aggregated opinions. As the latter can be computed without endogenous discounting, the model need not apply endogenous discounting.

Concerning claim H2, note that any (meta) model, introduces some relations and assumptions. The implications of our meta-model are only as good as its assumptions. The notion that agents have strategies is straightforward, as is the notion that these strategies can only use facts that are observed by the agent. An apparent weakness is our assumption that agents operate independently (non-collusion); perhaps decompositionality is a trivial consequence of this. However, all that we assume is that an agent cannot access another agent's private information, except through the system. Thus, two agents can cooperate (benevolently or maliciously) in our meta-model, as long as they only coordinate using the system. In reality, collusion – users coordinating outside of the system – does occur. It is important to notice that although decompositionality breaks, weak decompositionality (Lemma 1) remains true. We discuss collusion in more detail in Section 5.2. Thus, we see no reason to believe that decompositionality is an artificial result of our formalism, affirming claim H2.

[2]Meaning the model used in reputation systems, in recommender systems or in trust management systems.

Per decompositionality, pointwise multiplication forms the provably correct semantics of aggregation (Theorem 1). Pointwise multiplication does not actually evaluate the evidence. Thus, if two opinions are aggregated, and the individual opinions are free of endogenous discounting, then the result is also free of endogenous discounting. The opinion based on an individual piece of evidence, particularly a single recommendation, can trivially not apply endogenous discounting. Therefore, decompositionality precludes endogenous discounting, affirming claim H3.

5.2 Collusion

Agents may collude for all kinds of reasons. Usually agents attempt to increase their gains. A recommender does not need to be part of a coaction to provide informative recommendations. Thus, when recommenders collude, they typically mislead other agents[3]. Moreover, it is reasonable to assume that colluding agents are sufficiently smart to avoid lying in a way that is detrimental to their goals.

The proof of Theorem 1 obviously fails in the presence of colluding agents, since $p(\overline{f}, \overline{g}|f^x) \neq p(\overline{f}, \overline{g})$, invalidating decompositionality. However, since Lemma 1 still holds, weak decompositionality holds even under collusion. Since weak decompositionality holds even in systems where collusion occurs, endogenous discounting based on the user's direct observations remains fallacious regardless of collusion. Without loss of generality, we consider only endogenous discounting where recommendations are compared to each other.

A naive endogenous discounting approach weights recommendations with how common they are. A rational coalition lets all its members provide similar recommendations, to increase their weight. Naive endogenous discounting, therefore, becomes less effective as the size of the coalition increases – and we have proven it fallacious for coalitions of size one.

More sophisticated notions of endogenous discounting have to take into account possible strategies of the target and of the coalition (members). If the target strategy is not taken into account, the target may discriminate a group of agents to make it appear they are colluding. If the coalition strategy is not taken into account, the coalition may exploit the inner workings of the endogenous discounting mechanism. Therefore, we need to model the probabilities of strategies of agents and coalitions, given the user's observations. Thus, simply comparing the contents of the recommendations is insufficient.

The argument does not constitute a proof that endogenous discounting is fallacious, in systems with collusion. However, the argument is sufficiently powerful to indicate that endogenous discounting should be avoided, even in systems with collusion. Collaborative filtering algorithms and clustering algorithms have proven to be effective tools to find coalitions – see, e.g., [20]. We, therefore, advise using these algorithms, and to avoid endogenous discounting.

5.3 Intuition behind Endogenous Discounting

The reason why endogenous discounting is occasionally applied, is the intuition that deviant recommendations are probably lies. Our results do not disprove this intuition. The effect that deviant recommendations are likely lies exists. Our results, however, do imply that we do not need to filter this effect.

Intuitively, there is little information conveyed by a fake recommendation r. According to information theory, the valuation of

[3]In atypical situations, coalitions lie to assist the user [4]. In our domain, however, it is assumed that helpful agents simply provide truthful recommendations.

a known fake recommendation is intuitively close to the uniform distribution, call it p_u. The valuation of a known true recommendation intuitively matches the valuation based on the evidence from the claim, call it p_r. Let q be the probability that the recommendation is true. Then the probability that the strategy of the target equals f is $q \cdot p_r(f) + (1-q) \cdot p_u(f)$.

Now, let p_o be a valuation, such that r is a conflicting recommendation. In other words, p_o and p_r do not agree on which strategies are probable; for arbitrary f, $p_r(f) \approx 0$ or $p_o(f) \approx 0$. (Since, if neither $p_r(f)$ nor $p_o(f)$ is close to 0, then they agree that f is probable.) Aggregating p_o and p_r, we get $p_t(f) \propto p_o(f) \cdot (q \cdot p_r(f) + (1-q) \cdot p_u(f))$. By distributivity, we obtain two summands, $q \cdot p_r(f) \cdot p_o(f)$ and $(1-q) \cdot p_u(f) \cdot p_o(f)$. Since, for all f, $p_r(f) \approx 0$ or $p_o(f) \approx 0$, the first summand is close to zero. Thus $p_t(f) \approx c \cdot (1-q) \cdot p_u(f) \cdot p_o(f)$, and as p_u is close to uniform, $p_t(f) \approx p_o(f)$. In conclusion, when the conflict is sufficiently high, the aggregated opinion p_t tends towards the original opinion p_o, meaning that the impact of the conflicting recommendation diminishes.

Nowhere did we need to apply endogenous discounting, to obtain the result that conflicting recommendations should have a reduced impact. That is how it is possible that on one hand, the intuition behind endogenous discounting is valid, while on the other hand, endogenous discounting is fallacious.

5.4 Implications on Existing Work

Endogenous discounting is applied in classical trust models [27, 30, 32], recent trust models [22, 23] and in domain specific work [6, 8, 9, 14, 31], but – fortunately – not typically applied in recommender systems. The fallacy can enter a model in subtle ways, particularly if the primary objective of the model is to find the reliability of the recommender (rather than the target), as in [22, 29]. In fact, the fallacy may lay outside of the work in question, but arise when the work is naively applied, as possible with [19, 24].

Endogenous discounting is first defined in the Beta Reputation System [30]. The BRS shares many of the assumptions of the stateless meta-model. However, being one of the first Beta models, the authors do not yet attempt to capture recommendations using Bayesian methods. Rather, their chained opinions are based on intuition, and, fallaciously implement the intuition described in the previous section. The notion of endogenous discounting, as a heuristic, was still an improvement over no filtering. Thus, endogenous discounting gained some traction.

Prob-Cog is an example of a system with a wider scope than just evaluating and combining recommendations. Its primary goal is to encode human dispositions into a probabilistic setting [19]. The authors introduce a notion of credibility for recommenders to deal with unfair ratings. Credibility is computed using a variation of collaborative filtering [2], with modifications to allow tendencies such as optimism. Naively, the most precise value for credibility would use all recommendations. Our results help avoid this pitfall, and, thus, assist in applying systems like Prob-Cog correctly.

HABIT [26] is a probabilistic model that uses Bayesian inference. HABIT, like many Bayesian and beta models, does not support endogenous discounting. The reason is that such models are a direct sub-model of our meta-models. For HABIT, all assumptions of the stateless meta-model, are assumptions of HABIT. Therefore, Theorem 1 can be directly translated into HABIT's formalism.

In conclusion, our results prove that work based on endogenous discounting is flawed. Using our results, one can avoid incorrectly interpreting work that focusses on specialised issues. We corroborate existing work that does not apply endogenous discounting, by showing that it can be correct without endogenous discounting.

5.5 Recommendations for Future Work

The obvious implication of our work, is that future work should avoid endogenous discounting solutions. As we saw in the previous section, some heuristics can be improved with endogenous discounting (i.e., BRS with endogenous discounting). Therefore, we do not oppose future work that uses endogenous discounting per se. However, we advise authors to establish that endogenous discounting is the better alternative in their heuristic. It is insufficient to show that endogenous discounting provides better results than no filtering.

A major implication of our work, is the proof that for non-colluding recommenders, their recommendations can be evaluated in isolation. This means that divide-and-conquer approaches, such as Subjective Logic [10], are viable approaches. Divide-and-conquer methods are efficient, adaptive and easy to parallelise or decentralise. We advise researchers, therefore, to use decompositionality in their favour.

Bayesian and other probabilistic models are powerful, well-studied and general. In particular, recent work allows agents with increasingly rich behaviour; dynamic [5, 7], adaptive [21] or interactive [26]. Strategies straightforwardly model dynamic, adaptive and interactive behaviour. Therefore, we believe that strong theoretical and practical results can be obtained using our semantics.

6. CONCLUSION

When there is no collusion, it is fallacious to discount deviating recommendations, despite the fact that they more probably are lies (e.g. despite the majority opinion being more likely true). The reason is that, given a correct aggregation procedure, all recommendations can be treated in isolation and then aggregated, without post-hoc correction for lies. Filtering out deviating recommendations may, therefore, lead to confirmation bias.

When there is collusion, recommendations cannot be treated in isolation. We argue that this is a limitation in our approach to disproving endogenous discounting, rather than a scenario where endogenous discounting may work. Naive endogenous discounting is fallacious even under collusion.

An important necessity in proving our result is a sufficiently general meta-model. The meta-model is kept abstract, and assumes the bare minimum. What it does assume is in-line with modern probabilistic models. It assumes that agents have a local view that determines their actions, and that a user tries to determine the probability of agents' strategies. We considered both stateless and perfect-recall strategies.

We have formal proofs that recommendations can be treated in isolation and aggregated naively. We refer to this result as decompositionality. Decompositionality indicates that models that apply divide-and-conquer techniques are feasible. Therefore, our work impacts current work that applies endogenous discounting; current work that uses divide-and-conquer methods; and Bayesian or probabilistic models.

Acknowledgements

This research is supported (in part) by the National Research Foundation, Prime Minister's Office, Singapore under its National Cybersecurity R&D Program (Award No. NRF2014NCR-NCR001-30) and administered by the National Cybersecurity R&D Directorate. This research is also partially supported by "Formal Verification on Cloud" project under Grant No: M4081155.020, and the ASTAR/I2R - SERC, Public Sector Research Funding (PSF) Singapore (M4070212.020) awarded to Dr. Jie Zhang.

7. REFERENCES

[1] Karl Aberer and Zoran Despotovic. Managing trust in a peer-2-peer information system. In *International conference on Information and knowledge management*, pages 310–317. ACM, 2001.

[2] 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 Publishers Inc., 1998.

[3] Chrysanthos Dellarocas. Immunizing online reputation reporting systems against unfair ratings and discriminatory behavior. In *ACM conference on Electronic commerce*, pages 150–157. ACM, 2000.

[4] Naipeng Dong, Hugo Jonker, and Jun Pang. Enforcing privacy in the presence of others: Notions, formalisations and relations. In *European Symposium on Research in Computer Security (ESORICS)*, pages 499–516, 2013.

[5] Ehab ElSalamouny, Vladimiro Sassone, and Mogens Nielsen. Hmm-based trust model. In *International Workshop on Formal Aspects in Security and Trust (FAST2009)*, volume 5983, pages 21–35. LNCS, Springer, July 2009.

[6] Wenjuan Fan and Harry Perros. A reliability-based trust management mechanism for cloud services. In *Trust, Security and Privacy in Computing and Communications (TrustCom)*, pages 1581–1586. IEEE, 2013.

[7] Yasaman Haghpanah and Marie Desjardins. Prep: a probabilistic reputation model for biased societies. In *International Conference on Autonomous Agents and Multiagent Systems*, volume 1, pages 315–322. International Foundation for Autonomous Agents and Multiagent Systems, 2012.

[8] Naima Iltaf, Abdul Ghafoor, and Uzman Zia. A mechanism for detecting dishonest recommendation in indirect trust computation. *EURASIP Journal on Wireless Communications and Networking*, 2013(1), 2013.

[9] Mr Rahul A Jichkar and MB Chandak. An implementation on detection of trusted service provider in mobile ad-hoc networks. *International Journal of Engineering Trends and Technology (IJETT)*, 11(2):64–74, 2014.

[10] Audun Jøsang. A logic for uncertain probabilities. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 9(03):279–311, 2001.

[11] 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, Inc., 2005.

[12] Reid Kerr and Robin Cohen. Smart cheaters do prosper: defeating trust and reputation systems. In *Autonomous Agents and Multiagent Systems*, volume 2, pages 993–1000. International Foundation for Autonomous Agents and Multiagent Systems, 2009.

[13] Joseph A Konstan, Bradley N Miller, David Maltz, Jonathan L Herlocker, Lee R Gordon, and John Riedl. Grouplens: applying collaborative filtering to usenet news. *Communications of the ACM*, 40(3):77–87, 1997.

[14] Mohamed MEA Mahmoud and Xuemin Sherman Shen. Secure routing protocols. In *Security for Multi-hop Wireless Networks*, chapter 4, pages 63–93. Springer, 2014.

[15] Sergio Marti, Thomas J Giuli, Kevin Lai, and Mary Baker. Mitigating routing misbehavior in mobile ad hoc networks. In *International conference on Mobile computing and networking*, pages 255–265. ACM, 2000.

[16] Tim Muller, Yang Liu, Sjouke Mauw, and Jie Zhang. On robustness of trust systems. In *Trust Management VIII*, pages 44–60. Springer, 2014.

[17] Tim Muller, Yang Liu, and Jie Zhang. The fallacy of endogenous discounting of trust recommendations with dynamic agents. Technical report, Nanyang Technological University: http://pat.sce.ntu.edu.sg/tim/papers/fallacytechreport.pdf, 2015.

[18] Tim Muller and Patrick Schweitzer. On beta models with trust chains. In *Trust Management VII*, pages 49–65. Springer, 2013.

[19] Zeinab Noorian, Stephen Marsh, and Michael Fleming. Prob-cog: An adaptive filtering model for trust evaluation. In *Trust Management V*, pages 206–222. Springer, 2011.

[20] Grzegorz Orynczak and Zbigniew Kotulski. On a mechanism of detection of coalitions for reputation systems in p2p networks. In *Performance Evaluation of Computer and Telecommunication Systems (SPECTS 2014), International Symposium on*, pages 578–584. IEEE, 2014.

[21] Kevin Regan, Pascal Poupart, and Robin Cohen. Bayesian reputation modeling in e-marketplaces sensitive to subjectivity, deception and change. In *National Conference on Artificial Intelligence*, volume 21, page 1206. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2006.

[22] Raha Sadeghi and Mohammad Abdollahi Azgomi. A method for fair propagation of user perceptions for trust management in composite services. *Service Oriented Computing and Applications*, pages 1–20, 2014.

[23] Murat Şensoy, Geeth de Mel, Lance Kaplan, Tien Pham, and Timothy J Norman. Tribe: Trust revision for information based on evidence. In *Information Fusion (FUSION)*, pages 914–921. IEEE, 2013.

[24] Murat Şensoy, Jie Zhang, Pinar Yolum, and Robin Cohen. Poyraz: Context-aware service selection under deception. *Computational Intelligence*, 25(4):335–366, 2009.

[25] Xiaoyuan Su and Taghi M Khoshgoftaar. Collaborative filtering for multi-class data using belief nets algorithms. In *Tools with Artificial Intelligence (ICTAI)*, pages 497–504. IEEE, 2006.

[26] WT Teacy, Michael Luck, Alex Rogers, and Nicholas R Jennings. An efficient and versatile approach to trust and reputation using hierarchical bayesian modelling. *Artificial Intelligence*, 193:149–185, 2012.

[27] WT Luke Teacy, Jigar Patel, Nicholas R Jennings, and Michael Luck. Travos: Trust and reputation in the context of inaccurate information sources. *Autonomous Agents and Multi-Agent Systems*, 12(2):183–198, 2006.

[28] George Vogiatzis, Ian MacGillivray, and Maria Chli. A probabilistic model for trust and reputation. In *International Conference on Autonomous Agents and Multiagent Systems*, volume 1, pages 225–232. International Foundation for Autonomous Agents and Multiagent Systems, 2010.

[29] Yonghong Wang, Chung-Wei Hang, and Munindar P Singh. A probabilistic approach for maintaining trust based on evidence. *Journal of Artificial Intelligence Research*, 40(1):221–267, 2011.

[30] Andrew Whitby, Audun Jøsang, and Jadwiga Indulska. Filtering out unfair ratings in bayesian reputation systems. In *Proc. 7th Int. Workshop on Trust in Agent Societies*, volume 6, 2004.

[31] Qingui Xu and Guixiong Liu. Weakening unreliable ratings in p2p reputation systems based on an honest majority. *Advances in information Sciences and Service Sciences*, 4:418–425, 2012.

[32] Bin Yu and Munindar P Singh. Detecting deception in reputation management. In *International joint conference on Autonomous agents and multiagent systems*, pages 73–80. ACM, 2003.

[33] Kai Yu, Anton Schwaighofer, Volker Tresp, Xiaowei Xu, and H-P Kriegel. Probabilistic memory-based collaborative filtering. *Knowledge and Data Engineering, IEEE Transactions on*, 16(1):56–69, 2004.