ABSTRACT

HAZARD, CHRISTOPHER J. Trust and Reputation in Multiagent Systems: Strategies and Dynamics with Reference to Electronic Commerce. (Under the direction of Professor Munindar P. Singh.)

In multiagent interactions, such as e-commerce and peer-to-peer file sharing, being able to accurately assess the trustworthiness of other agents is important for agents to protect themselves from losing utility. We focus on an agent's *discount factor* (time preference of utility) as a direct measure of the agent's trustworthiness in a number of settings. We prove that an agent's discount factor, when in context of the agent's valuations and capabilities, is isomorphic to its trustworthiness for a set of reasonably general assumptions and definitions. Further, we propose a general list of desiderata for trust systems and show how discount factors as trustworthiness meet these desiderata. We also show how discount factors are a robust measure of trustworthiness when entering commitments with adverse selection and moral hazards.

When agents can significantly increase each other's utility at a moderate cost, the socially optimal outcome is for the agents to provide favors to each other. However, when agents cannot support or enforce a market system, the favor environment forms a situation similar to the repeated prisoner's dilemma because each agent can unilaterally improve its utility by refusing to help others. We present an adaptive tit-for-tat strategy that provides a mutually beneficial equilibrium when agents may have differing private discount factors and when favor costs and benefits are stochastic and asymmetric. This strategy enables agents to treat previously unencountered agents with caution, communicate about the trustworthiness of other agents, and evaluate past communication for deception. We discuss the details of our simulation results and the impact of various parameterizations and communication.

Building from the favor model, we examine more complex transactions with private discount factors as a model for trustworthiness. We closely examine the case of simultaneous favors, which comprise a single market transaction where two parties perform an exchange. Further, we investigate more complex market models, where agents directly compete on price and quality. We derive a number of methods that agents can use to obtain and aggregate information of other agents' discount factors and valuations.

Despite the large body of work in reputation and trust in dynamic multiagent environments, no metrics exist to directly and quantitatively evaluate and compare reputation systems. We present a common conceptual interface for reputation systems and a set of four *measurable* desiderata, inspired by dynamical systems theory, that are broadly applicable across multiple domains. We discuss the implications, strengths, and limitations of our desiderata. Our

discount factor as trustworthiness model performs well across the desiderata when measured against other established reputation models from the literature. We apply our desiderata to empirically evaluate the Amazon reputation mechanism in terms of actual ratings data obtained by sellers on Amazon's marketplace.

Trust and Reputation in Multiagent Systems: Strategies and Dynamics
with Reference to Electronic Commerce

by
Christopher J. Hazard

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Computer Science

Raleigh, North Carolina

2010

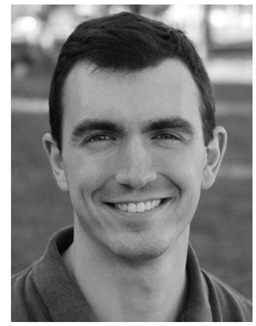
APPROVED BY:

Dr. Jon Doyle	Dr. Ting Yu
Dr. R. Michael Young	Dr. Munindar P. Singh Chair of Advisory Committee

DEDICATION

To those who have yet to be named, for soon it will be their turn to be trusted.

BIOGRAPHY



Christopher J. Hazard was born in 1981 in Wisconsin, growing up in a rural surrounding of forests, farms, and hills. He skipped the seventh grade, being the first student to ever skip a grade in the school district. In high school, he became the network administrator of one of the largest secondary education computer networks in the state of Wisconsin. Due to the lack of easily available learning resources, he became self-taught in computer science, physics, and mathematics by borrowing books from interlibrary loan and browsing the newly created World Wide Web.

In 1998, Chris began attending Valparaiso University in Indiana, skipping over many of the first two years of introductory courses. He majored in computer science and minored in mathematics, physics, and computer engineering. In December of 2001, he graduated with departmental honors, having built a distributed programming language and platform as his honors thesis. He also participated in numerous undergraduate research projects throughout his studies, including topics as diverse as quantum computing, cellular automata, and procedural music generation.

After an internship at Motorola as the lead developer of an existing 4-person team creating a CDMA infrastructure simulation and testing platform, he resumed his tenure upon graduating from Valparaiso. He was quickly promoted to a software architect role, and directed development teams in Illinois, Arizona, and Poland.

Chris enrolled in North Carolina State University in the fall of 2004. He participated in computer architecture research involving multicore processor caching algorithms, and began working with Dr. Peter Wurman on using microeconomic principles and game theory for robot coordination in warehousing environments with high combinatorial complexity. Chris interned at Dr. Wurman's warehouse robotics company, Kiva Systems, performing research on algorithm performance.

In 2007, Dr. Wurman left academia to work on his company full time, so Chris began working with Dr. Munindar Singh on a new dissertation on trust and reputation in multiagent systems and e-commerce, applying game theory and dynamical systems theory. Chris also took over the role of teaching upper-level undergraduate e-commerce classes. He received an award for his outstanding teaching. He also taught game programming to gifted and talented high school students through the TIP Summer Studies program.

Chris is also the founder and president of Hazardous Software, which makes multitemporal simulation software for training and entertainment. This technology is being showcased in their upcoming PC game Achron. He began work on the technology while an undergraduate student. He unveiled it to the public in 2009, receiving significant praise from game industry media and garnering interest from large technology firms and the US Army.

Chris and his wife Susan are both avid hikers and international travelers. Chris has visited approximately 20 countries on 6 continents.

ACKNOWLEDGEMENTS

First I would like to thank my advisor, Munindar Singh. Our advisor-advisee relationship is a classic example of two differing schools of thought combining to produce ideas and results that neither of us would have envisioned without the other. Munindar's positive attitude magnifies the successes and makes the tough times manageable. His keen understanding of what is important and what is not has been invaluable in directing the maturation of my research for both hard and soft skills.

I also would like to thank the other members of my committee: Jon Doyle, for the numerous discussions that were always insightful; Ting Yu, for practical feedback regarding trust and security; and Michael Young, for support of my work on Achron outside of the scope of my dissertation.

This dissertation was partly sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-09-2-0053. The views and conclusions contained in this document are those of the author and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

I would like to thank James Lester for all of his encouragement through my years at NCSU, David Thuente for all of his insightful advice, and Yan Solihin for all the engaging discussions at the beginning of my graduate studies. I would also like to thank Peter Wurman for making me realize how interesting, complex, and applicable e-commerce is to many other fields. I would also like to thank all of my fellow students in the labs who have shared ideas, listened to my practice talks, and proofread my papers.

While at Valparaiso University, David Johnson, James Caristi, and Doug Tougaw were instrumental in whetting my appetite for research and increasing my desire to pursue graduate school.

I would like to thank Mike Resnick for feedback on my many random ideas and for his support. I would also like to thank Brent Shumard for bouncing around many ideas and always being a sounding board for any obscure math problem I came across. Thanks to Phil and Kristin Barletta for their hospitality in coming to North Carolina.

My experience at NCSU wouldn't have been nearly as much fun without all the friendships I have made throughout my studies. I want to thank James Niehaus, Prachi Gauriar, and Paul Breimyer for laughing (or not laughing) at my dumb jokes and for all of the adventures getting lost in Raleigh and elsewhere. I especially want to thank Travis Breaux for his friendship and the many interesting and insightful conversations we have shared.

I would like to thank Mom for her never-ending support. I would also like to thank Dad for helping me become who I am today, even though he was not alive to see me complete this dissertation.

Finally, I would like to thank my wife Susan for her love and immense support throughout the marathon of graduate school. I look forward to meeting our unborn son.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

Trust is an important aspect of interaction, from a child to a parent, from a buyer to a seller, from one government to another. According to the Merriam-Webster dictionary, the term dates back to the 13th century from Old English, stemming from a word that meant faithful and is the root of the word true.

Economists have long examined the definition of trust, as markets depend on it to function. In his survey, James [2002] consolidates many of the questions economists have debated about the term, finding that definitions of trust are often paradoxical. In order for A to trust B, A must believe that B will not exploit A, that is, B should not be incentivized to negatively impact A's total expected utility for B's own gain. However, if B does not have an incentive to exploit A, then A does not need to trust B because B will behave in a desirable manner.

Within computer science, trust has also been given many different meanings. In their survey, Jøsang et al. [2007] break trust into two main definitions: whether an agent is reliable and whether an agent will exploit another, echoing the sentiments from economics. The level of trust falls on a continuous spectrum as well, whether the trust is based on reputation, such as whether a seller has offered quality products, or on policy, such as whether a specific agent is who it claims to be [Artz and Gil, 2007]. Policy-based trust mechanisms are typically employed in the disciplines of computer security, whereas softer security mechanisms use reputation measurements. Softer security systems involving trust and reputation are dealt with more frequently in multiagent systems, artificial intelligence, and e-commerce literature.

1.1 Trust in Multiagent Systems and E-commerce

Many kinds of business transactions can now be automated or semi-automated. Automation is being used for transactions in procurement, low-touch sales, and stock and commodity trading. As additional types of business-to-business transactions become automated [Subramaniam and Shaw, 2002] and autonomous agents become more crucial components of business, determining effective strategies for the agents becomes increasingly important in the study of these multiagent systems.

One primary question in dealing with such interactions is how much one agent should trust another. Doing business with an agent that has a reputation for being trustworthy generally has the benefit of reducing the risk of a poor outcome. Agents' reputations and perceived trustworthiness can significantly affect the demand and price an agent will receive. This holds in domains such as online auctions [Houser and Wooders, 2006] and supply chains [Sen and Banerjee, 2006]. Autonomous agents and humans often have differing trust models [Lewandowsky et al., 2000], and we focus primarily on autonomous agents.

Ideally, from a trust perspective, the mechanisms under which agents interact would be incentive compatible (IC), meaning agents' optimal strategies would be to be honest and truthful. Whereas IC mechanisms can be designed for a variety of interaction models [Jurca and Faltings, 2007], often maximizing profit for the agent running the mechanism is a higher priority. For example, eBay's (`http://ebay.com`) reputation system exhibits a bias toward transaction volume to maximize profit [Rubin et al., 2005], because sellers can game the reputation system [Khopkar et al., 2005]. Additionally, implementing an IC mechanism can be infeasible in certain settings in terms of computation or communication [Conitzer and Sandholm, 2004].

Trustworthiness reflects the worthiness of a *trustee* to aid or protect a *trustor*. For example, a trustworthy trustee will properly fulfill some task for a trustor or refrain from inappropriately revealing a secret. As trustor a learns more about trustee b, the amount of trust that a places in b should ideally approach the amount of trust of which b is truly worthy.

An agent's reputation is the aggregation of publicly available information about the agent. Such information is not necessarily accurate. Trust and reputation are often used in a complementary fashion: an agent expects positive outcomes when interacting with another agent that has a reputation for being trustworthy. Some systems are best described as trust systems because therein agents determine whether another agent will do what it says it will, whereas others are best described as reputation systems because therein agents determine and propagate their beliefs about other agents. The mechanics of the two kinds of systems exhibit considerable overlap [Ramchurn et al., 2004].

The need for trust systems arises in two situations: *adverse selection* and *moral hazard* [Dellarocas, 2005]. Adverse selection occurs with *typed* agents, meaning an agent is predisposed to some course of action due to its one or more (fairly constant) attributes. In particular, adverse selection refers to the case when agents have information asymmetries about others' intrinsic types such that an agent or item can masquerade as having a desirable attribute when it actually does not. An agent's type can range from a strict behavior regimen, such as accepting every offer or always producing high-quality items or being patient, to a parameter the agent uses in evaluating its utility, such as its willingness-to-pay for some item. The presence of typed agents means that agents may be able to improve their utility by determining which agents are of what type, and interacting only with agents of a favorable type. An example of a typed agent would be an agent selling faulty electronics at high prices. The agent may be unable to change the quality or price of the goods it sells, and other agents may do best to avoid purchasing from this agent. The ability of an agent to improve its utility by choosing with whom to interact is strongly affected by the interaction mechanism. An example of where an agent may not be able to choose the agents with which it will interact is an auction setting with perfectly substitutable goods where buyers and sellers are randomly matched by the auctioneer at a set price.

Conversely, in a setting where agents choose trading partners, if agent a manufactures poorer quality items than the other agents, knowing that a manufactures poorer quality items can enable some other agent, b, to increase its own utility by not purchasing from a. Determining trust with adverse selection can be framed as a multiagent learning problem, as the agents signal each other to increase the accuracy of their beliefs of other agents' types.

Moral hazards are created when agents do not bear the full cost of their actions and are thus incentivized to perform actions that may harm the utility of others. For example, a seller who deals with a gullible buyer has the moral hazard of falsely advertising its goods. To address moral hazards, trust systems attach sanctions to unwanted behavior. If agent a performs some unwanted behavior, then a trust system can attach some information to a. This information can be used by a centralized mechanism or individual agents to sanction or avoid interacting with a, with the effect that a would have an incentive to alter its behavior. An example of moral hazard is an agent manufacturing an item of low quality but advertising it as high quality, where an unsuspecting buyer would overpay the seller for the low quality item.

1.1.1 Trust Surveys

Ramchurn, Huynh, and Jennings [2004] separate trust into two main categories: system-level trust and individual-level trust. They define system-level trust as a mechanism that forces agents to be trustworthy. System-level trust encompasses incentive compatible models, where agents are incentivized to be trustworthy, and also security models, where typically an agent uses cryptographically secure mechanisms to authenticate itself, access certain information, or perform certain actions. Individual-level trust is where an agent has beliefs and acts upon those beliefs. Ramchurn et al. separate individual-level trust into models that gather ratings, aggregate ratings, and promote ratings, and also mention sociocognitive agents (e.g., belief-desire-intention agents).

The theme of dividing trust into its moral hazard and adverse selection components is the primary categorization by Dellarocas [2006]. Like Ramchurn et al., Dellarocas divides reputation systems into those that are incentive compatible (elicit honest feedback), and those that do not. Dellarocas also looks at reputation as a dynamic property, and breaks reputation phases into the initial, steady-state, and end-game, and discusses decentralization of reputation systems. He examines various problems of reputation systems, such as Sybil attacks, strategic manipulation, and how statistically, some agents will have a bad reputation simply because of noise in observations.

Artz and Gil [2007] group related work into four categories. The first is policy-based trust, which encompasses credentials and building on trusted authorities. The second is reputation-based trust, where agents use past interactions to predict future actions. Artz and Gil group the remaining work into what they call general models of trust, which include game theory as well as affective models, and trust in information resources, to denote the work of applied trust systems, particularly with the web, the semantic web, and collaborative filtering. The dimensions Artz and Gil use to classify trust systems are the targets of trust assessment, the representation of trustworthiness, methods of modeling the trust, management and aggregation of the data, computation aspects, and the purpose of the trust system.

The dimensions that Mui, Halberstadt, and Mohtashemi [2002] describe are more of a classification system than a means of clustering related work. Their dimensions are the breadth of simultaneous contexts that a trust system is applicable, whether the system uses individual or global aggregations of trust measures, whether the trust system assigns a reputation from one individual across all individuals who share an attribute or credential, whether and how individuals bias their observations, and how strongly a priori beliefs versus broad aggregate results versus deep propagation of reputation across agents affect reputation.

Sabater and Sierra [2005] classify related work on a number of dimensions. First is the conceptual model, whether the agents use cognitive beliefs or apply game-theoretic or Bayesian models. Another dimension is the information source, whether the information is directly observed, indirectly witnessed, assumed based on the agent's social roles, or an a priori prejudice based on observable information. The agents can obtain this information from their visibility type, which can be individual, meaning that the agent's observations are relative to itself typically from frequent interaction (moral hazard situations), or global, meaning observations that are common across agents and the interactions are infrequent for a given agent pair (adverse selection situations). The granularity of the measurements can be categorized by the number of contexts in which the trust is measured and whether information is aggregated between the contexts. Sabater and Sierra make a further distinction as to whether the trust information observed and exchanged between agents is Boolean or continuous, and whether the model uses a single value for a measure of reliability. They group the related work by how it handles agents' behaviors, whether the system depends on honest agents, assumes agents may bias information but not outright lie, or whether the system handles lying agents.

Jøsang, Ismail, and Boyd [2007] describe trust as a soft security mechanism. Their detailed view of the trust literature includes the following dimensions. First is reliability trust versus decision trust, which closely resembles the division between adverse selection and moral hazard as described by Dellarocas [2006]. This distinction is further split into collaborative filtering, which begins with the assumption that all information is truthful and filters accordingly, and reputation systems, where agents are assumed to be possibly untrustworthy and agents build trust. Jøsang et al. enumerate trust into the following types: provision trust, meaning quality of service; delegation trust, which is like provision trust, but performed on behalf of the agent; access trust, where an agent obtains access to information, resources, or control; identity trust, where an agent makes a claim about its identity; and context trust, which is provision trust at a certain task. They group trust systems into those that deal with qualified, affective ratings, those that use specific, single attribute ratings, and those that use general overall ratings. The final classification dimensions of trust systems are the extent that the reputation architecture is decentralized, the trust aggregation method (i.e., summation, Bayesian, discrete values, belief models, fuzzy models, flow models), and the application (i.e., eBay, expert sites, product review sites, Epinions, BizRate, Amazon, Slashdot, Kiro5hin, Google, supplier reputation systems, scientometrics). Jøsang et al. also discuss related work based on the problems managed, such as the low incentive for agents to rate others, biases toward positive ratings, unfair ratings, detecting unfair ratings, Sybil attacks, quality variations over time, discrimination, and ballot box stuffing.

1.1.2 Common Dimensions

In this section we examine the commonalities between the ways the other surveys classify trust systems. We group the dimensions as follows.

Incentive Compatibility. Incentive compatible means whether or not it is an agent's optimal strategy to play honestly as to its own type. This means that a rational agent will truthfully reveal any information necessary for an interaction. Given this property, a system generally does not need an additional reputation system unless it is part of the mechanism itself. A trust system can potentially be incentive compatible in only one direction. For example, an interaction mechanism where an agent pays for a commercial airline ticket upfront is generally incentive compatible for the buyer because the buyer is not given the opportunity to take the ticket without paying for it. The dimension of incentive compatibility is discussed by Ramchurn et al. [2004] and Dellarocas [2006].

Access versus Action. Access trust means verifying an agent's identity and granting it permission to perform some task, typically involving security and encryption. Access trust enables action trust, which involves agent interactions that depend on an agent knowing another, such as provision, delegation, and reciprocation. These two types of trust are of different scopes and are usually disjoint in the related literature. Trust systems are classified into these two categories by Ramchurn et al. [2004], Artz and Gil [2007], and Jøsang et al. [2007].

Focus on Adverse Selection. Adverse selection arises when an agent advertises a product or service at a quality different from the actual quality. This discrepancy between reported quality and actual quality does not necessarily mean that the agent is intentionally deceiving; adverse selection indicates nothing about intent. Adverse selection involves agents with a fixed type or attribute that the agent cannot easily change, for example, the quality of good that the agent can deliver. The solution to this problem of asymmetric information is for agents to learn about other agents' types and to be able to signal an agent's type to other agents. Statistics, data mining, and machine learning are useful tools for an agent to learn the quality of another agent's goods or services, but given a noisy measurement process and enough agents, some outlier agents will receive incorrect reputations. Adverse selection often arises when agents interact infrequently. For example, if an agent makes a one-time purchase from an online store, the rate of interaction between the agent and the online store is relatively low with respect to the transaction volume of the online store. Adverse selection is a trust dimension in the work of Sabater and Sierra [2005], Jøsang et al. [2007], Dellarocas [2006], and Ramchurn et al. [2004].

Focus on Moral Hazard. When an agent has the ability to increase its own utility at another agent's expense, the system is said to exhibit moral hazard. Moral hazard is different from adverse selection because the agent is in control of its actions. Statistics generally do not work well for modeling moral hazard, and the strategies must instead be examined in a game-theoretic manner. Moral hazard can be managed by credible sanctioning, where one agent can decrease another's utility if the other agent is not cooperating. Most real-world situations contain a mixture of adverse selection and moral hazard. However, few trust and reputation systems explicitly handle moral hazard, which is usually left to game-theoretic solutions, such as contrite tit-for-tat [Wu and Axelrod, 1995]. Moral hazard is supported as a dimension of trust by Sabater and Sierra [2005], Jøsang et al. [2007], Dellarocas [2006], and to a lesser extent by Ramchurn et al. [2004].

Dimension Dependency. Different trust systems measure trust in a different number of dimensions, such as performance or quality. For example, a simple measure of the probability of a positive interaction [Jøsang, 1998] would have one dimension, a measure of discount factor and reliability [Smith and desJardins, 2009] would have two dimensions, and a detailed review of a video game, including the graphics, sound, story, and gameplay qualities might have four dimensions. Under other names, dimension dependency is covered as a trust system property by Sabater and Sierra [2005], Jøsang et al. [2007], and Mui et al. [2002]. Artz and Gil [2007] break dimension dependency down into simple reputation versus separation into beliefs, risk, and utility.

Aggregation Breadth. Aggregation breadth is the extent by which a trust system is centralized or distributed, and governs the mechanism by which trust information is aggregated. A fully centralized trust system would accurately record information about transactions in one location and disseminate this information. In contrast, agents must keep track of all information on their own in a fully distributed trust system. Prior beliefs that agents have about other agents, particularly agents that have obtained credentials and conditional probabilities, factor into the initial biases and affect aggregation. Dividing trust systems by their breadth and method of aggregation is discussed by Ramchurn et al. [2004], Jøsang et al. [2007], Mui et al. [2002], Artz and Gil [2007], and Dellarocas [2006].

1.2 Approach: Trust Dynamics and Patience

Throughout the trust and reputation system literature, two techniques that stem from game theory are commonly applied for designing such systems. Signaling models are those in which agents attempt to assess private attributes about other agents, whereas sanctioning models are those in which agents behave strategically in an attempt to maximize their utility [Dellarocas, 2006].

In real-world environments where agents must decide whether or not to trust one another, clean distinctions between signaling and sanctioning are rare. For example, an agent that allocates its own bandwidth and other resources may have little influence over the amount of resources it has available. Yet, it may be strategic and rational within those constraints. A manufacturer can acquire a good reputation for having tight quality controls, but new management may wish to see larger profit margins and may strategically slowly cut back on the quality controls as long as it remains ahead of its competitors.

Despite the complexity of the real world, few reputation systems are specifically designed to address both sanctioning and signaling. Typically, authors of reputation systems that involve signaling devise a variety of malicious behaviors to test their system against. Examples of the adversary agents include randomized acts of unfavorable behavior [Kamvar et al., 2003, Huynh et al., 2006], building up and spending of reputation [Srivatsa et al., 2005, Kerr and Cohen, 2009, Salehi-Abari and White, 2009], Sybil attacks where an agent creates multiple identities [Kerr and Cohen, 2009, Kamvar et al., 2003, Sonnek and Weissman, 2005], and collusion with other agents [Kamvar et al., 2003, Sonnek and Weissman, 2005, Srivatsa et al., 2005]. Other systems

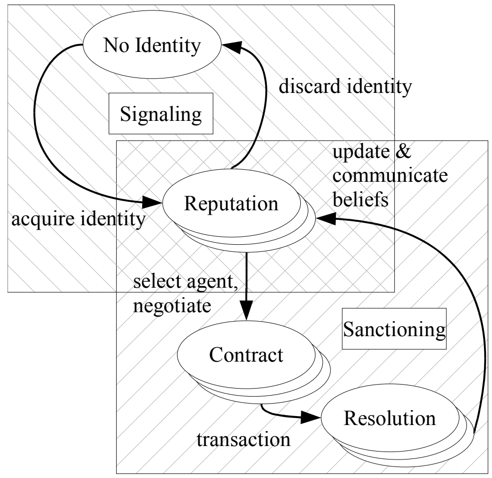


Figure 1.1: Trust and reputation life cycle from an agent's perspective.

are designed specifically around strategic agents to ensure good behavior, but do not attempt to measure attributes of the agents [Jurca and Faltings, 2007, Hazard, 2008]. A minority of reputation systems, such as that by Smith and DesJardins [2009], examine both signaling and sanctioning explicitly.

Our general view is that trust is looking forward in time with respect to sanctioning and strategy, whereas reputation is looking backward in time with respect to signaling and determining agents' types. We discuss this dichotomy in further detail in Section 2.1. Our primary focus is on rational agents and e-commerce settings, rather than on modeling human behavior. Emotional and cognitive factors of trust are thus outside of the scope of our study.

1.2.1 Trust and Reputation Life Cycle

Although the specifics of particular trust and reputation systems can differ greatly, they all share some commonalities. In this section, we unify the systems to a common set of states and actions as outlined in Figure 1.1.

The following are the different states an agent can go through in a transaction in the presence of an open reputation or trust system. An agent is not limited to being in one state at a time, but can maintain multiple accounts and participate in multiple transactions simultaneously.

No Identity The agent begins without an identity or account in the system. This state is applicable for open systems where agents may enter or leave. From this state, an agent may acquire an identity and move to the reputation state. Acquiring an identity may be as trivial as using a nonvalidated screen name in an open text field where the agent simply claims to have some identity. Alternatively, the system may require extensive background checks, verifications from official organizations, or significant payments to create the account. An agent may asynchronously acquire multiple identities, and may acquire identities in different domains or with different populations of agents.

Reputation Each identity that the agent has created will have its own reputation within the community. An agent may discard an identity, either actively by deleting an account or passively by simply no longer using an identity. When an agent decides to (or is forced to) interact with another agent, it must select an agent (or agents) with which to interact. It may communicate with this *target* agent, performing extensive negotiations and setting up a formal contract. Alternatively, the agent may simply rely on norms or not actively communicate with the target prior to the transaction.

Contract A contract expresses a promise or commitment to engage in some behavior. Contracts may be well-defined and policed by an external system or may be as ill-defined as the agents' a priori assumptions. From a contract, the agents involved undergo some transaction with the other agents involved. The transaction can involve active participation, such as exchanging money for an item, or a transaction can be passive, such as all agents timing out and not performing any task.

Resolution After a transaction has taken place, an agent will update its own beliefs about the agents involved in the interaction. The agent can evaluate, report, and communicate its new beliefs about another agent based on the results of the transaction, either directly to other agents or via a centralized reputation reporting mechanism. Concurrently, the agent may revisit the results and decide that further transactions are required. To set up future transactions, the agents may renegotiate to a new contract after having observed the other agents. A renegotiation can have positive connotations, such as providing additional services to supplement a previous transaction, or the renegotiation can have negative connotations, such as an agent demanding reparations from a transaction that did not fulfill the contract.

1.2.2 Rational Agents

Our approach to trust is from the standpoint of a self-interested rational agent. A rational agent maximizes its expected utility when planning and executing actions. In environments with uncertainty, rational agents must model uncertainty in order to compute their valuations. This extends to multiagent settings, where rational agents need to model other agents' behavior to devise a utility optimizing strategy. When all agents are rational, the best response strategies of all the agents are known collectively as Nash equilibria.

In general, we hold that rational agents can be described by the following attributes:

Valuations. A rational agent employs utility theory, assigning a utility value for every possible state and a utility cost for every possible action. The valuations may be a function of the state of the world. A goal could be described as a state with high utility. A valuation can be further used to model risk aversion by using a nonlinear function. For example, an agent could value having \$0 with utility of -100, having \$10 with utility of 1 and having \$100 with utility of 2.

Capabilities. An agent has certain capabilities, that is, ways to affect other agents and the environment. Capabilities may be probabilistic or have probabilistic chances of success. We include *reliability* within an agent's set of capabilities. We define reliability as the probability distribution of how a capability will impact another agent, that is, the probability distribution over the valuations that another agent will receive.

Initial Beliefs. When an agent begins operating in an environment, it has certain beliefs about the state of the world and about other agents. Such beliefs may be correct or incorrect, as well as probability distributions or single data points.

Computational Bounds. An ideal rational agent has infinite computational power, that is, it can solve the Nash equilibria of any game instantly before it needs to perform an action. Realistically, computational power is limited in some way, and so an agent may only have an approximately best strategy under some bounds.

Time Preference. An agent's patience, which we will refer to formally as *time preference*, means the mechanism by which the agent's perceived value of something changes as a function of the time that the value will be realized if all else is held constant. For example, a greedy agent may not spend money to maintain equipment because it does not value its future utility nearly as much as its present utility. Time preference can reflect the expected lifetime of the agent, uncertainty, and external factors.

Rational agents with different values of the aforementioned attributes can behave differently in otherwise identical situations. Due to the differing behavior, one agent could be considered more trustworthy than another. But which of these attributes define an agent's trustworthiness?

Our working hypothesis is that time preference, particularly when evaluated in light of valuations and capabilities, is what determines whether a rational agent is trustworthy. In treating trustworthiness as an attribute of an agent independent of its capabilities and valuations, we differentiate trustworthiness from uncertainty in the same way that moral hazard is differentiated from adverse selection. Time preference is a measurable and quantifiable attribute for evaluating an agent's optimal strategy.

We primarily consider interactions between only two agents, with multiparty interaction generally beyond the scope of this work. Further, we focus on trustworthiness with respect to rational agents and therefore exclude the emotional aspects of trust.

1.2.3 Discount Factors as Trustworthiness

A key intuition is that a trustworthy agent is patient, i.e., interested in long-term relationships: for example, we expect a store for local residents to sell better wares than a tourist trap. In general, anything is worth less in the future than now. With exceptions such as for storage, degradation, and depreciation, having money or a usable item is generally worth more now than later for reasons such as the uncertainty of the future and opportunity to use the item or money in the mean time. For example, most people would prefer $100 today over $100.01 next week. But one's premium for immediacy is bounded: typically, most people would prefer $1,000 tomorrow to $10 today. An agent's *intertemporal discount factor* reflects its break even point for the present versus the next time unit. For example, if you are neutral between $90 today and $100 tomorrow, then your discount factor is 0.90 (per day).

Further, trustworthiness and patience can vary with the *context*: a nearly bankrupt business facing its creditors may sell items without sufficient quality checks. We use context to refer to the risk environment that an agent facing, such as facing a pending bankruptcy or succeeding in a steady market. Outside of the mathematical use with respect to variables, we use *domain* to refer to a type of interaction, such as the role of a provider in a web services market versus the role of a seller in an online auction.

Definition 1 *An agent employs* exponential intertemporal discounting *in some context when there exists some* $\gamma \in [0, 1]$ *such that for all future times,* $t \in [0, \infty)$*, the agent's utility gain,* U*, from some event in that context at time* t *is* $U = \gamma^t u$*, where* u *is the utility the agent would have perceived had the event occurred at the present time (*$t = 0$*).*

An agent's discount factor captures how much it would value something at future points in time relative to the present. To put it into colloquial terms, an agent with a low discount factor will "take the money and run" whereas an agent with a high discount factor is "in it for the long haul." A higher discount factor can yield a greater payoff because the agent is not myopically optimizing, but this rule has exceptions [Ely and Välimäkiz, 2003, Hazard, 2008].

An agent's discount factor captures how much it would value something at future points in time relative to the present. Internally, an agent's discount factor can be influenced by intrinsic factors, such as uncertainty [Rubinstein, 2003], patience, or expected lifetime in a situation for an individual, or extrinsic factors, such as the cost of capital for a firm driven by market rates. The observable value that an agent uses for its discount factor can also be affected by external factors, such as time pressure and competition as discussed in Section 5.1.5, which may be different from a more consistent internal discount factor.

Knowing other agents' discount factors is important in determining an opponent's optimal strategy. In more complex models, discount factors can be used along with costs and valuations to explain agents' reliability and quality of goods and services. Assuming an agent can measure its own reliability, an agent with a high discount factor may take steps to increase its reliability if that meant that other agents would stop using its services if it did not.

Even though we intuitively associate trustworthiness with the expectation of future long-term relationships, most current approaches do not necessarily reflect this intuition. Existing measures of trustworthiness [Ramchurn et al., 2004] typically use arbitrary ratings or are highly dependent on the domain, distribution, and manner of interactions. A small body of related work has discussed some aspects of the relationship between discount factors and trust [Addison and Murshed, 2002, Deutsch, 1973, Whitmeyer, 2000]. However, with two exceptions [Hazard, 2008, Smith and desJardins, 2009], we are unaware of related work directly employing discount factors as a measure of trustworthiness.

Although discount factors are widely used in economics, finance, behavioral sciences, game theory, and artificial intelligence, each agent or firm is typically assumed to have a publicly known discount factor. In economics and finance, a convincing argument is made that a discount factor is common among all of the participants in a given situation. The argument is agents can alternatively choose to invest their money in large markets with public and competitive rates of return for a given level of risk, and therefore discount rates are fairly consistent for a given venture. Assuming discount factors are public is reasonable for certain areas such as finance, but even professional economists' opinions of an appropriate discount factor can have a wide distribution [Weitzman, 2001], further suggesting an agent's discount factor is specific to each agent. Discount factors are influenced by preferences of the person, agent, or organization. Though discount factors may be used strategically by opponents, many models require publicly known discount factors. Conversely, we focus on agents maintaining private discount factors and measuring others' discount factors.

1.2.4 Comparing Reputation Systems

Reputation is an important concept and computational reputation systems are popular primarily because there are strong intuitive connections between an agent's reputation and both the utility that it obtains and the utility another agent obtains when interacting with it. For example, an agent can obtain more money for the same products on eBay[1] simply by having a more positive reputation [Houser and Wooders, 2006]. A rational agent would only build and maintain a positive reputation if doing so maximizes utility. For example, in commerce environments, an agent can strategically build up and then expend its reputation in order to monopolize a market [Sen and Banerjee, 2006].

[1] `http://ebay.com`

Many authors propose desiderata to motivate their trust and reputation systems [Huynh et al., 2006, Kamvar et al., 2003, Zacharia and Maes, 2000]. However, we are unaware of a general characterization of desiderata for reputation systems that are quantitative, objective, and applicable across a wide range of domains. We present four desiderata, focusing on what quantitative properties make one reputation system more effective than another. Devising widely applicable metrics for trust is considered an important open problem [Barber et al., 2003] and is the focus of this work.

Our approach to comparing reputation systems involves examining the dynamics of a strategic agent's reputation. That is, we subject each reputation system to a utility maximizing agent with intertemporal discounting. We determine if and how the utility maximizing agent is able to circumvent the reputation mechanism's ability to measure whether the agent is desirable for interaction.

1.2.5 Challenges to Approach

A major challenge to studying trust and reputation of autonomous agents in e-commerce is that real-world data is practically nonexistent. Autonomous agents do not yet make critical evaluations and decisions in real-world settings in e-commerce. Online markets involving individuals, such as eBay, are largely driven by people and so human psychology makes the environment different compared to one run by rational agents. On the business scale, powerful automated markets for procurement and logistics, such as that provided by CombineNet,[2] are at the forefront of technology. However, even in the most advanced systems, most pricing, quality, reputation, and trust evaluations are not yet performed by autonomous agents. We are therefore limited to mathematical proofs, simulations, and analogous experimentation using rationally thinking, business-minded people.

Another technical challenge is computational complexity. Determining which agents are lying, computing optimal strategies and Nash equilibria for a rational agents in complex environments, and combining observations to maximize information gain and minimize assumptions can all be intractable for complex problems. Despite these challenges, we are still able to compute approximate results for many complex situations.

From a philosophical perspective, an argument is that trustworthiness is more than just patience. For example, consider a patient agent (one with a high discount factor) that is incompetent at all tasks it can perform. In a multiagent system, this agent would still attempt to take advantage of any other agent that did not know of its incompetency. Whereas one may argue that such an agent should be labeled untrustworthy, we instead denote this behavior under uncertainty; the other agents will soon learn that either the agent is either impatient

[2] `http://combinenet.com`

or incompetent, leaving uncertainty between the two choices. We further argue that the agent under question is, at its core, trustworthy, but is simply in a desperate situation.

The exact nature of time preference that should be used is a matter of some debate in economic settings [Groom et al., 2005]. We use the method of discounting from Definition 1 for one primary reason and two secondary reasons. The primary reason is that it is dynamically consistent, whereas hyperbolic discounting is not [Rubinstein, 2003]. The secondary reasons are more for convenience. Exponential discounting is widely used in economics and artificial intelligence literature. The other secondary reason is that exponential discounting only uses one parameter; dealing with trustworthiness as a single number and yields more tractable results and analysis.

1.3 Summary of Contributions

Motivating Question 1 *In an e-commerce setting with rational, self-interested agents, wherein discount factors are a common tool for modeling agent patience, are discount factors effective as a trustworthiness measure, that is, is modeling another agent's discount factor, valuations, capabilities, and environment sufficient to judge whether the agent will behave in a manner that can be described as trustworthy?*

Contributions

Claim 1 *Discount factors are isomorphic to trustworthiness given an agent's valuations, capabilities, and environment when trustworthiness is defined as a single scalar value that characterizes the long-term favorability of one agent over another to some third agent that will be engaging in interaction with one of the two agents being judged for favorability.*

We develop a model of trustworthiness as a discount factor that naturally captures the future long-term relationship intuition. Formalizing the intuition between trustworthiness and discount factors requires key assumptions that are not generally made explicit in related work. In Chapter 2, we formalize these important technical assumptions, which characterize trustworthiness from the perspectives of the trusting agent (preference property), the trusted agent (strength property), and the stability of the situation over time. In doing so, we not only delineate some key assumptions regarding trustworthiness but also obtain an objective trust measure in the nature of the discount factor, isolating an agent's objective trustworthiness from subjective effects. Our primary result is a proof that, given our widely applicable assumptions and definition of trustworthiness, any trust measure that collapses trustworthiness down to an individual objective scalar (real value) is isomorphic to intertemporal discounting.

We find that the discount factor model satisfies additional properties. Specifically, we identify crucial desiderata for a computational approach to trust. Our discount factor model meets these desiderata whereas other approaches generally fail some of them.

A significant advantage of a formal objective basis for trustworthiness is that it supports powerful approaches for reasoning about trust and for one agent to infer the trustworthiness of another based on the latter's actions. In Chapter 5, we consider a series of e-commerce situations where buyers and sellers estimate each other's trustworthiness based on signals such as the quality of products sold, prices offered and accepted, and eagerness to conclude a transaction. We show how information on trustworthiness may be aggregated and estimated, and conclude with a discussion of the implications of discount factors as trustworthiness measures.

Motivating Question 2 *How can reputation systems be evaluated and compared when faced with strategic agents for a given environment? More specifically, what measures can characterize how a reputation system will behave when faced with strategic agents, what measures indicate which reputation system will be more effective at correcting agents' beliefs, and what measures indicate which reputation system will be best at distinguishing between different agent types?*

Contributions

Claim 2 *Measurable dynamic properties of a reputation system, namely, monotonicity, accuracy, convergence, and unambiguity, are useful for characterizing a reputation system. Measuring and comparing these properties comprises an effective method to evaluate and compare the resilience of reputation systems against exploitative strategic agents.*

We approach reputation from a dynamic systems perspective. In Chapter 6, we motivate and formalize the following quantifiable desiderata.

Monotonicity. Agents who would provide favorable interactions should acquire better reputations than agents who would provide less favorable interactions. For example, a seller who always offers high-quality items at a low price should have a better reputation than an agent who produces defective items that it advertises as being of high-quality (and thus sells at a high price).

Unambiguity. An agent's reputation should be asymptotically unambiguous, meaning an agent's asymptotic reputation should be independent of any a priori beliefs about the agent held by some observing agent. An unambiguous reputation system would, as the number of interactions tends toward infinity, always yield the same reputation for a given agent regardless of the specific interactions. Consider two otherwise identical buyers (that is, identical in their valuations for goods of a given quality, utility functions, capabilities, influence over peers, and so on) who initially disagree about a seller's reputation. Both buyers should converge to an agreement about the seller's reputation after a sufficiently large number of interactions, assuming the seller behaves steadily in the same manner with each buyer.

Convergence. Agents' reputations should converge quickly. For example, it is preferable to be able to learn after a smaller number (rather than a greater number) of interactions whether a seller offers high or low-quality products, regardless of past beliefs, provided the seller keeps to its type. A reputation system must meet unambiguity for convergence to be meaningful.

Accuracy. Reputation measurements should be accurate regardless of prior beliefs. For example, if a buyer incorrectly believes that a seller produces high-quality items, the buyer should quickly learn an accurate reputation value for the seller.

Our desiderata apply to both adverse selection and moral hazard, with or without the propagation and aggregation of reputation information. The measurements from the desiderata can answer a wide range of questions, such as whether one reputation system is better than another, whether agents would benefit from using a reputation system, how stable the system is, and how quickly agents can build up or lose their reputation. Rather than examine and compare reputation systems against a list of possible attacks [Huynh et al., 2006, Kerr and Cohen, 2009, Kamvar et al., 2003], our desiderata compares general dynamical properties of the system as affected by strategic agents.

We apply our desiderata to a diverse group of trust and reputation mechanisms from the literature. In each case, we pair off an agent against an ideal rational agent. We primarily focus on the interaction between two agents, but we examine a few larger settings. We find a variety of desirable and undesirable behaviors across the models, finding the general mechanism proposed by both Hazard [2008] and Smith and desJardins [2009] to exhibit the most favorable results of those studied, although this mechanism does not adapt to a continuous range of behaviors as easily as some other systems. We analyze some real-world data retrieved from Amazon, and find that the data shares strong similarity to that shown by rational agents on its underlying Beta reputation model. We discuss the strengths and limitations of our desiderata.

Motivating Question 3 *Given rational agents with initially private discount factors in a setting where the agents can only offer favors to one another at a cost to themselves, such as peer-to-peer file sharing, unenforced market transactions, or resources within cloud computing, what rational yet mutually beneficial strategies can agents use to choose whether to offer favors* *to other agents given different interaction models, such as where the favors are alternating or simultaneous, and follow a simple stochastic process?*

Contributions

Claim 3 *Although many repeated games involving trust settings have an infinite number of equilibria, many strategies are both mutually beneficial and intuitive, such as strategies that balance agents' abilities to sanction one another and strategies that incentivize agents to remain in a relationship.*

In some environments, agents have recourse when one agent does not follow through with its commitments as part of an agreement, such as using a legal system to prosecute a breach of contract. However, using legal means or external enforcement mechanisms to seek retribution for an unfulfilled commitment is often more expensive than simply ostracizing the agent that did not fulfill its commitment. For example, an agent may deliver an item that is of low quality but technically meets the specifications of the contract, or an agent may simply not reciprocate uploading data to another agent in a peer-to-peer computing.

We examine various models of agent interaction stemming from favor reciprocity. The first is when agents stochastically and unilaterally decide whether to offer the other a favor, where the cost and value of the favor is chosen from a known distribution in Chapter 3. The second situation is when the agents are paired for a transaction, that is, both agents are simultaneously deciding whether or not to uphold their end of an exchange, which is the basis for many e-commerce transactions. This work is described in Chapter 4.

From these models, we find equilibria where agents hold sufficient ability to sanction each other such that rational agents sustain mutually beneficial relationships. We employ the basic principles of Chapter 2 in that agents are evaluating each others' discount factors. In Chapter 6, we compare our discount factor favor reciprocity model from Chapter 3 to other reciprocity models in the literature. We find that, aside from one limitation of our reciprocity model on continuous domains, our model outperforms all the others when faced against rational agents.

Chapter 2

Defining Trustworthiness and Trust Systems

To derive and state formal results about trust, we need to have a specific definition. As trust definitions are often mathematically informal, we create our own formal definition of trust in the context of multiagent systems, motivated by e-commerce, that enables us to derive useful results. In this chapter, we begin with common definitions of trust and reputation. Then, we show that a general definition of trustworthiness is isomorphic to discount factors in the context of agent's valuations. Further, we present general desiderata for trust systems rooted in common dimension classifications of trust systems.

2.1 Signaling Versus Sanctioning

The game-theoretic designations of signaling and sanctioning games are relevant to trust and reputation systems because they address the key mechanism of whether an agent must decide who to choose or how to act [Dellarocas, 2006, Jurca and Faltings, 2007]. In this section, we propose a way of determining the influence of signaling versus sanctioning and how these properties affect the design of a trust or reputation system.

In a signaling setting, agents have private information that they may use to their advantage. The asymmetric information can be used strategically to cause adverse selection, where agents perform transactions with agents they believe to be desirable but end up with an undesirable interaction. An example of a signaling situation is where agents are purchasing mass-produced products and deciding whether to buy the product from one manufacturer or another based on quality, price, and features. In this case, agents signal to each other what they believe about other agents (specifically, the manufacturers). Statistical and probabilistic measures are most effective at measuring agents' behaviors in the signaling setting.

Sanctioning mechanisms are useful in cases of moral hazard. Moral hazard occurs when agents' utilities are uncorrelated, meaning that one agent's gain may yield another's loss, and one agent can directly exercise control over another's utility. A purchase where a buyer pays the seller and then the seller has the option of not sending the product to the buyer is an example case of moral hazard. If the seller will not be sanctioned for its behavior and will have no future relations with the buyer, then it has no incentive to send the product. Sanctioning must be credible for the agents involved to be successful, and may be performed by the agent affected by refusing future transactions, or by other agents policing the system. Modeling behavior in a sanctioning environment with rational environments means employing game theory techniques to find Nash equilibria.

As we remarked above, many real-world situations do not fall cleanly into either signaling or sanctioning situations. An agent may have some control over the quality of its products. In real-world scenarios, it is unlikely that an agent would be unable to make any changes to quality (pure adverse selection) or for an agent to have perfect control over quality (pure moral hazard). This distinction is blurred further by agents having differing levels of patience that influence their strategic behavior [Hazard, 2008, Smith and desJardins, 2009] and also by the blurred distinction of whether an observation was intentionally communicated [Castelfranchi, 2006]. The amount of sanctioning comes down to how much explicit control an agent has over its communications, and also intent, which may be subtle.

In broad terms, we can distinguish two varieties of trust that apply in many computational settings with intelligent agents. We abstract the terms *Competence* and *Integrity*, as described by Smith and DesJardins [2009], into *Capabilities*, which are what an agent *can* do, and *Preferences*, which are what an agent *will* do. From these definitions, it is clear to see that when agents want to determine which other agents have capabilities, they need a signaling system that looks into what the agents have done before. Agents need a sanctioning system to determine another agent's preferences and ensure that the agent will perform a desirable behavior in the future when it has the choice. This is consistent with the notions of **reactive** and **anticipatory** coordination [Castelfranchi, 1998].

To examine the role of signaling versus sanctioning on reputation systems, it is instructive to consider three interrelated terms—trust, trustworthiness, and reputation—that are used in nonstandardized ways in the literature. We begin from basic definitions in order to capture the general intuitions about them.

Trust is an agent's assessment of another party along some dimension of goodness leading to expected outcomes.

Trustworthiness is how good a party is in objective terms. In other words, this is a measure of how worthy the party is to be trusted.

Reputation is the set of general beliefs (among the agents in a society or community) about a party.

Specifically, Alice may or may not trust Bob for possessing desirable attributes (these could be capabilities, resources, bandwidth, and such). Alternatively, Alice may or may not trust Bob for having his preferences aligned with hers or rather for having his preferences aligned with hers under a particular incentive mechanism. Bob may or may not be worthy of any trust Alice may place in him. Bob may or may not have a reputation for being trustworthy in the specified ways. And such a reputation may or may not be well earned.

Reputation and trust therefore can be fit into our dual categorization. Reputation involves what an agent is, as measured from its past; an agent has a reputation of having some attribute or capability, and so a reputation system in this sense is a signaling system. Trust is concerned with what an agent will do in a future situation, which concerns the agent's preferences and must be handled by a sanctioning system. However, as trust and reputation have other connotations in specific domains, such as emotion, we will maintain the distinction using the terms signaling and sanctioning.

2.1.1 Measuring Influence of Signaling and Sanctioning

Consider agents a and b that have fixed behavior, behaving virtually the same way regardless of the situation (e.g., by offering products of some specific quality). An example of such an agent is one that controls a high-volume web service with specific offerings and finite bandwidth with little autonomy and limited reasoning capabilities. Consider an agent c that is deciding whether to interact with agent a or agent b. If c chooses a, then c will receive some benefit (or loss) of utility, u_a. If c chooses b, then c's utility would be changed by u_b. Since the agents have fixed behavior, c's behavior other than choosing a or b will not make much difference. To maximize utility, c should use a reputation system (using statistics, for example) to measure how a and b's behave before making the decision.

Conversely, consider that agents a and b are rational, have full and precise control over each of their actions, and may change their behavior without any switching costs. An example of these agents would be low-volume reseller agents that have a sufficient supply of resources offering substitutable products or services. In this case, whether c chooses a and b matters little to c's utility. Instead, c's choices in negotiation and behavior with respect to a or b dominates c's change in utility. Finding an optimal interaction strategy is how c can maximize its utility, and c should use a trust system (game theoretic modeling) to determine its best course of action.

If we write the benefit some agent c will gain with behavior x when choosing agent a as $u_{a,x}$, then the magnitude of the difference of utility change between choosing agents a and b while c maintains consistent behavior is $|u_{a,x} - u_{b,x}|$. Using c's utility maximizing behavior, this difference can be written as $\max_x |u_{a,x} - u_{b,x}|$. When evaluated against every agent in the set of agents available for c's consideration, S, agent c's maximum difference in utility between interacting with any two agents, $d_{\text{selection}}(c)$, can be written in terms of the rate of interaction between c and another agent a, $r_{a,c}$, and the set of all of c's possible behaviors, H, as

$$d_{\text{selection}}(c) = \max_{a \in S, b \in S, x \in H} |r_{a,c} \cdot u_{a,x} - r_{b,c} \cdot u_{b,x}|. \tag{2.1}$$

We may write the maximum utility difference between any two behaviors, $d_{\text{strategy}}(c)$, as

$$d_{\text{strategy}}(c) = \max_{a \in S, x \in H, y \in H} |r_{a,c} \cdot u_{a,x} - r_{a,c} \cdot u_{a,y}|. \tag{2.2}$$

The maximum effect of the choice of either agents and strategies on utility, $d_{\text{total}}(c)$, can be expressed by

$$d_{\text{total}}(c) = \max_{a \in S, b \in S, x \in H, y \in H} |r_{a,c} \cdot u_{a,x} - r_{b,c} \cdot u_{b,y}|. \tag{2.3}$$

As $d_{\text{selection}}$ measures the impact of an agent's type and d_{strategy} measures the impact of an agent's strategy, we can use these values to determine the impact of signaling and sanctioning on a multiagent interaction relative to the total effect of choice on utility. In aggregation, we express the average value of each of the values across all agents as $\overline{d_{\text{selection}}}$, $\overline{d_{\text{strategy}}}$, and $\overline{d_{\text{total}}}$. The fraction of agents' total utility in a system that can be affected by signaling, $i_{\text{signaling}}$, can be represented as

$$i_{\text{signaling}} = \frac{\overline{d_{\text{selection}}}}{\overline{d_{\text{total}}}}. \tag{2.4}$$

The fraction of utility that can be affected by sanctioning, $i_{\text{sanctioning}}$, can be represented as

$$i_{\text{sanctioning}} = \frac{\overline{d_{\text{strategy}}}}{\overline{d_{\text{total}}}}. \tag{2.5}$$

The contributions of selection and strategy must account for the total range, thus $\overline{d_{\text{total}}} \leq \overline{d_{\text{selection}}} + \overline{d_{\text{strategy}}}$.

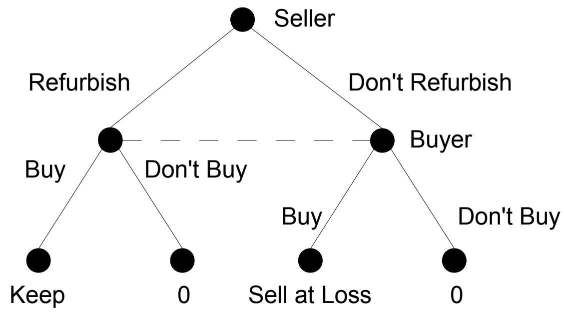


Figure 2.1: Online auction extended form game.

Seller Agent	Refurb. Value	Refurb. Market Price	Unrefurb. Price	Refurb. Cost
A	\$500	\$400	\$200	\$150
B	\$490	\$350	\$250	\$80

Table 2.1: Online auction refurbished laptop example data.

2.1.2 Example: Online Auction Representation

We use a simplified online auction interaction to show an example of applying our signaling versus sanctioning measure. Suppose agents are participating in an online market for refurbished laptops as depicted by the extended form game in Figure 2.1 with prices outlined in Table 2.1.

Suppose a buyer agent, C, values the refurbished laptop from A at \$500 and the refurbished laptop from B at \$490. Agent C needs to decide whether to buy from A or B for the market price of \$400 or \$350, respectively. It costs A \$150 to refurbish its laptop that it bought unrefurbished at \$200, and costs B \$80 to refurbish the laptop it purchased at \$250. Both A and B are claiming that the laptop on sale is refurbished, but C does not know for sure.

First, we investigate the case of selection. Agent C can select to buy from A or B, but A and B have no choice in the matter because of the online auction format. The rates of interaction from A's perspective are $r_{A,A} = 0$, $r_{A,B} = 0$, $r_{A,C} = 1$, and B is analogous. The rates from C's perspective are $r_{C,A} = 1$, $r_{C,B} = 1$, $r_{C,C} = 0$.

Agent A only can interact with C, and the maximum profit A could make while still providing a laptop at the market price of \$400 is \$200 if it did no refurbishment on the \$200 original unrefurbished laptop. Similarly, $d_{selection}(B) = \$100$. To compute $d_{selection}(C)$, we must first evaluate which strategy yields the greatest difference between choosing A or B. When the seller performs the refurbishment, C's difference in utility between choosing seller A and B is $|(\$500 - \$400) - (\$490 - \$350)| = \$40$. When the seller does not perform the refurbishment, the difference becomes $|(\$200 - \$400) - (\$250 - \$350)| = \$100$. As the rates of interaction are symmetric, the larger of these two yields $d_{selection}(C) = \$100$. The average value of the difference of selection across all three agents is the average expressed as $\overline{d_{selection}} = (\$200 + \$100 + \$100)/3 \approx \$133.3$.

Next we investigate the case of sanctioning for each agent. Agent A can choose whether or not to refurbish at the cost of \$150. Therefore, we find $d_{\text{strategy}}(A) = |(\$400 - \$200 - \$150) - (\$400 - \$200)| = \$150$, which is the cost of refurbishing the laptop, and accordingly $d_{\text{strategy}}(B) = \$80$. To find $d_{\text{strategy}}(C)$, we also examine the sellers' behavior. If A does not refurbish the laptop before shipping it, but instead delivers a broken laptop, then C regains only \$200 from externally selling the laptop in another market at the unrefurbished price and loses its \$400 payment. Applying this evaluation with both A and B, $d_{\text{strategy}}(C) = \$300$ because A deciding whether or not to refurbish is the largest difference in values. Putting the three of these agents' results together, we obtain
$\overline{d_{\text{strategy}}} = (\$150 + \$80 + \$300)/3 \approx \$176.7$.

Despite the multitude of combinations of agents and strategies, finding $\overline{d_{\text{total}}}$ is easy for this example because the problem is small and the extrema are easy to intuitively find. As A and B do not have a choice in which agents they interact, their values for $\overline{d_{\text{total}}}$ are the greater of $\overline{d_{\text{strategy}}}$ and $\overline{d_{selection}}$. The largest utility gain C can obtain from a transaction is \$140 if B provides a refurbished laptop, and C's the worst case is if it unknowingly buys a unrefurbished laptop from A and loses \$300. The total difference for C is thus \$440, and the average across all agents is $\overline{d_{\text{total}}} = (\$200 + \$100 + \$440)/3 \approx \$246.7$.

The system has $i_{\text{signaling}} = \frac{\$133.3}{\$246.3} \approx .54$ and $i_{\text{sanctioning}} = \frac{\$176.7}{\$246.6} \approx .68$. An effective reputation system for this system should emphasize sanctioning mechanisms slightly over signaling mechanisms. For agents within this system, these values mean that utilities can be affected more strongly by what other agents do and the attributes of relations rather than the particular agents involved in a relation.

2.2 Defining Trustworthiness

To define trustworthiness, we first must have definitions of how agents interact. We define an *event*, i, as a pair $\langle u_i, t_i \rangle$ consisting of a change in utility, $u_i \in \Re$, to some agent at a specified time, $t_i \in \Re$. We define an event as an isolated, independent change in utility, given all externalities, conditions, and decisions that create the event. An event may have additional side effects, such as altering the utility of another agent, but as these are not essential to our discussion and formalisms, we exclude them in our notation and define an event as a pair for clarity.

We use the following notation. Each agent, a, has a total expected utility function, U, that yields the agent's total utility given its trustworthiness and a set of events of utility changes. The function may be written more formally as $U : \Gamma \times I \mapsto \Re$, meaning that the total utility function takes in a real value of trustworthiness, $\gamma_a \in \Gamma$, and a set of events, $I = \{i_1, i_2, \ldots i_n\}$, and yields a number for the total utility of the events. We write it in the form $U(\gamma_a, I)$.

In our running example, an event is a cash flow or an change in ownership or status of a good. At the time when a seller transfers the ownership of the item to the buyer, the buyer receives some utility at that time. The utility gain that the buyer receives may be an expected value if the buyer is planning on reselling the item, perhaps after additional manufacturing or configuration, for a profit. When the seller receives money for the good or service, the event is to add money to the seller's account at the time when the buyer pays.

As in the running example, we restrict our attention to trust with respect to future actions. This would eliminate some English uses of the word "trust" such as "I trust book reviews on Amazon," because there is no future action there. It would allow "I trust Amazon to send me the book on time," which involves a future action.

2.2.1 Assumptions

We assume trustworthiness is reasonably fixed for the time frame in which the agents act. This is reasonable because if trustworthiness changed quickly, for example, if sellers frequently and unpredictably changed their type, a measure of trustworthiness would not be useful for predicting outcomes.

This does not mean that trustworthiness is fixed for a given agent. Models in which agents' types change [Mailath and Samuelson, 2006] are compatible with our approach.

Assumption 1 *An agent's trustworthiness is consistent enough to be meaningful across interactions; recent measurements of an agent's trustworthiness, if accurate, should usually reflect the agent's current trustworthiness.*

Assumption 1 merely requires that the rate of change for agent types is sufficiently lower than the rate of interactions so that knowing another agent's type is useful in an agent's decision model.

Utility theory lies at the core of e-commerce and postulates that agents have valuations for goods or services. A common currency is obviously desirable for commerce [Willmott et al., 2002], and enables agents to compare their valuations.

Assumption 2 *A utility loss or gain by one agent can be directly compared to the utility loss or gain of another agent.*

Quasilinearity means that the utility gained from isolated independent events is additive over the range of utilities involved such that an agent's total utility is closely approximated by the sum of all of its utility changes. The property of quasilinearity is frequently assumed in consumer theory and e-commerce [Walsh and Wellman, 2003].

Assumption 3 *Each agent has quasilinear utility; given two events yielding utilities at the present time of* u_1 *and* u_2*, the agent's total utility,* U*, is* $U = u_1 + u_2$*.*

Individual rationality means that an agent will not enter into nor fulfill a commitment unless doing so maximizes the agent's utility. A buyer will not purchase an item that is greater than its willingness-to-pay for that item, assuming that willingness-to-pay accounts for any expected benefits or losses indirectly associated with purchasing the item, such as when an otherwise unnecessary purchase is made to improve a relationship. Individual rationality is a core foundation of autonomous agents in much of the e-commerce literature [Wellman, 1996].

Assumption 4 *Agents are individually rational.*

2.2.2 Intuitions about Trustworthiness

Trustworthiness inherently involves settings where agents directly or indirectly engage in behavior that affects each others' utilities. The concept of a *commitment* helps capture this relationship. A *debtor* (agent) commits to a *creditor* (agent) to bring about an *event* [Singh, 1999]. In essence, a commitment reflects a dependence of the creditor on the debtor.

Definition 2 $\mathsf{C}(\mathrm{b}, \mathrm{a}, \mathrm{i})$ *is a* commitment *from debtor* b *to creditor* a *that* b *will bring about an event* i *at time* t_i *yielding a positive utility to* a *and a negative utility,* u_i*, to* b*.*

We restrict attention to commitments that require a negative utility for the debtor simply because commitments that yield positive utility to all parties with no risks does not require trust in our sense. In other words, we seek to capture the intuition about a debtor's trustworthiness based on the troubles it will go through to fulfill its commitments.

Often, in e-commerce, commitments would occur in complementary pairs so the overall situation would be win-win. For example, when a buyer commits to paying a seller and the seller to providing goods to the buyer, both benefit from the transaction. Indeed, given individual rationality (Assumption 4), every commitment that an agent enters must entail the expectation of a complementary commitment, such that the expected sum of the utilities is positive. Agent a may have beliefs as to how it will be repaid, such as having a 50% chance of b deciding on event i and a 50% chance of b deciding on event i'. When evaluating its total utility function, a should evaluate this as the expected value $\frac{U(\gamma_a, i) + U(\gamma_a, i')}{2}$, which holds due to Assumption 3.

The success or failure of a commitment provides a basis for the creditor to measure the trustworthiness of the debtor. For example, b may commit to deliver an item of a specified

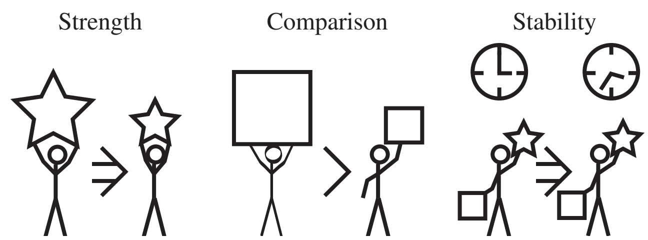


Figure 2.2: Illustration of intuitions about trustworthiness.

quality to a. If b fulfills a commitment $\mathsf{C}(b, a, i)$, a neutrally or positively updates its view of the relationship between a and b. If b fails to fulfill this commitment, a negatively updates its view of the relationship between a and b.

We now motivate some key intuitions regarding trustworthiness, which we then combine in our proposed definition of trustworthiness. Figure 2.2 illustrates the intuitions except *scalar*.

Scalar. Representing trustworthiness as a single value in a given context is a natural convention. One may ask, "How much do you trust b to produce and deliver some item with quality of at least X?" and receive a reply of "a lot." Such a value can be quantified; many online services provide ratings as points or percent of customers satisfied. A scalar representation does not preclude an agent from holding additional beliefs of the value or accuracy of trustworthiness, such as a probability distribution, nor from requiring additional information when making a decision of whether to trust, such as how much the trustee values something. Further, we can use different scalars for each context. More formally, we say that the set of trustworthiness values is the set of real numbers, $\Gamma = \Re$.

Comparison. A trustor a can compare two trustees b and c. Specifically, a considers b more trustworthy than c if, all else equal, b would be willing to suffer a greater utility loss than c would to fulfill the same commitment to a. In essence, a must know something about the valuations and costs incurred by both b and c and be able to compare these values as supported by Assumption 2. This does not mean that a will receive more utility from b's commitment than c's commitment, only that b is fulfilling a more costly commitment. Formally, agent a would consider agent b more trustworthy than agent c if, all else equal, for some event i with positive utility to a, there exist commitments $C_b = \mathsf{C}(b, a, i)$ and $C_c = \mathsf{C}(c, a, i)$ such that b would fulfill C_b and c would not fulfill C_c.

If c does not fulfill its commitments to a, by our definitions, this necessarily entails the loss of expected utility by a. If a pays c to deliver an item at a specified quality and c fails to deliver the item or provides an item of low quality, a will have gained less utility than it expected and incurred a negative net utility. This decrease in net utility causes strain on the relationship, causing a to either retaliate against c, such as by posting negative comments about c causing other agents to avoid transactions with c, or to avoid future loss by reducing its involvement with c by not making further purchases from c. In either case, c will initially have greater utility from incurring less cost by providing a lower quality item, but possibly lose more utility over the long term.

Strength. The behavior of each agent is internally consistent. Given equal impact on a relationship, if an agent is willing to do something difficult to keep a commitment, it should be willing to do something easy. If an agent is willing to deliver 1,000 gallons of kerosene to fulfill a commitment, then the agent should be willing to deliver 600 gallons of kerosene if everything else in the overall commitment stays the same (provided that storing or disposing of the other 400 gallons is not more difficult or costly than delivering it). From the perspective of the debtor, this property does not require actual fulfillment, it only requires that the agent be willing to exert the effort (sacrifice utility). If an item arrives late due to extenuating circumstances, this does not mean that the seller is necessarily less trustworthy. However, the creditor may only lessen its negative interpretation of an unfulfilled commitment if the creditor has some belief of noise in the signal of whether commitments are fulfilled. Formally, consider events i, j where $u_i \leq u_j$ and agents a, b. If b fulfills $\mathsf{C}(b, a, i)$ then b fulfills $\mathsf{C}(b, a, j)$.

Stability. The idea of stability is that agents should tend to behave in a manner that reflects a consistent underlying level of trustworthiness, which stems from Assumption 1. This essentially means that an agent, at the present time, considers its trustworthiness to be consistent for modeling future interactions. Using our online market example, an agent should be approximately equally trustworthy if a commitment will be set up now or one month from now, presuming the agent and market remain constant with regard to price, demand, supply, reputations, and reliability of available information. For example, suppose a firm can be trusted now to successfully deliver an order of 20 microphones of a certain quality within two weeks of payment. Then, if all else (e.g., external prices, internal staffing, and such) remains consistent, the firm can be trusted to deliver the same order if it were placed several months later again within two weeks of payment. Suppose the same firm is indifferent to committing to a delivery of 20 microphones and a delivery of 5 speakers today. If again, the environment and agents' valuations stay the same, the firm should be indifferent to those two commitments if asked again in a month. More formally, if an agent is indifferent between two commitments or sets of events, I_1 and I_2, then it should also be indifferent if the time is shifted by some arbitrary s. This may be expressed as

$$U(\gamma, I_1) = U(\gamma, I_2) \Rightarrow U(\gamma, \{\langle u_i, t_i + s\rangle : i \in I_1\}) = U(\gamma, \{\langle u_i, t_i + s\rangle : i \in I_2\}). \quad (2.6)$$

Stability means that an agent should tend to behave in a similar manner across a period of time, but this does not mean that an agent is indifferent between when an event or commitment may happen. An agent may prefer to receive an item sooner rather than later. We are simply stating that, given identical circumstances, an agent would enter the same commitments if they were shifted by some time because the agent is stable. If properties of the environment, agents' valuations, or agents' trustworthiness change, the agents may model such changes and factor them into their decision making however appropriate.

Definition 3 *The* trustworthiness *of agent b from a's perspective is a scalar value that a believes to be an accurate projection of an b's attributes that a can use to compare b to other agents (*comparison*) with respect to the utility that b would be willing to sacrifice to fulfill a commitment to a (*strength*), that is also relatively stable across time (*stability*).*

2.3 Trustworthiness and Discount Factor Isomorphism

We now derive our main result: an agent's discount factor is a direct measure of its trustworthiness given assumptions.

Because previous changes of utility are accounted for in an agent's current utility, it is only useful to evaluate the impact of future changes to utility. We therefore restrict the domain of t_i to $[0, \infty)$.

Theorem 1 *Given commitment as in Definition 2, trustworthiness as in Definition 3, and Assumptions 1, 2, 3, and 4, the representation of trustworthiness satisfying these definitions is isomorphic to an intertemporal discount factor.*

Proof 1 *By Definition 2, the utilities of any two events i and j are independent. This definition, coupled with Assumption 3 of quasilinearity, implies that an agent's total utility, U, is a summation of some utility function for each event, f, over all of the events, with $\frac{\partial f}{\partial u_i} > 0$. With trustworthiness γ and the set of events I, this is given by*

$$U(\gamma, I) = \sum_{i \in I} f(\gamma, u_i, t_i), \tag{2.7}$$

Given comparison *(supported by Assumption 2) and* strength*, an agent, b, is considered more trustworthy than another, c, if b will fulfill a commitment requiring a larger expenditure than c. This implies there is a commitment of some cost that b will fulfill and c will not; below this cost, both agents would fulfill the commitment. We only need to examine an individual event, and can restate this property using the event utility function, f.*

Let us evaluate agents b and c with trustworthiness γ_b and γ_c, respectively. Let agent a expect a commitment, $\langle u_1, t_1 \rangle$, to be fulfilled by the agent in question where, by Definition 2, $u_1 < 0$. Further, suppose that if the commitment is fulfilled, a will provide some utility back to the respective agent in the continued relationship: as Section 2.2.2 explains, at least two complementary commitments are required for agents to enter into commitments. We examine the simplest case, where this returned utility is expressed by a single event, $\langle u_2, t_2 \rangle$, such that $u_2 > 0$ and $t_2 > t_1$.

From Assumption 4, $f(\gamma, u_1, t_1) + f(\gamma, u_2, t_2) > 0$ for b and c; otherwise the relationship is destructive and rational agents would not engage in the commitments. Suppose b chooses to fulfill its commitment and c chooses to not fulfill its commitment. Their decisions show $U(\gamma_b, \{\langle u_1, t_1 \rangle, \langle u_2, t_2 \rangle\}) > U(\gamma_b, \emptyset)$ and $U(\gamma_c, \{\langle u_1, t_1 \rangle, \langle u_2, t_2 \rangle\}) \leq U(\gamma_c, \emptyset)$. If no events occur to change an agent's future utility, the agent's utility does not change, so $U(\gamma_b, \emptyset) = U(\gamma_c, \emptyset) = 0$. This implies, given the above assumptions of the two-event interaction set, that

$$U(\gamma_b, \{\langle u_1, t_1 \rangle, \langle u_2, t_2 \rangle\}) > U(\gamma_c, \{\langle u_1, t_1 \rangle, \langle u_2, t_2 \rangle\}). \tag{2.8}$$

Because b fulfilled a commitment that was larger than c would fulfill, by comparison *and* strength*, b is more trustworthy than c. If b is more trustworthy than c, then its trustworthiness value is higher, meaning $\gamma_b > \gamma_c$. We can take the limit as $(\gamma_b - \gamma_c) \to 0$, to find that*

$$\frac{\partial U}{\partial \gamma} \geq 0 \tag{2.9}$$

holds in this scenario with two events. This means that more trustworthy agents, when their trustworthiness is known to each other, attain higher expected utility than untrustworthy agents in two-event scenarios, all else being equal.

Stability*, supported by Assumption 1, entails that agents are consistent in their trustworthiness. The outer operation of U in (2.6) is a summation, and the number of terms in each summation (the number of events in each set of events) are not necessarily equal. Therefore, the only two possibilities that allow both equalities to hold are that time has no effect on events' utilities or that a change in time results in a constant multiplicative factor across all terms in a summation independent of the utilities.*

First, we consider the case where a change in time results in a constant multiplicative factor. The event utility function f must contain a multiplicand of the form x^t. This is because, given $x \geq 0$, x^t exhibits the appropriate behavior of $x^{t+s} = x^s \cdot x^t$ with x^s being constant for a constant time s. The first case, where time has no effect on f, can be represented by the second case with $x = 1$.

At this point, x remains an undefined attribute that affects the utility evaluation. Supposing

x did not affect the trustworthiness of an agent, if b is more trustworthy than c, then (2.8) must hold. Setting $x = 0$ for agent b would violate this inequality. As this contradicts the assumption that x cannot affect the trustworthiness of the agent, x therefore directly affects the trustworthiness of an agent.

Given scalar*, only one attribute may affect the trustworthiness of an agent. We now check to make sure that x satisfies the constraints of γ. In the two-event scenario, when $U > 0$ as given by Assumption 4, $\frac{\partial U}{\partial x} = t_1 u_1 x^{t_1 - 1} + t_2 u_2 x^{t_2 - 1}$. Because $x \geq 0$, $t_2 \geq t_1$, $u_1 < 0$, $u_2 > 0$, and $U = u_1 x^{t_1} + u_2 x^{t_2} > 0$, we can solve U for $u_2 > -u_1 x^{t_1 - t_2}$, and substitute the infimum of u_2 in this expression (and any greater number) into the expression for $\frac{\partial U}{\partial x}$ to find $\frac{\partial U}{\partial x} \geq 0$. This satisfies (2.9), thus satisfying* strength *and* comparison *(x came out of a derivation of* stability*).*

Substituting γ for x and rewriting in the form of (2.7), we find $U(\gamma, I) = \sum_{i \in I} \gamma^{t_i} u_i$. Revisiting (2.9), $\frac{\partial U}{\partial \gamma} = \sum_{i \in I} t_i \cdot \gamma^{t_i - 1} u_i$. To prevent imaginary terms for events with $t_i < 1$, the constraint of $\gamma \geq 0$ is required. This final utility equation coupled with the domain of γ is, by Definition 1, exponential intertemporal discounting.

2.4 Desiderata for Trust Systems

Devising optimal designs of general-sum multiplayer games is a difficult and domain-dependent problem. However, general desiderata can help guide interaction design. Such desiderata include individual rationality, guarantee of attaining a minimum payoff, guarantee of payoff to be within some ϵ within a best response strategy, and Pareto optimality when an agent is playing against its own strategy [Vu et al., 2006]. However, the desiderata for trust and reputation systems are not quite as straightforward [Dingledine et al., 2000] because trust and reputation are supplemental to *primary interaction mechanisms*. A primary interaction mechanism is one, such as a market, that affects agents' utilities directly.

A key motivation for work on trust is that the primary interaction mechanism is not incentive compatible (IC). Were it so, the agents would act honestly out of self interest. Our desiderata not only apply well when the primary mechanism is not IC, but also work when it is IC. Incentive compatibility is highly desirable for mechanism design, but achieving incentive compatibility may not be computationally feasible [Conitzer and Sandholm, 2004]. Further, an IC mechanism may not be in the best interest of the agent or firm running the mechanism, because an IC mechanism may not maximize profit.

Many papers on trust propose desiderata [Huynh et al., 2006, Kamvar et al., 2003, Ramchurn et al., 2004, Teacy et al., 2006, Zacharia and Maes, 2000]. Dingledine et al.'s [2000] desiderata list is the only comprehensive one we have found, but even their desiderata list focuses on aspects that are specific to certain kinds of reputation systems. We now propose desiderata that apply even when no central authority is available to enforce interactions or sanctions, and which focus on top-level goals that directly benefit the agents or system. A desirable system must be:

Evidential. An agent should use evidence-based trustworthiness measurements to predict future behavior. This is the essence of a trust system, with an agent rationally assessing others' behavior and acting upon its knowledge. Evidence also includes temporal relevance; new evidence that an agent has successfully changed its type, if credible, should indicate to another agent that old evidence may no longer be relevant. In the online market example, an agent should measure trustworthiness in a quantifiable and repeatable manner based on the quality of goods and timeliness of their offers, to determine how to best engage in future interactions.

Aggregable. Trustworthiness measurements should be accurate, precise, and possible to aggregate. This is key because aggregation enables an agent to communicate about trustworthiness and to put together indirect information obtained from other agents to increase knowledge of other agents' trustworthiness. In the market model, this aggregation involves reading others' comments, albeit with skepticism, to maximize the information considered.

Viable. The system should be practical in its computation and communication requirements. An approach that requires an exponentially large number of messages among buyers and sellers or requires each agent to perform an NP-Hard computation on a large dataset would not be tractable.

Robust. Measurements should be robust against manipulation; agents may signal or sanction to determine which agents are of what type and to resist strategic manipulation of the measurements. Manipulation can come in many forms, such as building up a reputation and then spending it, opening many pseudonymous accounts to communicate an inflated reputation to legitimate agents (Sybil attack), and opening a new account to expunge a bad reputation. We do not assume an incentive compatible reputation mechanism [Jurca and Faltings, 2007]. IC would be ideal for ROBUSTNESS, but can be impractical in some problem domains, either because of computational or communicational complexity conflicting with VIABILITY, or because of unenforceability if agents can deviate from the specified mechanism without credible consequences.

Flexible. Trustworthiness should be applicable across multiple situations within the same context. Trustworthiness measurements should carry over across products, services, and even interaction mechanisms. Suppose a seller is running a web service from which buyers can purchase directly, but also sells some of its items in a simultaneous ascending auction run by a third party. A buyer should be able to carry over knowledge of trustworthiness about the seller from direct sales to infer information about the quality of the items sold on the third party's auction and vice versa, even though the mechanisms are different. If a buyer becomes a seller, its reputation as a buyer should be indicative of its behavior as a seller, provided other agents can infer some knowledge of valuations, capabilities, and beliefs in the new domain.

Privacy enhancing. The system should maximize agents' privacy by minimizing the collection of information. The implications on a system can be quite broad. We use privacy in this sense to indicate that the public exposition of an agent's attributes is minimized. We differentiate privacy from anonymity. Anonymity is the antithesis of reputation; an agent must be (at least pseudonymously) identifiable in order for others to learn about its trustworthiness. Privacy can prevent an agent's identity outside of the system from being known. Thus maximal privacy would reduce the burden of an agent entering or leaving the system. This is because some cost is incurred by an agent divulging its identity in the system, such as the opportunity cost of preventing the agent from assuming a new identity within the system when its reputation is bad. Less privacy can also imply that the agent has some external account or information that the system could use to sanction it. In this sense, privacy acts as a liability limitation much in the way that a firm partially disassociates liability from its employees. The benefits of privacy are that agents have reduced friction of entering and leaving the system. The drawbacks include 1) a possible influx of unfavorably typed agents and 2) agents with bad reputations reverting to a neutral reputation. Both drawbacks are dependent on how other agents measure and handle trustworthiness.

2.4.1 Existing Trust Systems

Yu et al. [2004] provide a method for discovering peers and communicating reputations that maintains accuracy against noisy ratings and malicious peers. However, Yu et al.'s mechanism is weak against ROBUSTNESS because it measures other agent's quality of service (QoS) and only requires that the aggregate QoS be above a certain threshold. This creates a moral hazard wherein strategic agents will maintain reputations just above the threshold. Their mechanism does not meet FLEXIBILITY well, because it is not clear how to weight and aggregate QoS across domains of interaction.

Teacy et al. [2006], Jøsang [1998], and Huynh et al. [2006] present methods of aggregating trustworthiness from peers that can account for uncertainty. Kamvar et al. [2003] propose a self-policing peer-to-peer reputation system that is highly distributed. Like the work of Yu et al. above, the trust measurements and communications of these three works take into account neither the possibility of different domains nor of different utilities involved, thus violating FLEXIBILITY. For example, their methods do not account for whether an agent is trustworthy enough to deliver a single order of a million items if the agent was known previously to be trustworthy to deliver one item. Similarly, these methods assume agents have a specific type and always perform the same actions, at least on a probabilistic basis, regardless of the other agents and situations involved, thus violating ROBUSTNESS. Such an assumption can be reasonable when one agent is interacting with many anonymous agents, such as a company selling a particular brand of food, but often do not hold under nonanonymity when the agents are rational and fewer, or can precisely control their interactions with others.

Zacharia and Maes' [2000] mechanism seeks to achieve low-level behavioral goals, such as enabling agents with higher reputations to have more influence on others' beliefs. However, their subjective trustworthiness measures only weakly achieve AGGREGABILITY. Like the aforementioned trust and reputation systems, their measures are highly specific to the interaction domain, which does not meet FLEXIBILITY. Zacharia and Maes tested their system only against malicious agents that build up reputation and then spend it, and do not examine strategic agents, so we are unable to assess how well their system meets ROBUSTNESS.

Saha et al. [2003] support EVIDENTIALITY, because their method uses agents' reputations to directly evaluate the future expectations of utility that would be achieved by each possible interaction. However, their method does not meet AGGREGABILITY because agents cannot aggregate information from sources other than their own interactions. Saha et al.'s method is also potentially weak against ROBUSTNESS if agents can easily change identities and exploit favors offered to unknown agents. Further, their method does not meet VIABILITY because agents cannot communicate their knowledge.

Resnick and Sami [2007] focus on preventing various types of reputation manipulations, supporting ROBUSTNESS. Whereas their model appears to meet most of the rest of the desiderata, their model discards potentially useful information, partially conflicting with AGGREGABLE. This is particularly limiting in the case when information on a particular product or agent can change, and the system is slow to adapt because of the sudden increase in information entropy.

2.4.2 Discount Factor and Desiderata

Using an agent's discount factor as its trustworthiness meets EVIDENTIALITY, because each agent can measure others' discount factors and apply them in a direct manner to evaluate its optimal strategy.

Discount factor measurements meet AGGREGABILITY because they can be combined to increase accuracy and precision. The measurements consist of a range or PDF of another agent's discount factor, and can be combined via probability theory to yield further accuracy [Hazard, 2008]. The only difficulty with discount factor measurements is that the measuring agent must account for its best understanding of what the measured agent is experiencing, and must account for the measured agent's best response. Computing the best responses to find the

Bayes-Nash equilibria can be a hard computational problem [Conitzer and Sandholm, 2004]. In our model, we have found the computational complexity of some discount measurements to be relatively simple or readily approximable, such as when sellers are slowly dropping their prices in a market with more demand than supply. However, in other situations, such as when an agent is aggregating and deciding the validity of many conflicting reports about one agent from other agents, the computational complexity may be high, yielding a potential conflict with computational efficiency in VIABILITY. Further study is required to find the computational complexity for computing other agents' discount factors in various interactions and to determine whether efficient algorithms exist.

In general, agents would not demonstrate a discount factor lower than their actual unless they are competing with others for limited resources. Agents have difficulty credibly demonstrating discount factors above their own because their impatience prevents them from waiting for the postponed, larger utility. For these reasons, discount factors as trustworthiness measures are aligned with ROBUSTNESS. Further, discount factors are strongly influential in many different domains and situations, such as an agents' desire for quality, the rate at which sellers drop their prices, and how quickly agents come to an agreement in negotiation, discount factors. Whereas the exact method of measuring discount factors changes with the role and situation, discount factors as trustworthiness can maintain their strengths with other desiderata across these domains, regardless of domain-specific valuations and capabilities, thus satisfying FLEXIBILITY.

Discount factors' ability to cope with an open system facilitate PRIVACY ENHANCING in the sense that they offer a low barrier to entry and generally do not require external information to be revealed. If one agent knows nothing about another agent, the maximum entropy distribution of the other agent's discount factor is a uniform distribution on $[0, 1)$, which offers some protection against unknown agents as the expected discount factor is $\frac{1}{2}$. If an agent has a priori knowledge of the distribution of discount factors of agents to be encountered, it may use that distribution instead. If unfavorably typed agents repeatedly assume new identities to expunge poor reputations, or attempt to open a large number of pseudonymous accounts to bolster their own reputation (Sybil attacks), then an a priori distribution can be sufficiently pessimistic in a new agent's discount factor at the expense of how quickly an agent can recognize a new but favorably typed agent. Using discount factors as trustworthiness does not prevent implementations from requiring agents to reveal valuation information, and agents may have some ability to evaluate others' valuations. Therefore, discount factors do not maximize this desideratum, but do not directly violate it.

2.5 Discussion

Trustworthiness is objective because it means how deserving or worthy an agent is of trust. Of the properties of trust we introduced, *scalar* is merely a convention. *Strength* and *stability* are *absolute* aspects in that they pertain to an agent in itself, whereas *comparison* is *relative* in that it pertains to an agent in reference to other agents. Notice that even *comparison* indicates an objective measure, because it compares the trustworthiness of trustees from the perspective of a rational trustor.

Agents with low measured discount factors behave in ways that are generally considered untrustworthy. An agent with a low discount factor would produce poor quality items, exert low effort on service tasks, and not offer or return favors. In each case, the agent will prefer smaller utility gain now to a larger gain in the future. If an agent a with a low discount factor were entrusted with a secret by agent b, perhaps for mutual benefit, a would not have a strong incentive to keep the secret. Agent a would not put much value on its future relationship with b, and would reveal the secret to some third agent, c, if agent c offered a little short term gain. Having a low discount factor means an agent is myopic and impatient. Under our definitions and assumptions, trustworthiness is therefore roughly equated to patience.

Agents with high measured discount factors often behave in a trustworthy manner. However, the way discount factors as trustworthiness can depart from intuition is when an agent with a high discount factor faces a moral hazard where it does not expect sanctioning to be effective. The agent with the high discount factor would not necessarily be honest when it is not being observed. It is possible for an agent that steals items from other agents to have a high discount factor if the agent believes that the probability of being caught or the utility loss due to punishment will be sufficiently low. One scenario is the agent's beliefs are wrong and other agents observe the undesirable behavior, attributing the behavior to lower valuations or a lower discount factor. Conversely, if the agent's beliefs are accurate and other agents cannot differentiate an agent that is always altruistic (strongly typed) from an agent that is only altruistic when observed (purely utility maximizing), then no objective trust system could measure this.

The discount factor method requires each agent to model another agent's valuations in addition to its trustworthiness. This model affords the first agent an analytically predictive model of the second. Almost any trust model can be tailored to different domains and contexts, such as automobile repair and cooking. However, discount factors can model a single trustworthiness value across the domains, as long as sufficient information is available about the agent's valuations and capabilities (as defined by the value an agent will receive from another's action) in the different domains. This means even if an agent repairs automobiles well but cooks poorly, its trustworthiness can be consistent across the domains as long as the contexts are equivalent and the agent's valuations and beliefs can be modeled. Even if information is scarce, agents can have mutual information about the information scarcity and attribute nontrusting behaviors to the scarcity of information.

Using its expectation of another agent's valuations in decision models helps an agent evaluate the trustworthiness of agents in complex situations. Suppose b regularly purchases cheap office supplies from s, and always finds them to be of good quality. In this context, s is trustworthy. Because the profit margins on the items are small, b is only able to know that, for example, $\gamma_s > 0.9$. Now suppose a is looking to buy an expensive office chair. The discount factor that b reports may not indicate that s will sell a high-quality office chair in the different setting, depending on the possible profits. If s focuses on office supplies, it may not have the economies of scale to make larger profits on high-quality office chairs, increasing the incentive to provide one of low quality. Note that discount factors coupled with valuations also can work in the reverse; a supplier of expensive niche items may not be able to efficiently offer cheap bulk goods, and may not experience much sanctioning if it were to provide poor quality goods to an unknown single-transaction customer.

Agents' discount factors may change along with Assumption 1 due to various reasons. External factors include a change in the market or the agent's ownership, and internal factors include an agent deciding to leave a given market at a specified future date.

Practical examples of multiagent interactions involving many individuals, firms, and other organizations can exhibit a range of behaviors, including agents that act strategically, agents that behave in a consistent manner directed by a rigid set of beliefs, and agents that fall between the extremes. Discount factors offer a measurement of trustworthiness that is applicable to the range of agent behavior where both adverse selection and moral hazards exist.

Chapter 3

Favor Reciprocation With Private Discounting

3.1 Introduction

People and organizations routinely perform favors in a variety of settings built on norms, empathy, and trust. However, a self-interested agent acting on behalf of a person, organization, or itself, only has whatever intrinsic empathy and trust towards others with which it was designed. In many situations, the optimal outcome is sought by agents exchanging resources and services.

While market-based resource allocation is often an effective tool for social optimization [Wellman, 1996, Golle et al., 2001], markets may not be effective with self-interested agents without common currency, sufficient liquidity, means transferring resources, or effective methods to enforce a fiat currency. In such cases, agents can reciprocally perform favors instead of using markets to improve their own welfare and thus improve social welfare, albeit with generally less efficient outcomes. Agents that can trust one another to reciprocate favors to form a gift economy will have a better ability to smooth out inefficient allocations over time.

Many environments lend themselves to such favor-based interactions. One widely studied example is peer-to-peer file sharing as a decentralized means of distributing data and software [Kamvar et al., 2003, Banerjee et al., 2005]. Complex tasks in multi-robot systems often require coordination to increase utility [Gerkey and Matarić, 2002]. Favor-based mechanisms are particularly useful in situations where robots are self-interested but do not have resources to enforce trade. In business-to-business settings, agents involved in procurement can decide whether to put forth extra effort to deliver goods or services exceeding the contract when the customer is in need to foster the relationship for reciprocatively beneficial behavior. Personal assistant agents may engage in similar interactions, such as transferring reservations that offer more utility to a recipient who has a history of positive reciprocations. Gift economies may also be used to augment market-based transactions to circumvent market friction, such as a burdensome taxation system, or to increase the risk of switching to another provider with an unknown reputation.

In this chapter, we build a simple but applicable favor-based interaction model in which agents attempt to maximize their own utility based on their discount factor and what they expect to gain in the future. We employ the commonly used exponential discounting, net present value, although other decreasing discount models may be substituted. In our model, agents interact via a stochastic process, and can only choose whether to reduce their own utility in order to increase that of another. Our model exhibits individual rationality (expected positive utility by participating in the game) and in most cases attracts agents to approximate incentive compatibility (optimal strategy is to perform honestly).

We begin by presenting related work, followed by our stochastic interaction model, and then present our adaptive reputation method, which we then use to construct our primary result: a reciprocity general strategy that works in stochastic environments. We investigate how agent communication impacts the model and evaluate the strategies via simulation. We find analytically and experimentally that agents with discount factors that allow them to retaliate the most effectively in a tit-for-tat style equilibrium achieve the highest utility, which are not always the most patient agents. Finally, we draw some conclusions from our analysis and simulations about how communication affects agents' behavior in our model.

3.2 Related Work

Our work is similar to Sen's work [Saha et al., 2003, Sen, 2002] in that we build a reciprocity model on future expectations, but we allow for the discovery of private discount factors, observe possible ranges of responses rather than point values, and do not use randomization to communicate signals. Our model also resembles that of Buragohain et al. [2003] in the way we are using incentives to build trust in an environment with favors, but the primary differences are that their model has continuous interaction and does not deal with discount factors.

Many others have proposed various methods of quantifying and communicating reputation and trust. Sierra and Debenham [2005] describe an information theoretic model of trust where more information yields less uncertainty in decision making. Because they use ordinal preferences instead of utility, their model works well only if all items being negotiated about are of sufficient and comparable worth. Teacy et al. [2006] use a beta distributions to model positive and negative interactions, but do not take into account the magnitude of the interaction nor strategic behavior from agents. Along similar lines of trust measurement, Yu and Singh [2002] use Dempster-Shafer theory, which represents belief and plausibility in probabilistic terms, to model trust and reputation. However, like the work of Teacy et al. and Sierra and Debenham, this model does not deal with strategic behavior nor interactions of different ranges of utility, potentially leaving agents' utility unprotected against a strategic agent.

Hermoso et al. [2007] investigate the relationship between trust and organizations. While their trust metrics are somewhat ambiguous, their models are capable of handling an agent's varying reliability across different roles. Our model focuses on strengthening the rigor of trust measurements, but does not yet account for reliability across different roles.

Ramchurn et al. [2006] develop a finite-horizon negotiation mechanism based on repeated games. Because the agents are negotiating about the near future, their differing discount factors are implicitly accounted for in the negotiations, resembling the effects of our model. However, their model does not deal with agents that break promises, and thus needs exogenous enforcement.

Azoulay-Schwartz and Kraus [2004] present a favor reciprocity model of information exchange and use a punishing trigger strategy with forgivement. While their method of interaction and mechanisms resemble ours, they assume that discount factors are public and their punishment mechanism does not account for the effect on the opposing agent in relation to the cost of the punishment.

While explicitly dealing with the desiderata of incentive compatibility and individual rationality are generally considered important in game theory and auction literature, dealing with strategic behavior is more rare in the trust and reputation literature. Jurca and Faltings [2007] develop an exchange model where the client can sanction the provider if a refund is not given for a bad interaction. Their model achieves similar goals to ours, except that their model is built on a more complex refund-based interaction rather than simple reciprocity, and their model assumes discount factors are publicly known. While our model does not exhibit perfect incentive compatibility, we are able to leverage approximate incentive compatibility in equilibrium as an attractor to gain accuracy in modeling private discount factors.

3.3 Favor Model

In our favor model, each agent $a \in A$ encounters other agents in pair-wise interactions with two roles: offering a favor and asking a favor. Each round, agents are paired with other agent and possibly given a chance to act in both roles, depending on a stochastic process. The probability that agent a_1 encounters agent a_2 in the round as offer and ask roles, respectively, is $r_{a_1 \to a_2}$; this is the rate of interaction. Similarly, the probability that a_1 encounters a_2 as ask and offer roles is $r_{a_2 \to a_1}$.

When agents encounter one another, they play a game, Γ, chosen randomly from the set of possible game parameterizations, $\mathfrak{G}$, as follows. The agent in the ask role knows its willingness-to-pay for a particular favor, w_Γ, and asks the other agent for the favor. While the asking agent could choose not to ask for the favor, the strategy of asking a favor always dominates not asking the favor because asking incurs no cost, reveals nothing to the opponent, and the asker will either receive a favor or gain information about the other agent. When the asking agent asks for the favor, the cost of the offering agent to perform the favor, c_Γ, is revealed to both agents. The agent in the offering role then decides whether to provide the favor, P, or to reject the request, R. The values of c_Γ and w_Γ are drawn from the public, non-negative distributions of the random variables C and W respectively. Each agent may have a unique distribution, but we assume they all share the same distribution for clarity. For brevity, we will drop the Γ subscript for c and w when the subscript is obvious or irrelevant.

Each agent's type is comprised of its discount factor, previous observations of interactions with other agents, and information acquired from communication with other agents. As these attributes of an agent's type are private, agents must analyze other agent's actions and strategize about information revelation in order to maximize utility.

This repeated game has the obvious Nash equilibrium of offering agents always playing R. Given sufficient discount factors, the repeated game also has the multitude of Nash equilibria of trigger strategies, that is, playing some sequence of P and R but playing permanently R if the sequence is ever broken by either agent. However, more interesting equilibria and behaviors emerge when reputation is taken into account. An agent has no direct control over gains of its own utility, and is thus subject to the actions of other agents.

In order to deal with agents' changing preferences over time, we discount agents' histories by using a replacement process for the agents. When a replacement occurs, an agent is effectively removed and replaced with a new agent; its discount factor is redrawn from the distribution of discount factors. The agent's observations and information of other agents may be cleared when it is replaced. Agent replacements have been shown to be an effective tool for modeling how agents change over time [Mailath and Samuelson, 2006]. This replacement process forces agents to focus on recent observations more than old observations and allows agents to change over time. An agent's *replacement rate* is the expected number of replacements that the agent will undergo per unit time. We will denote agent a's replacement rate as λ_a, and use a Bernoulli process for replacement. Replacements can be justified by a change in the market, agent's ownership, information, or other factors in a dynamic environment. For the model to be meaningful to real-world scenarios, the replacement rate should be sufficiently low for reputation to be significant.

3.3.1 Strategies With Known Discount Factors

As an agent's discount factor increases, its willingness to give favors to other agents in return for greater reward later increases. An agent with a high discount factor therefore would desire a mechanism that rewards its patience and prevents other agents with lower discount factors from taking advantage of it. Under such a mechanism, agents will reciprocate favors according to discount factors of their own and of the opposite player.

Suppose agents a_1 and a_2 are exchanging favors, where a_1 is offering all favors to a_2 that cost a_1 less than a_1's current maximal favor offering, $\overline{c}_{a_1 \to a_2}$. Agent a_1 would choose which favors to perform, and also control its cost, by adjusting $\overline{c}_{a_1 \to a_2}$. We can think of the value $\overline{c}_{a_1 \to a_2}$ as how much a_1 trusts a_2 to repay the favors. Similarly, a_2 is offering a_1 all favors that cost a_2 less than $\overline{c}_{a_2 \to a_1}$. As the only positive payoff to a_1 is controlled by a_2 and a_1 incurs cost for providing favors to a_2, a_1 has a direct incentive to reduce its costs by refusing to provide favors. When a_1 is playing in an offer role in game Γ, we can thus write a_1's expected total future utility, U_{a_1}, of interacting with a_2 discounted by γ_{a_1} for each time step, t, given the cost of the current favor c_Γ as

$$U_{a_1} = \sum_{t=1}^{\infty} \gamma_{a_1}^t r_{a_2 \to a_1} PE(W|C < \overline{c}_{a_2 \to a_1}) - c_\Gamma - \sum_{t=1}^{\infty} \gamma_{a_1}^t r_{a_1 \to a_2} PE(C|C < \overline{c}_{a_1 \to a_2}) \ . \quad (3.1)$$

We define the shorthand notation $PE(Y|X) \equiv P(X) \cdot E(Y|X)$ with $P(X)$ being the probability of event X occurring and $E(Y|X)$ is the expected value of Y given that X occurred. Equation 3.1 may be easily extended to a total utility by a summation over all agents.

One primary criterion of an effective interaction mechanism is *individual rationality*, that is, an agent a_1 will expect to gain utility by participating, expressed formally as $U_{a_1} > 0$. If an a_1 will not expect to gain from a relationship of exchanging favors, the agent will not enter the relationship, setting $\overline{c}_{a_1 \to a_2} = 0$. By applying this principle to a given pair of agents, we can find the maximum $\overline{c}$ for which each agent would be willing to provide a favor to the other while keeping the utility non-negative. We can find these values by simply setting $U_{a_1} = U_{a_2} = 0$, setting c_Γ to the corresponding $\overline{c}$ that the agent pays out in each equation, and solving to find $\overline{c}_{a_1 \to a_2}$ and $\overline{c}_{a_2 \to a_1}$.[1] By assuming the worst case of $c_\Gamma = \overline{c}$ for each agent, all smaller values of c_Γ will yield positive expected utility and thus meet the criterion of individual rationality.

Our favor model exhibits a moral hazard, that is, each agent can directly increase its utility by reducing its corresponding $\overline{c}$. For many parameterizations, our favor model is closely related to the repeated prisoner's dilemma; the Pareto efficient outcome, $\forall a, a' \in A : \overline{c}_{a \to a'} = \infty$, is

[1] An agent a could also have some minimum utility threshold, $k_a > 0$, which it must receive in expected utility gain in order for it to participate, particularly if a relationship requires some opportunity cost, overhead cost, or communication to simply maintain. In this case, $U_a = k_a$ would be solved instead.

not necessarily a Nash equilibrium because agents have incentive to defect from this strategy. However, Pareto efficient outcomes can be achieved as a subgame-perfect Nash equilibrium if agents can credibly punish others for not offering favors. One such example is the *grim trigger* strategy, where agent a permanently sets $\overline{c}_{a \to a'} = 0$ if agent a' does not offer a favor. If significantly many agents use the grim trigger, then agents with sufficiently large discount factors are reluctant to ever not offer a favor, bringing about the Pareto efficient outcome. However, the grim trigger strategy is ineffective unless agents will credibly commit to it and can be extremely pessimistic when agents and circumstances may change [Axelrod, 2000].

Tit-for-tat strategies are similar to grim trigger, except that the punishment is not as long lived and agents eventually forgive others. Generally, tit-for-tat entails one agent punishing a defecting agent at least as much as the utility that the offending agent gained by refusing to provide a favor. If a_2 gains some utility x by reducing $\overline{c}_{a_2 \to a_1}$ and bringing a_1's utility down, then a_1 would like to bring a_2's utility down by at least x by reducing by reducing $\overline{c}_{a_1 \to a_2}$. Theorem 2 shows that in our favor model, if a relationship meets the criterion of individual rationality, then a trigger strategy can be credibly employed to yield a perfect Bayesian equilibrium.

Theorem 2 *If agents a_1 and a_2 are engaged in a relationship where each agent fulfills favors of some set of costs, $\overline{C}_{a_1}$ and $\overline{C}_{a_2}$ respectively, and the relationship meets the individually rational criterion for each agent for the supremum of $\overline{C}_{a_1}$ and $\overline{C}_{a_2}$, then the agents' relationship of fulfilling the favors of $\overline{C}_{a_1}$ and $\overline{C}_{a_2}$ permits trigger strategies as perfect Bayesian equilibria.*

Proof 2 *Rewriting Equation 3.1 in terms of gains from favors given by the other agent, $G \geq 0$, losses from providing favors to the other agent, $L \geq 0$, and cost of the current game, $c_\Gamma \geq 0$, an agent's utility becomes $U = G - L - c_\Gamma$. Fulfilling the maximum value of $c_\Gamma \in \overline{C}_{a_1}$ for a_1 meets individual rationality yielding $0 < G_{a_1} - L_{a_1} - \sup \overline{C}_{a_1}$. Because only a_2 can control the value of G_{a_1}, a_2 can use a trigger strategy to reduce G_{a_1} to 0 if a_1 deviates from the strategy of offering all favors of values in $\overline{C}_{a_1}$. As $G_{a_1} > L_{a_1} + \sup \overline{C}_{a_1}$, a_2's trigger strategy would nullify any utility gain that a_1 had obtained by deviating. The same logic holds for a_1 controlling G_{a_2}. Therefore, because no agent can increase its expected utility by changing its strategy, and because the repeated game is a stationary process, the trigger strategy is a perfect Bayesian equilibrium.*

The rate of change of a_2's utility, U_{a_2}, with respect to a_2's rate of change of $\overline{c}_{a_2 \to a_1}$ can be written as the partial derivative $\frac{\partial U_{a_2}}{\partial \overline{c}_{a_2 \to a_1}}$. Because an agent's utility increases when its costs are reduced, this partial derivative is always negative. Agent a_1 is also able to affect a_2's utility; changing $\overline{c}_{a_1 \to a_2}$ increases a_2's utility by the rate of $\frac{\partial U_{a_2}}{\partial \overline{c}_{a_1 \to a_2}}$.

In steady-state, such as by trigger strategies as in Theorem 2, agents will maintain constant values of $\overline{c}$. If an agent makes a small change to $\overline{c}$, its opponent can retaliate the same amount to nullify the gain. However, retaliating may be particularly costly to the agent performing the retaliation, particularly if the agent has a lower discount factor. We can express an equilibrium where agents have equal control over each others' utility as

$$\frac{\partial U_{a_1}}{\partial \overline{c}_{a_2 \to a_1}} = -\frac{\partial U_{a_1}}{\partial \overline{c}_{a_1 \to a_2}} \text{ and} \tag{3.2}$$

$$\frac{\partial U_{a_2}}{\partial \overline{c}_{a_1 \to a_2}} = -\frac{\partial U_{a_2}}{\partial \overline{c}_{a_2 \to a_1}} \ . \tag{3.3}$$

The right hand side of each equation is negative because decreasing one's own $\overline{c}$ increases one's own utility. Except when the rates of encounter, r, or the maximum favor willing to be offered, $\overline{c}$, are extremely different between the agents, a pair of discount factors and $\overline{c}$ values exists that satisfies this equality. Theorem 3 shows that there is always a discount factor that satisfies Equations 3.2 and 3.3 given that the rates of encounter are equal.

Theorem 3 *Given agents a_1 and a_2 with equal encounter rates, $r_{a_1 \to a_2} = r_{a_2 \to a_1}$, and equal maximum favor limits, $\overline{c}_{a_1 \to a_2} = \overline{c}_{a_2 \to a_1}$, there exists a discount factor for each agent where each agent can reduce its opponent's utility at the same rate as its opponent can increase its own utility.*

Proof 3 *Given the probability density function (PDF) of C, $f_C(\cdot)$, $\frac{\partial U_{a_2}}{\partial \overline{c}_{a_1 \to a_2}} = \frac{\gamma_{a_2}}{1-\gamma_{a_2}} r_{a_1 \to a_2} E(W) f_C(\overline{c}_{a_1 \to a_2})$ and $\frac{\partial U_{a_2}}{\partial \overline{c}_{a_2 \to a_1}} = -1 - \frac{\gamma_{a_2}}{1-\gamma_{a_2}} r_{a_2 \to a_1} \overline{c}_{a_2 \to a_1} \cdot f_C(\overline{c}_{a_2 \to a_1})$. $\lim_{\gamma_{a_2} \to 0} \left\{ \frac{\partial U_{a_2}}{\partial \overline{c}_{a_1 \to a_2}} < -\frac{\partial U_{a_2}}{\partial \overline{c}_{a_2 \to a_1}} \right\}$ because all terms become 0 except for the 1, leaving $0 < 1$. $\lim_{\gamma_{a_2} \to 1} \left\{ \frac{\partial U_{a_2}}{\partial \overline{c}_{a_1 \to a_2}} > -\frac{\partial U_{a_2}}{\partial \overline{c}_{a_2 \to a_1}} \right\}$ because the constant 1 becomes irrelevant and the remaining values can be divided off leaving $E(W) > \overline{c}$, which was assumed in the problem for individual rationality. Because $\frac{1-\gamma_{a_2}}{\gamma_{a_2}}$ is continuous and differentiable over the interval of $[0, 1)$, by the intermediate value theorem, there exists a γ_{a_2} that satisfies the equality $\frac{\partial U_{a_2}}{\partial \overline{c}_{a_1 \to a_2}} = -\frac{\partial U_{a_2}}{\partial \overline{c}_{a_2 \to a_1}}$. The same derivation holds for the opposite agent.*

When Equations 3.2 and 3.3 hold, the two agents can easily nullify their opponents' utility gains by applying a small change to their own $\overline{c}$ values. This further strengthens the agents' incentives not to deviate from their current values of $\overline{c}$, even in the absence of a publicly known trigger strategy. Because this, we expect the most effective cooperation would occur with agents that have the most appropriate discount factor for the given parameterization, that is, discount factors that satisfy Equations 3.2 and 3.3 using the most probable values of $\overline{c}$ because they have the most leverage to affect their opponents. We will revisit this notion when discussing the simulation results.

3.4 Modeling Reputation

We denote other agents' reputations from the vantage of agent a_1 as a set including a_1's direct observations combined with the information communicated to a_1 by other agents as $\mathcal{I}_{a_1}$. An observation, $i \in \mathcal{I}_{a_1}$, consists of the tuple $(o_i, o'_i, t_i, \gamma^*_i)$, where o_i is the agent that made the observation, o'_i is the agent the observation is made about, t_i is the time of the observation, and γ^*_i is the observation range. We define an observation range as a tuple of the upper and lower bound of the discount factor, γ_i, given an observation of the action the agent took in a given game. For the purposes of this section, we assume that agents' observations are accurate.

New observations can increase the precision and accuracy of a reputation estimate or alternatively invalidate previous observations if they are conflicting. Conflicting observations typically occur because an agent has undergone replacement, but may also occur due to an agent finding out that some observations communicated by other agents were incorrect. Given a set of observations, $\mathcal{I}$, and a new observation, i', we define the function $\mathcal{X}$, which returns the set of all observations in $\mathcal{I}$ which conflict with i', as

$$\mathcal{X}(\mathcal{I}, i') = \{i \in \mathcal{I} : o_i = o_{i'}, o'_i = o'_{i'}, \gamma^*_i \cap \gamma^*_{i'} = \emptyset\} \ . \tag{3.4}$$

When agent a_1 makes a new direct observation of a_2, i', we denote the resulting relevant history of observations as $\mathcal{I}_{a_1} \oplus i'$. We define this operation of accommodating a new observation as

$$\mathcal{I}_{a_1} \oplus i' = \begin{cases} \mathcal{I}_{a_1} \cup \{i'\} & \text{iff } \mathcal{X}(\mathcal{I}_{a_1}, i') = \emptyset, \\ \{i \in \mathcal{I}_{a_1} : t_{i'} \geq \max_{j \in \mathcal{X}(\mathcal{I}_{a_1}, i')} t_j\} \cup \{i'\} \ . \end{cases} \tag{3.5}$$

If agents' strategies are consistent and prevent conflicting observations, we can denote the expected number of direct observations of a_2 between replacements by a_1 as $E(|\{i \in \mathcal{I}_{a_1} : o_i = a_2\}|)$. This value is maximized just prior to replacement and 0 after the replacement. By using the equality $\sum_{t=0}^{\infty}(1-\lambda_{a_2})^t = \frac{1}{\lambda_{a_2}}$, the expression becomes

$$E(|\{i \in \mathcal{I}_{a_1} : o_i = a_2\}|) = \frac{r_{a_2 \to a_1}}{2\lambda_{a_2}} \ . \tag{3.6}$$

As the replacement rate drops to 0 and the observation history length becomes infinite, the only conflicting observations that would occur would be due to agent strategies such as misinformation or collusion. All else being equal, lower replacement rates would not decrease the expected number of *relevant observations*, that is, observations that occur at or after the last conflicting observation. Given an arbitrary additional observation, i', some replacement rate x, and a small change in replacement rate, ϵ, this can be expressed as

$$E(|\{i \in \mathcal{I}_{a_1} : o_i = a_2\} \oplus i'| \quad |\lambda_{a_2} = x) \geq E(|\{i \in \mathcal{I}_{a_1} : o_i = a_2\} \oplus i'| \quad |\lambda_{a_2} = x + \epsilon) \ . \quad (3.7)$$

With a longer relevant observation history, a conflicting observation probabilistically makes more of the observation history irrelevant. If agents bias reputation toward a poor reputation when few or no relevant observations are available, then this mechanism translates into more difficulty gaining reputation than losing it. In such a system, agents would have increased value for maintaining a positive reputation because of this bias.

3.4.1 Agents' Utilities Under Incomplete Information

To denote the results from simultaneously solving the aforementioned constraint of individual rationality for the maximum allowable $\bar{c}$, $U_{a_1} = U_{a_2} = 0$, we introduce a function, $\gamma^*_{\text{offer}} : \mathfrak{G} \times \{P, R\} \rightarrow ([0, 1) \times [0, 1))$, that solves these simultaneous equalities to find the range of possible discount factors for the offering agent.[2] This function returns the observation of the discount factor for the offering agent based on the action performed by the offering agent, either P or R, and the parameters on the game, $\Gamma \in \mathfrak{G}$, where $\mathfrak{G}$ is the set of all possible games that the agent could play. Each game consists of the favor cost c_Γ and the agents involved. The range returned for γ_{offer} will be one of $[\gamma, 1)$ or $[0, \gamma]$, depending on whether the offering agent played P or R, respectively. We will refer to this range as γ^*_i for the outcome of game Γ_i. Further, we introduce the function $c^* : [0, 1) \times [0, 1) \rightarrow \mathbb{R}$, which takes in discount factors (or estimations thereof) of two agents, γ_{a_1} and γ_{a_2}, and returns the maximum value of a favor the first agent would offer to a_2, $\bar{c}_{a_1 \rightarrow a_2}$, under individual rationality for the parameters given. We use the shorthand notation $c^*_{a_1 \rightarrow a_2}$ to indicate $c^*(E(\gamma|\{i \in \mathcal{I} : o'_i = a_1, o_i = a_2\}, E(\gamma|\{i \in \mathcal{I} : o'_i = a_2, o_i = a_1\}))$. Both functions γ^*_{offer} and c^* are used as inputs to Equation 3.1 when the corresponding values are unknown.

An agent's expected utility in the current game for a given action is the sum of the value obtained by the action plus the expected future value, V, of all future games given the agent's reputation after having performed the action. In order to find V, an agent a_1 must evaluate other agents' discount factors from the observations, $\mathcal{I}_{a_1}$, it has made. When a_1 is playing in the

[2]While we only focus on two-agent interactions, a system of equations of total utility may be used to include all agents with a separate equation for each agent within the connected components of the communication graph. These connected components would include the agents of interest, all other agents that those agents trust sufficiently to accept communicated observations, all agents trusted by those agents to accept an observation, etc., so all agents may be included. These equations are evaluated from an individual agent's perspective, so the values are accurate only to the accuracy of the agent's information. While these systems of equations can be difficult or impossible to solve in closed form, numerical methods such as multivariate secant or bisection are effective.

offer role, given that a_2 has chosen to ask the favor by playing P, we can rewrite a_1's expected discounted utility from Equation 3.1 relative to its discount factor for the current game, Γ, as a function of a_1's action, $s \in \{P, R\}$, as

$$U_{a_1}(s) = V_{a_1}(\mathcal{I}_{a_1} \oplus (a_1, a_2, t, \gamma^*_{\text{offer}}(\Gamma, s))) - \delta_{s,P} \cdot c_\Gamma \ , \tag{3.8}$$

where the result of function V yields the future value of a_1's reputation given that a_2 will observe a_1 playing s in a game with the value of c_Γ. If a_1 plays P, then its utility will be reduced by c_Γ.[3]

Each observation i loses potency as elapsed time increases since the observation was made, with loss rate based on the replacement rate of the agent observed. The uniform distribution satisfies the principal of maximum entropy given the maximum and minimum value of an observation, meaning that given only the information that the actual value is between two endpoints, the maximum likelihood distribution is uniform. Immediately after an observation, the discount factor must lie within the range observed, meaning that the probability the discount factor is outside of that range is 0. After infinite time has passed, the observation becomes irrelevant and a uniform distribution of belief is resumed. We find the value of the probability density function (PDF) of an agent's discount factor γ, given a single observation i and current time T, discounted by the replacement rate as

$$f_i(T, \gamma) = \begin{cases} \frac{1-(1-\lambda^{T-t_i})(1-(\sup \gamma^*_i - \inf \gamma^*_i))}{\sup \gamma^*_i - \inf \gamma^*_i} & \text{if } \gamma \in \gamma^*_i \, , \\ 1 - \lambda^{T-t_i} & \text{if } \gamma \notin \gamma^*_i \ . \end{cases} \tag{3.9}$$

We can then use Bayesian inference to combine the PDFs of the relevant observations to find what a given agent will expect another agent to believe of its discount factor, $E(\gamma | T, \mathcal{I})$, as

$$E(\gamma | T, \mathcal{I}) = \int_0^1 x \frac{\prod_{i \in \mathcal{I}} f_i(T, x)}{\int_0^1 \prod_{i \in \mathcal{I}} f_i(T, y) dy} dx \ .^4 \tag{3.10}$$

To find the total future utility for a given reputation, an agent needs to determine its expected gain from encounters with every agent. By combining relevant observations, finding the corresponding maximum favor value c^* using the expected discount factor via Equation 3.10 for each situation, and using the results in the manner of Equation 3.1, we find

$$V_{a_1}(\mathcal{I}) = \frac{\gamma_{a_1}}{1 - \gamma_{a_1}} \sum_{a \in A} \left(r_{a \to a_1} P E(W | C < c^*_{a \to a_1}) - r_{a_1 \to a} P E(C | C < c^*_{a_1 \to a}) \right) \ . \tag{3.11}$$

[3]The Kronecker delta, $\delta_{i,j}$, yields 1 if $i = j$, 0 otherwise.

[4]We derive and discuss this type of aggregation in further detail in Sections 4.5.1 and 5.2.

In addition to using observation and communication to evaluate others' discount factors, agents can also strategize how to influence others' perception of their own discount factor. An agent would prefer to have other agents overestimate its own discount factor, as the agent could take advantage of other agents that are willing to give up short-term gains for larger long-term gains. Similarly, agents do not want others to underestimate their own discount factor, because they would be missing gains for which they would be willing to reciprocate favors.

Despite the incentive to convince others of an artificially high reputation, our model is approximately incentive compatible in the steady state because the cost to convince another agent of a better reputation is more than the expected future gain. Incentive compatibility is important because without it, agents cannot accurately deduce discount factors from other agents that strategically give larger favors than they should. While our model does not ensure incentive compatibility, it generally ensures that the region of incentive compatible state space is an attractor. When agents are not in a region of the state space that is incentive compatible, agents are incentivized to correct others' beliefs.

The three exceptions where our model is not approximately incentive compatible are as follows. First, while agents prefer opponents to overestimate their discount factors, spending utility to inflate reputation costs more than an agent will receive from the inflated reputation. However, if an agent already has a stronger reputation than its discount factor, then the agent is incentivized to use the reputation and play R for games below $\overline{c}$ while obtaining favors from the other agent. Second, if an agent's reputation is much lower than its actual discount factor, then the agent may not always offer small favors for small values of $\overline{c}$ because they may not sufficiently increase reputation to be worthwhile. While not incentive compatible, these two cases correct other agents' beliefs.

The third exception to incentive compatibility is if agents' discount factors are high and the value received from a favor is disproportionately large relative to the cost of offering a favor. In this case, an agent may know that its opponent's high discount factor will prevent a decrease in reputation from dramatically decreasing utility because its opponent's expectations of the future far outweighing the utility lost. In these cases, agents with the highest discount factors may not necessarily end up with the highest utility.

3.4.2 Communicating Reputation Information

Agents have a variety of ways to communicate information about other agents' reputations. An agent could send another agent its entire list of observations for a given agent, or alternatively just an estimate of a discount factor. While supplying more detailed information can be more helpful to the recipient, more degrees of freedom of this information make it more complex for the recipient to evaluate whether the information is accurate and truthful, particularly if other agents are colluding. Because agents can gain utility when others overestimate their discount factors, a self-interested agent may be reluctant to divulge extra information that might reduce another agent's belief of its discount factor.

Our communication model offers similar effects to that of the model proposed by Procaccia et al. [2007], although their method uses randomization to communicate reputation instead of ranges of discount factors. Agents maintain observations and communications which are used in aggregation to evaluate each other and give future recommendations.

We focus on the following simple yet plausible forms of communication. Suppose agent a_1 asks a_3 a question about whether a_2 is trustworthy. Agent a_3 can answer yes, no, or refuse to answer. Similarly, a_1 can choose to use or ignore a_3's advice and may solicit advice from other agents.

Agent a_1 could ask a_3 whether a_3 would provide a favor to a_2 for the current game or whether a_3 would recommend a_1 provide a favor to a_2. These two questions can yield different answers. For the former question, a_3 obviously has more information about itself, but a_1 might not have much information about a_3. If a_1 does not have much information about a_3, then a_1 should not ask a_3 because a_1 cannot effectively evaluate a_3's answer. For the latter question, a_3 has less information about a_1, and could be punished for giving a bad recommendation only for having insufficient information about a_1.

Because a_1 knows what information it has about a_3 better than a_1 knows what information a_3 has on a_1, asking the question of whether a_3 would provide a favor to a_2 will provide a_1 with the most reliable information. In choosing this question, a_1 is incentivized to ask advice from agents it knows the most about, but may also choose to ask advice from other agents for the purpose of learning about them. Agent a_1 is further incentivized to ask advice from agents with similar discount factors, because their answers would be similar. However, if a_3's discount factor is different from a_1's and a_1 has some knowledge about a_3's discount factor, then a_1 can still use a_3's advice to learn more about a_2.

When a_1 asks a_3 whether a_3 would offer a_2 the favor in the current game, a_1 would like to make an observation about a_3 in addition to the observation of a_2. Agent a_1 can combine a_3's advice with a_2's action and utilize future observations of a_2 to more accurately reevaluate the observation made about a_3's answer. If a_3 refuses to answer when asked, a_1 can assume a_3 is not confident in its information about a_2, and cannot make an observation. If, on the other hand, a_3 answers, a_1 will judge a_3 based on the recommendation either positively or negatively.

Now we examine the implications of how a_3's recommendation will affect a_1's perception of a_3. Suppose a_3 answered it would play P in the queried game. In this case, a_1 will compute the observation of a_2 as if it were from a_3's perspective using its belief of a_3's discount factor. Later, a_1 may reevaluate this observation when deciding to offer a favor to a_3 by using a_1's current knowledge of a_2's discount factor and the parameters to the game in which a_3 had given its recommendation. Agent a_1 will believe that a_3's recommendation is accurate if a_1 believes that a_2's discount factor is within the bounds of a_1's observation of a_3's recommendation. Similarly, if a_3 recommends R to a_1 and a_1 later finds out that a_2 had had a low discount factor, then a_3's reputation will be increased by the P observation in a_1's hypothetical game between a_2 and a_3.

Suppose a_3 answered it would play R with a_2 and a_1 finds out later that a_2 had had a high discount factor. Agent a_1's interpretation of the observation of a_3's recommendation would be that a_3 has a discount factor below that required to play P with a_1, yielding an upper bound on a_3's discount factor.

Finally, the most complex case is when a_3 answers it would play P with a_2, but a_1 later finds out that a_2 had had a low discount factor. If a_1 finds that a_2's discount factor is low and does not return favors, then a_1 can reason that a_3 gave the answer P so that a_1 would increase its expected value of a_3's discount factor. This answer would indicate that a_3's discount factor is so low that it was not concerned with a_1 other than extracting a favor that it would not have to repay. Because a_3's discount factor cannot be measured with respect to this false information for the hypothetical game, and because this false answer indicates that a_3's discount factor is arbitrarily low, a_1 would be prudent to throw away its previous observations of a_3 to reset its expected discount factor to a low value.

Note that these hypothetic observations from recommendations should be observed as when the recommendation was given, not at the time of computation. If a_1 believes a_2 underwent replacement since the observation of a_3's recommendation, but a_3 has not undergone replacement, a_1 should only use its observation history about a_2 from before a_2's replacement when computing a_3's expected discount factor.

The number of observations from communication can scale up with the number of agents as $|A|^3$ because each agent can communicate with all others about all others. However, in scale-free networks and other network topologies found in real-world applications, relevant communication usually scales under that bound. An agent can use various techniques to determine which agents to ask advice. A simple technique is for an agent to ask other agents that it has interacted with at least a minimum number of times; we use this in our simulations. For agents with low replacement rates, more complex evaluations may be used, such as asking agents with similar discount factors, or agents whose answers are believed to offer the maximum amount of information entropy about a given agent.

When an agent is reevaluating the impact of observations from recommendations on other observations from recommendations, the order of evaluation will have some impact on the results. For example, if a_1 accepts conflicting recommendations from a_3 and a_4 about a_2, a_3

or a_4 may be incorrectly punished depending on the order of evaluation. This impact can be minimized by reevaluating observations using the agent's most recent knowledge. Because agents need not be truthful, an agent does not know which ordering is best. Reasonable solutions include evaluating observations in chronological order or minimizing conflicting observations. We use the chronological ordering in our simulations.

3.5 Simulation Results

We conducted simulations to assess our model's affects. For the simulations without communication, we used 32 agents and ran each experiment with 100 rounds. For simulations with communication, we used 16 agents and 50 rounds (due to the more significant simulation time). Random numbers were only used to set up the games and agents, so using the same seed with different algorithms provided the same set of games. While we examined the behaviors across randomized results, for the graphs in this section we used the same seed for randomizing the games across variations to remove visible noise and make the results easier to see. The replacement rate was chosen uniformly to be .02 to reflect a mean expected agent life of 50 rounds. The choice of the replacement rate did not appear to affect the shape of the results provided it was sufficiently low. Globally high replacement rates remove the effectiveness of a reputation system because agents do not live long enough in the system to develop a reputation.

We chose the uniform distribution C on $[0, 200]$, the exponential distribution with $E(C) = 100$, and various values for W because they offered a wide range of behavior even when agents have full knowledge of each others' discount factors. From further experiments of different parameterizations and distributions, the behaviors discussed in this section appear typical.

We ran the experiments with two topologies for agents encounters. The first is a uniform topology, such that each $r_{a_1 \to a_2}$ was chosen from a uniform distribution on $[0, 1]$. The second topology is a scale-free network where agents commonly encounter a small set of agents and occasionally encounter an outside agent. To construct this network, we begin with two agents having corresponding encounter rates set to 1, and add agents individually, randomizing the new agent's rate of encounter with every other agent proportional to the sum of the other agent's rate of encounters relative to the sum of all encounter rates.

To determine the effect of the distribution of agents' discount factors, we examined four distributions: uniform, all the same, 4^{th} root of uniform distribution, and 1 minus the 4^{th} root of uniform distribution. The uniform distribution allowed us to see effects in a diverse population, whereas the second test made sure that the behavior was consistent when the agents all had the same discount factor. The the fourth root of uniform distributions, chosen to give a population of agents with generally high and low discount factors, are biased toward high

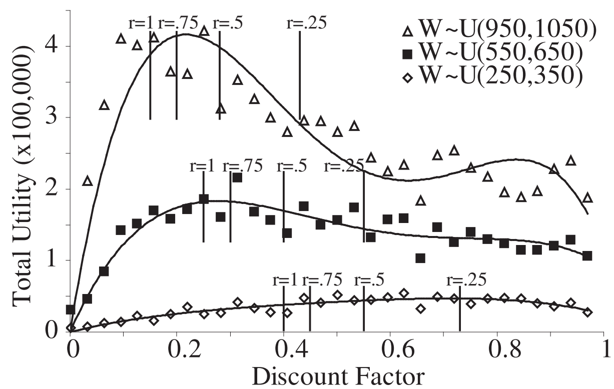


Figure 3.1: Typical simulation results of 32 agents each with a uniform distribution of encounter rates.

or low discount factors respectively, but with one agent that had an opposite discount factor.

Figure 3.1 shows a small but indicative subset of our results of agents' performance given different distributions of W using a uniform topology and uniform distribution of C. Each point represents an agent's final utility, and the trend lines are depicted by the best fit quartic polynomial. We note that the figure shows the total non-discounted utility of the agents so that they can be directly compared. Simulations with higher expected values of W obviously have higher final utilities, but the interesting feature is how the group of agents with discount factors that yield the highest final utility change with respect to W.

The vertical lines near each simulation set represent the discount factor that satisfies Equations 3.2 and 3.3 with the infimum of C, 0, solved for symmetric rates of encounter of $r \in \{1, .75, .5, .25\}$. Each of these lines represents the discount factor where two opposing agents can equally balance off their retaliations to each other, forming an even tit-for-tat strategy for the given symmetrical rate of encounter. Because we are setting $r = r_{a_1 \to a_2} = r_{a_2 \to a_1}$ in these derivations, the results are approximations to the actual interactions, which do not typically have symmetric rates.

The results of Equations 3.2 and 3.3 give an intuition of which regions of discount factors will receive the highest payoffs due to the ease and credibility of sanctioning. Agents with discount factors below this region will abstain from offering large favors to other agents because of their impatience and thus agents with higher discount factors will refuse large favors to them, lowering their utility. When agents with discount factors above this region attempt to sanction agents with lower discount factors, their sanctions must be strong because of the relative impatience of the other agents. Given the high utilities for receiving a favor and the low cost of providing a favor, the magnitude of these sanctions outweigh the benefits they would have gotten if they

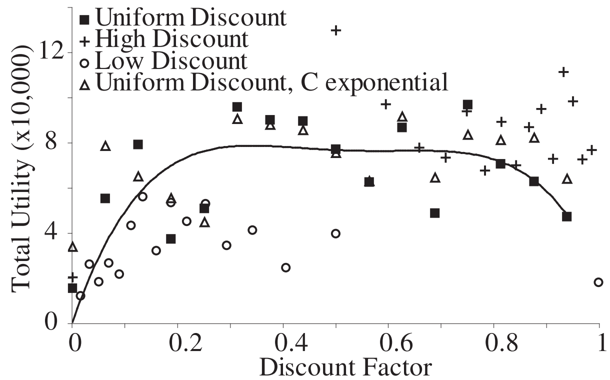


Figure 3.2: Various simulation results of 16 agents for 50 time steps each with a uniform distribution of encounter rates.

had not sanctioned. A social analogy for the lower utility of agents with high discount factors in these situations would be a stubbornly pedantic individual being ostracized by his peers because his peers do not find additional utility from the pedantry.

Figure 3.2 depicts variations with $W \sim U(550, 650)$ for 16 agents with 50 time steps. The points labeled Uniform Discount and the corresponding trend line are the same parameterization as in Fig. 3.1 as a reference point. The two 4^{th} root distributions (labeled High Discount and Low Discount) show results when agents have high or low discount factors. Groups of agents with high discount factors outperform the uniform distribution and groups with low discount factors underperform the uniform distribution. The trends from the distributions matched up regardless of the topology or communication; throughout all our data, the trend is that agents with discount factors that allow them to equally retaliate achieving higher payoffs, which may not necessarily be the highest discount factors. The results from the scale-free distribution (not depicted) were similar to that of the uniform distribution with the only notable differences being an increased variance and mean in payoffs. The results from using communication (not depicted) were that agents with low discount factors tended to have 10-30% lower payoffs. Agents with high discount factors were less affected, although some attained higher and lower utilities than without communication.

3.6 Stochastic Discrete Favor Reciprocation

In this section, we outline investigations of solution concepts for a pair of agents entering a relationship with a stochastic process of favor opportunities, where the favors are chosen randomly.

From Equation 3.1, agent a's expected future benefit from agent b is $\frac{\gamma_a}{1-\gamma_a} r_{b \to a} P E(W | C < \overline{c}_{b \to a})$, and its expected future cost is similarly $\frac{\gamma_a}{1-\gamma_a} r_{a \to b} P E(C | C < \overline{c}_{a \to b})$. One drawback of the favor model is that agents can enter an equilibrium where they must provide costly favors even if they offer little benefit to the other agent. This behavior is due to the equilibrium requiring each agent offering all favors below a certain cost. Here, we relax this assumption. In doing so, the agent now evaluates its own benefit of each potential favor before deciding whether it should ask the favor in the first place.

For every possible cost to b for providing a favor to a, there exists some benefit to a where a would ask b the favor, and some lesser benefit (possibly a cost, a negative benefit) where a would not ask b to perform the favor. Agent a may decide not to ask a favor because the cost is a burden to the b, and if the b fulfills the favor, then b will have committed more utility to a's benefit. To maintain a mutually beneficial relationship, a would need to provide more favors to b to make up for b's cost, otherwise b would need to reduce the favors it provides to a. If b reduced the favors to a, then a may also be forced to reduce its offerings, and may end up costing b more utility than b gained. If neither a nor b would benefit from changing their favor process, then the system is in a Nash equilibrium.

3.6.1 Discrete Favors

We now consider the case where an agent offers a favor to another occurs via a stochastic process, with the set of favors containing a finite number of favors, each with an expected value. Each favor performed costs the offerer and benefits the asker. An agent can only gain utility by favors provided by others.

The set of favors, including benefits and costs, as well as agents' discount factors are all common knowledge. We define a *favor relationship* between two agents, a, and b, as the sets of favors that each will always offer to the other, ξ_a and ξ_b respectively, when the favor in the set comes up due to the stochastic favor opportunity process. These sets of favors are subsets of the sets of possible favors, Ξ_a and Ξ_b.

The process of favor opportunities is stochastic with a discrete time step. Each favor, i, has a rate, r_i, which indicates the probability that the favor will arise during the given timestep. Agents' decisions of whether to offer a favor occur one at a time; the game is asynchronous. A *favor relationship* is a steady state equilibrium, a solution concept where each agent involved will always provide all of the favors in its set and assumes that the other agent does as well. If an agent fails to provide a favor, then the equilibrium will be broken. Following the grim trigger strategy, the other agent will terminate the favor relationship.[5]

[5] A less harsh approach might be that the agent lowers the probability of the favor being offered, for example, if there is noise in the signal of whether a favor was provided. In no noise is assumed, another possibility is that

Table 3.1: Example set of favors.

Favor Name	A utility	B utility	rate
$A1$	-1	3	0.2
$A2$	-2	5	0.7
$B1$	3	-1	0.2
$B2$	7	-3	0.7

Table 3.2: Total future expected utility of agent A and agent B for each possible favor set.

	none	$B1$	$B2$	$B1\&B2$
none	0.0, 0.0	1.2, -1.5	9.8, -8.3	11.0, -9.8
$A1$	-1.4, 1.5	-0.2, 0.0	8.4, -6.8	9.6, -8.3
$A2$	-4.8, 8.8	-3.6, 7.3	5.0, 0.5	6.2, -1.0
$A1\&A2$	-6.2, 10.3	-5.0, 8.8	3.6, 2.0	4.8, 0.5

As agents are rational, they will only enter a favor relationship if it is beneficial. This puts a lower bound of future expected utility to 0 for each agent.

3.6.2 Sustainable Favors Solution Concept

We can abstract this stochastic relationship into an agreement between the agents to model equilibria. Each agent assumes that the other will provide the corresponding favors for the current equilibrium. However, if one agent does not offer a favor required to maintain the current relationship, then both agents must adopt a new equilibrium.

Consider agents A and B with the favor possibilities outlined in Table 3.1. Using discount factors of 0.5 and 0.6 for A and B respectively, we can compute the expected future utility of each set of favors, as shown in Table 3.2.

Each agent can only decrease its own utility, and each agent has the ability to terminate the relationship at any time by offering no favors. A rational agent will not accept a favor set that offers a negative expected payoff.

To find the equilibria, we can examine each allocation and see whether any agent could unilaterally improve its utility by offering a different allocation of favors while keeping the other agent's utility non-negative. For example, if we begin with the set of all four favors, $\{A1, A2, B1, B2\}$, agent B could improve its utility while still keeping A's expected utility positive by not offering $B2$. When A is offering $A1$ and $A2$, A can improve its utility while keeping B's utility positive by only offering $A2$. If either agent does not offer their respective

the agent will assume that only that favor will no longer be provided and find the equilibrium without that favor.

favor in the set $\{A2, B2\}$, then the favor relationship will be broken. Following this process is much the same way as finding iterated strict dominance of strategies in a classic normal form game except that the other agent's utility must be kept above 0.

Algorithm 1 returns *agent*'s best response to the set of favors, j, offered by *otheragent*. The variable *best* is the current best response, I_{agent} is the set of possible favors that the agent can offer, and $\mathcal{P}$ is a function which returns the power set of a given set. The function U_{agent} returns the total expected utility of *agent*, with the first parameter being the set of favors it will offer and the second parameter being the set of favors it will receive.

Algorithm 1 FindBestResponse($agent, otheragent, j$)

```
best ← ∅
for all i ∈ P(I_agent) do
  if U_agent(i, j) > U_agent(best, j)
  and U_agent(i, j) ≥ 0 and U_otheragent(j, i) ≥ 0 then
    best ← i
  end if
end for
return best
```

Algorithm 2 uses Algorithm 1 to find the equilibrium of the game starting with the specified agent from a specified favor set. The algorithm recursively alternates between each agent, finding each agent's best response and checking whether the other agent would prefer to change its strategy. Within the algorithm, $i_A \in I_A$ and $i_B \in I_B$ are the current favor set, $best_A$ and $best_B$ are the best strategies encountered, and *agent* is the agent that has the current decision.

Whereas the example in Table 3.2 only has one equilibrium, multiple equilibria may exist. Some may only be reachable if the agents start at specific favor sets rather than the full favor set, which is why Algorithm 3 examines all possibilities. The algorithm also looks at the possibilities of each agent being permitted the first move. With multiple equilibria, the agents choose favor sets from among equilibria strategies along the Pareto frontier. This set of favors may be used as a normal form game from which mixed strategies may be derived.

The algorithms we present in this section can be made more efficient by using dynamic programming or function memoization. We present the algorithms as they are for the sake of clarity. Like the complexity of traditional iterated dominance [Conitzer and Sandholm, 2005], these algorithms are NP-complete with respect to the size of the favor sets.

Algorithm 2 FindEquilibriumFromPosition($i_A, i_B, agent$)

```
if agent = A then
   best_A ← FindBestResponse(A, B, i_B)
   best_B ← FindBestResponse(B, A, best_A)
else
   best_B ← FindBestResponse(B, A, i_A)
   best_A ← FindBestResponse(A, B, best_B)
end if
if best_A = i_A and best_B = i_B then
   if U_A(best_A, best_B) ≥ 0
   and U_B(best_B, best_A) ≥ 0 then
      return (best_A, best_B)
   else
      return (∅, ∅)
   end if
end if
if agent = A then
   return FindEquilibriumFromPosition(best_A, i_B, B)
else
   return FindEquilibriumFromPosition(i_A, best_B, A)
end if
```

Algorithm 3 FindAllEquilibria

```
e ← ∅
for all (i_A, i_B) ∈ P(I_A) × P(I_B) do
   e ← e∪ FindEquilibriumFromPosition(i_A, i_B, A)
   e ← e∪ FindEquilibriumFromPosition(i_A, i_B, B)
end for
return e
```

3.7 Conclusions

Our favor model offers a mechanism for self-interested agents to achieve cooperation when agents can only decrease their own utility to increase others' utility. While it does not necessarily achieve the maximum possible social utility, it maximizes an agent's utilities under its own private discount factor while ensuring that agents can expect to not lose utility by helping others. Using adaptive discount factor modeling allows analysis to bridge the gap between reputation and rational strategy. This modeling also allows agents to use discount factors in other contexts besides favors. For example, if agents were performing market transactions or playing other repeated games with one another, our favor model can supplement such interaction systems.

Agents learn other agents' discount factors and exploit reciprocity. Agents also have the ability to avoid loss by refusing favors to agents with low discount factors or inconsistent strategies. Our strategy converges to a steady-state perfect Bayesian equilibrium. For these reasons, our model approximately meets the criteria described by Vu et al. [2006] for effective learning algorithms in multi-agent systems.

Chapter 4

Strategic Transactions With Private Discounting

4.1 Introduction

Many kinds of business transactions can now be automated or semi-automated, such as procurement, low-touch sales, and stock and commodity trading. As the use of these systems and agents becomes widespread and more competitive, automating strategy becomes increasingly important. One primary question in dealing with such interactions is how much one agent should trust another.

Our primary contribution is the foundation of a strategic transaction-based interaction model with agents that have private discount factors. The model is a stylized market in which a set of agents participate to gain utility. We make the strong assumption that all costs and valuations are common knowledge. While this assumption limits the direct applicability of the model in many practical settings, such as business-to-business transactions, most of our model may be extended to work in these settings by including beliefs about valuations which may be learned [Saha et al., 2003] or inferred. The assumptions of public valuations and costs allow us to more clearly describe and define our model and results by minimizing notational verbosity. We derive optimal and approximately optimal strategies and Bayes-Nash equilibria. Our model supports learning to trust other agents based on their observed discount factor in market interactions, and strategically communicating about other agents' discount factors.

Because modeling other agents' discount factors while strategically interacting with them has many subtleties, we proceed through this chapter by introducing a simple model and then iterate over several models, each adding their own complexity. After describing related work, we present the basic transaction model. We proceed covering the grim trigger strategy and introduce novel refinements to handle private discount factors. We next introduce the possibility that agents can undergo change of their own discount factors, and in doing so, turn the grim trigger refinements into effective and adaptive strategies. We then discuss interagent communication and present simulation results. Finally, we draw some conclusions.

4.2 Related Work

Trust, reputation, and reliability have been widely studied in multi-agent systems in a variety of contexts [Jøsang et al., 2007]. Most work from the computing and AI perspective focuses on measuring, communicating, and aggregating trust and beliefs to decide whether to begin a transaction with another agent. Though incentives are regarded as being important in much of this literature, many models do not exhibit explicit incentivization mechanisms nor study strategic interactions. Wang and Singh develop an observation-based model in which agents can model the probability and confidence that another agent will act in a positive fashion [Wang and Singh, 2007]. Their model only accounts for equally weighted positive and negative experiences, and does not model strategic interactions. Other models, such as those by Huynh et al. [Huynh et al., 2006] and Yu and Singh [Yu and Singh, 2002], provide empirical evidence from simulation that their models improve social welfare. However, the results may not hold with self-interested agents.

Jurca and Faltings present an incentive-compatible decentralized mechanism for communicating reputation by employing specialized reputation measuring agents [Jurca and Faltings, 2003]. Their work focuses on information brokers that buy and sell information on the reputation of another set of agents that use a different currency to play prisoner's dilemma games. The two tiers of agents differentiate the scope of their model from ours, though our model could conceivably be combined with parts of theirs. Sandholm and Lesser explore mechanisms of using divisible goods to encourage agents to exchange without external enforcement [Sandholm and Lesser, 1995]. While their approach solves similar problems to ours, it requires divisible goods and exchanges. Further, our model focuses on determining agents' differing discount factors and accordingly adapting strategic behavior.

Trust has also been widely studied in economics. A notable difference from the economic perspective, compared with the computing and multi-agent perspective, is that economics literature tends to primarily focus on incentives, strategic interactions, and consequences. In his survey of economics literature on trust, James [2002] offers a suggestion that the use of the word *trust* should be abandoned in the literature because of its ambiguities with regard to incentives and strategy. The use of the word *trust* is notably absent in much of the game-theoretic literature.

Our work is similar to Sen's work [Saha et al., 2003, Sen, 2002] in that we build a reciprocity model on future expectations, but we allow for the discovery of private discount factors, observe ranges of responses rather than point values, and do not use randomization to communicate signals. Our model also resembles that of Buragohain et al. [2003] in the way we are using incentives to build trust in an environment with favors, but the primary differences are that their model has continuous interaction and does not deal with discount factors.

Our agent communication model offers similar effects to the model proposed by Procaccia et al. [2007]. Agents maintain observations and communications which are used in aggregation to further evaluate each other and give future recommendations. The primary differences are that their method uses randomization to communicate reputation scores instead of observations of discount factors and our aggregation methods are based on ranges of observations.

Systems to incentivize agents to share resources in peer-to-peer (P2P) file-sharing networks are an application of reputation management that has received considerable attention. Kamvar et al. develop a distributed probabilistic measure of trust [Kamvar et al., 2003]. While they do not address the issues from a game-theoretic perspective, they argue how their model will hold up under different threats. Buragohain et al. [Buragohain et al., 2003] and Li et al. [Li et al., 2007] develop game-theoretic models to incent agents to share resources by using a probabilistic approach in trusting other agents. The primary difference between our assumptions and those of Buragohain et al. and Li et al. is that our model uses discount factors and agents are presented with a series of independent transactions instead of having time-invariant utilities in a steady-state trading of resources with explicit externalities. Golle et al. are able to circumvent some of the trust problem in centralized P2P systems by developing a centralized market-based transaction system, aided by delivery quality guarantees afforded by cryptographically secure hashes of commodity files [Golle et al., 2001].

At the core of our model is the iterated prisoner's dilemma. While derived in game theory and economics to model many situations involving trust, the strategic play in the prisoner's dilemma has been widely studied in multi-agent systems as well. Sandholm and Crites apply on-line reinforcement learning for determining optimal strategies in the prisoner's dilemma [Sandholm and Crites, 1995]. Birk shows that evolutionary game theory is useful to bring populations to trust other agents in an N-player prisoner's dilemma [Birk, 2001].

The problems of measuring individuals' discount factors and modeling their change over time have also been studied. However, these problems have been studied to a lesser degree than most of the aforementioned topics, particularly when both problems are featured together. Rust and Phelan use a search technique to approximate individual discount factors in their study on social security and medicare [Rust and Phelan, 1997]. Hazard constructs a favor exchanging model for gift economies, measuring discount factors and using them strategically [Hazard, 2008]. This

		a_2	
		Pay (P)	Not Pay (N)
a_1	Give (P)	$m - c, w - m$	$-c, w$
	Not Give (N)	$m, -m$	$0, 0$

Figure 4.1: Our two-player transaction model.

model is closest to the work we present, although we find that discount factor correlates better with utility with our transactions method than is shown in the gift economy. Baye and Jansen construct folk theorems for stochastic discount factors [Baye and Jansen, 1996]. In comparison, our work treats discount factor replacement as different agents, so individual agents are not anticipating their own future discount factors.

4.3 The Transaction Game

Each agent $a \in A$ is assigned a type including a discount factor, $\gamma_a \in [0, 1)$, which indicates the agent's time preference for utility. At every round, an agent is randomly paired off with other agents to play a randomized transaction game. Each possible pair of agents has a rate of encounter of r_{a_1,a_2} encounters per round, where $r_{a_1,a_2} \in (0, 1]$ signifies the probability of an encounter between the two agents with agent a_1 selling an item to agent a_2. An agent's goal is to maximize its utility with respect to its discount factor. Each agent is thus incentivized to learn other agents' discount factors to formulate a best response for each game played, and will play according to its current beliefs of the other player. After playing a game, each agent records an observation of the encounter to learn more about the other agent.

The games are a manifestation of the prisoner's dilemma, modeled as a market transaction where all prices and costs are public. Agent a_1 is the producer or seller of a good or service, which we shall refer to as an item, and agent a_2 is the consumer. We will use a_1 and a_2 to denote two generic agents as their roles may be reversed, but we will primarily use notation from a_1's perspective.

We consider three prices associated with this transaction: the cost for a_1 to produce the item, c, the market value of the item, m, and a_2's willingness to pay for the item, w. When referring to specific values, we will use the lower case c, m, and w and when referring to the random variable, we will use upper case C, M, and W. While we use the same price distributions for all agents for simplicity, this need not be the case. The model will work without modification if agents have different distributions, such as if $W_{a_1} \neq W_{a_2}$. The normal form of the game is depicted in Figure 4.1. If the game did not satisfy the constraint $w > m > c$, then at least one player would play N and thus reveal nothing of its discount factor to the other agent. Thus, we can deal with this in one of two ways: either both agents will play N, or we can choose distributions to satisfy the constraint.

The simplest way to choose distributions that satisfy the constraint is to choose the numbers from three non-intersecting ranges as $c \in (0, q]$, $m \in (v, x]$, and $w \in (y, z]$ with $q \leq v$ and $x \leq y$. If we were to instead use three random variables, X_1, X_2, and X_3 to construct the game as $c = X_1$, $m = X_1 + X_2$, and $w = X_1 + X_2 + X_3$, we would have distributions that satisfied the constraints and had easily discernable expected values. The drawback of adding three random variables together is that we push the distribution of w and m towards the normal distribution due to the central limit theorem. While our model can be modified to support overlapping distributions, where both players play N to ignore games that violate $w > m > c$, we use independent variables to keep the formulae more readable.

4.4 Grim Trigger Strategies

In this section, we introduce the classical *grim trigger* strategy (play P until opponent plays N, then play N forever), abbreviating grim trigger as GT. We modify GT to utilize knowledge of discount factors and define an optimistic version to discuss the strategy playing against itself. We assume that the population is not stuck in an evolutionary stable strategy of all playing N. Each agent is playing a strategy that involves protecting itself against N actions from other agents, which may include the possibility of punishing opponents that play N. Such strategies include variants of *tit-for-tat* (start with P and play opponent's previous move), and GT.

4.4.1 Grim Trigger Above Discount Factor

If a_2 is of a GT type, then a_1 needs to decide based on its discount factor whether a large short-term gain is better than a lower but longer term-gain. Agent a_1's expected gain from any given successful future transaction can be expressed as $E(M) - E(C) = E(M - C)$, where $E(M)$ is the expected market price of an item that is paid to a_1, and $E(C)$ is the expected cost for a_1 to produce or obtain the item. In the case of playing against a strict GT type, the discounted utility of playing P, $U_{GT}(P)$ may be found by taking the expected value of the geometric sequence of the discount factor in terms of the rate of encounter, r_{a_1,a_2}, as

$$\begin{aligned} U_{GT}(P) &= (m - c) \cdot \gamma_{a_1}^0 + \sum_{t=1}^{\infty} \gamma_{a_1}^t \cdot r_{a_1,a_2} \cdot E(M - C) \\ &= m - c + \frac{\gamma_{a_1}}{1 - \gamma_{a_1}} r_{a_1,a_2} \cdot E(M - C). \end{aligned} \tag{4.1}$$

Similarly, the expected utility of playing N, $U_{GT}(N)$ is

$$U_{GT}(N) = m \cdot \gamma_{a_1}^0 + \sum_{t=1}^{\infty} \gamma_{a_1}^t \cdot r_{a_1,a_2} \cdot 0 = m. \tag{4.2}$$

When playing against an agent of type GT, a_1 should play P if and only if $U_{GT}(P) \geq U_{GT}(N)$. From Equations 4.1 and 4.2, a_1 will play P if and only if

$$\frac{c}{c + r_{a_1,a_2} \cdot E(M - C)} \leq \gamma_{a_1}. \tag{4.3}$$

While the GT strategy can protect an agent from losses, GT agents can miss out on potential gains, particularly in stochastic games. When a_2 observes a_1 playing N, a_2 infers that a_1's discount factor is below the threshold required to play P. However, future games may contain immediate payoffs for a_1 not high enough to warrant playing N. We introduce a new strategy, *Grim Trigger above Discount Factor* (GTDF). Using this strategy, agent a_2 begins by playing P. Whenever a_1 plays N, a_1 will record the highest discount factor required for an agent to play N against a GTDF agent, and then a_2 will always play N for any game exhibiting payoffs that it expects a_1 to play N. For future games with payoffs that require a lower discount factor than recorded, a_2 will play P until it observes a_1 playing N below the corresponding discount factor.

When a rational, normal type a_1 is faced against a GTDF type a_2, a_1 must continually evaluate whether to play N or P based on a_2's knowledge of a_1's discount factor. If a_1 plays P, a_2 will not lower its expectation of a_1's discount factor. Agent a_1 knows that the lowest discount factor it has exhibited to a_2 in a game is $\underline{\gamma}_{a_1}$, where $\underline{\gamma}_{a_1} \geq \gamma_{a_1}$. We also introduce corresponding notation for c and m. We denote the smallest values where the corresponding agent played N as $\underline{c}$ for a_1 and $\underline{m}$ for a_2. Agent a_2 will play N for any game that offers an incentive for a_1 to play N given a_2's belief of a_1's discount factor, reducing a_1's outcome to 0 in those games.

When a_1 plays N, it will no longer be able to gain utility from any games that require a higher discount factor than it has exhibited. We can thus rewrite Equations 4.1 and 4.2 to account for the utility gained or lost by a GTDF agent playing N. Here we only need to consider games when $C \in [c, \underline{c}]$. The probability of a game meeting this criteria can be denoted as $P(C \in [c, \underline{c}])$. Values of C outside of this range are given 0 marginal utility since they will not be affected by the agent's action in this game, because a_2 will play N above $\underline{c}$ and a_2 will play P under c (assuming it is considering playing P in this game). If a_1 plays N, then a_2 will play N for all values greater than or equal to the current value of c, as a_2 will modify $\underline{\gamma}_{a_1}$ accordingly. For conciseness, we define the shorthand notation $PE(Y|X) \equiv P(X) \cdot E(Y|X)$, where $P(X)$ is the probability of event X occurred and $E(Y|X)$ is the expected value of Y given that X occurred. We can write a_1's expected discounted marginal utility of playing P, $U_{GTDF}(P)$, as

$$\begin{aligned} U_{GTDF}(P) &= (m-c)\cdot\gamma_{a_1}^0 + \sum_{t=1}^{\infty}\gamma_{a_1}^t \cdot r_{a_1,a_2}\cdot PE(M-C|C\in[c,\underline{c}]) \\ &= m-c+\frac{\gamma_{a_1}}{1-\gamma_{a_1}} r_{a_1,a_2}\cdot PE(M-C|C\in[c,\underline{c}]). \end{aligned} \tag{4.4}$$

Because a_1 does not gain utility above the initial m given the condition $C\in[c,\underline{c}]$, the expected marginal utility of playing N, U_{GTDF}, is the same as U_{GT}. The corresponding inequality for Inequality 4.3 with respect to GTDF is

$$c/\left(c+r_{a_1,a_2}\cdot PE(M-C|C\in[c,\underline{c}])\right)\leq\gamma_{a_1}. \tag{4.5}$$

The two components $PE(M-C|C\in[c,\underline{c}])$, $P(C\in[c,\underline{c}])$ and $E(M-C|C\in[c,\underline{c}])$, can be written in terms of the probability density function (PDF) of c, $f_C(x)$, as

$$P(C\in[c,\underline{c}]) = \int_c^{\underline{c}} f_C(x)dx \quad \text{and} \tag{4.6}$$

$$E(M-C|C\in[c,\underline{c}]) = E(M) - \int_c^{\underline{c}} x\cdot\frac{f_C(x)}{P(C\in[c,\underline{c}])}dx. \tag{4.7}$$

To ensure the consistency of the strategy given by Equation 4.7 against the GTDF strategy, we verify that our consideration that only games with $C\geq c$ will be affected by an N action with Theorem 4. We do this by showing that c and γ_{a_1} are bijective.

Theorem 4 *The discount factor at which a_1 would be indifferent between playing N and P for a given value of c against a GTDF type a_2, $\gamma_{a_1}^*(c)$, defined as $\gamma_{a_1}^*(c) = c/\left(c+r_{a_1,a_2}\cdot PE(M-C|C\in[c,\underline{c}])\right)$, is bijective on the domain $[0,\infty)$ and range $[0,1)$ for any differentiable PDF of C where $E(C)$ is defined.*

Proof 4 *Let $G(c)\equiv r_{a_1,a_2}\cdot P(C\in[c,\underline{c}])\cdot E(M-C|C\in[c,\underline{c}])$. From Equations 4.6 and 4.7, $G(c) = r_{a_1,a_2}\cdot E(M)\cdot\int_c^{\underline{c}} f_C(x)dx - r_{a_1,a_2}\cdot\int_c^{\underline{c}} x\cdot f_C(x)dx$. For $\gamma_{a_1}^*(c)$ to be strictly increasing, its first derivative must always be positive, that is, $\frac{d}{dc}\frac{c}{c+G(c)} > 0$. Using the quotient rule, eliminating the positive squared quotient, and rearranging yields $r_{a_1,a_2}\cdot P(C\in[c,\underline{c}])\cdot(E(M) - E(C|C\in[c,\underline{c}])) > r_{a_1,a_2}c\cdot(E(M)\cdot\frac{d}{dc}\int_c^{\underline{c}} f_C(x)dx + \frac{d}{dc}\int_c^{\underline{c}} x\cdot f_C(x)dx)$. We can eliminate r_{a_1,a_2} from both sides of the inequality because it is non-negative. Because the upper bound of the integrals is not a function of c, we can take the negative derivative at the lower bound to find $P(C\in[c,\underline{c}])\cdot(E(M)-E(C|C\in[c,\underline{c}])) > c\left(-E(M)\cdot f_C(c) - c\cdot f_C(c)\right)$. As f_C is a*

PDF and thus non-negative, c is positive, and $E(M)$ is positive, the right side is non-positive. Additionally, because $M > C$ always holds, the left side is positive. Therefore, $\gamma^*_{a_1}(c)$ is strictly increasing, thus injective.

To prove that $\gamma^*_{a_1}(c)$ is bijective, we also need to show it is surjective. Because $E(M - C|C \geq 0) > 0$ by definition of the game, $\gamma^*_{a_1}(0) = 0$. We employ l'Hôpital's rule to find $\gamma^*_{a_1}(\infty) = \lim_{c\to\infty} 1/(1 - E(M) \cdot f_C(c) - c \cdot f_C(c))$. Since $C > 0$, $E(C)$ is defined, and the area under f_C is 1, we know that $\lim_{c\to\infty} c \cdot f_C(c) = 0$ must hold and thus $\lim_{c\to\infty} f_C(c) = 0$ must also hold. Therefore, $\gamma^*_{a_1}(\infty) = 1$, so $\gamma^*_{a_1}$ spans the codomain of $[0, 1)$ and is surjective.

If a_1 plays N when a_2 was expecting a_1 to play P, then a_2 must update its belief about a_1's discount factor. In this case, a_2 can use Theorem 4 to find a new value of $\underline{\gamma}_{a_1}$, above which to always play N. By replacing the appropriate values, Theorem 4 holds for the buyer's role as well.

4.4.2 Equilibria of Optimistic GTDF Agents

Consider two agents playing the transaction game, where both agents are of an *optimistic* type (OPT). We define an optimistic type as employing a strategy where an agent believes that its opponent's discount factor is the maximum value possible given everything it has observed, but otherwise behaves in a strategic, rational manner. With no information about its opponent, an OPT type agent initially believes its opponent's discount factor is 1, thus the agent can only learn that the actual value of its opponent's discount factor is below the currently believed value.

When both agents are of OPT type, each needs to keep track of its opponent's belief of its own discount factor. While the game's strategies from both roles are similar, we shall again assume the perspective of a_1 in the sell role. Agent a_1's initial belief of a_2's discount factor is $\underline{\gamma}_{a_2} = 1$, the underline again representing the lowest currently believed value. Agent a_1 knows that a_2 initially believes a_1's discount factor is $\underline{\gamma}_{a_1} = 1$.

To maximize utility, a_1 needs to evaluate its expectations of a_2's discount factor as well as its expectations of a_2's exceptions of a_1's discount factor. Agent a_1 should play P only if P both offers a higher expected utility and it expects a_2 to also play P.

We can extend Equations 4.6 and 4.7 to include the PDF of m as

$$P(C \in [c, \underline{c}] \cap M \leq m) = \int_{-\infty}^{m} \int_{c}^{\underline{c}} f_C(x) \cdot f_M(y) dx dy \tag{4.8}$$

and

$$E(M - C|C \in [c, \underline{c}] \cap M \leq m) = \int_{-\infty}^{m} \int_{c}^{\underline{c}} (y - x) \cdot \frac{f_C(x) \cdot f_M(y)}{P(C \in [c, \underline{c}] \cap M \leq m)} dx dy. \quad (4.9)$$

We can then use Equations 4.8 and 4.9 to modify Inequality 4.5 to include the cases when a_2 will play N based on $\underline{\gamma}_{a_2}$ and the corresponding c and m values. To maximize utility, an OPT type a_1 should play P if and only if both of the inequalities

$$\gamma_{a_1} \geq \frac{c}{c + r_{a_1,a_2} \cdot PE(M - C|C \in [c, \underline{c}] \cap M \leq \underline{m})} \quad (4.10)$$

and

$$\underline{\gamma}_{a_2} \geq \frac{m}{m + r_{a_1,a_2} \cdot PE(W - M|M \in [m, \underline{m}] \cap C \leq \underline{c})} \quad (4.11)$$

are satisfied. If Inequality 4.11 is not satisfied, then a_2 will play N, and so a_1 should also play N. If Inequality 4.10 is not satisfied, then a_1 should play N to protect itself from loss. Note that the optimal behavior of a_1 playing N when a_2 plays P is covered by these two inequalities if $\underline{\gamma}_{a_1}$ were substituted in for γ_{a_1} in Inequality 4.10. Because each agent maximizes its utility based on knowledge of the other agent, and no agent can gain utility by playing differently, this is a unilaterally optimal strategy and thus a Bayes-Nash equilibrium given the assumptions of the OPT agent type.

4.4.3 Dynamic Behavior of GTDF Agents

The strategies we have presented for GTDF and OPT agents do not necessarily produce sustainable outcomes. Agents begin with an ideal reputation, but the subgame perfect Bayes-Nash equilibria we describe lead an agent to continually spend its reputation until it asymptotically achieves the worst possible reputation. Suppose a_2 currently believes that the maximum c for which a_1 will play P is $\underline{c}$. We define c^* as the maximum value of c that satisfies Inequality 4.10 and m^* corresponding to Inequality 4.11. We can obtain the expected value of $\underline{c}$ in the next time step, $\underline{c}'$, by subtracting the expected decrease in $\underline{c}$ when a_1 plays N and a_2 is expecting a_1 to play P as

$$\underline{c}' = \underline{c} - r_{a_1,a_2} P(C \in [c^*, \underline{c}]) \cdot (E(C \in [c^*, \underline{c}] - \underline{c}). \quad (4.12)$$

Equation 4.12 can be iterated to find the time to reduce $\underline{c}$ by a given amount or to find the rate at which a_1 gains from spending its reputation over a given time. Note that the term subtracted from $\underline{c}$ is nonnegative, as an agent will not increase another agent's perception of its own discount factor. As long as $P(C \in [c^*, \underline{c}]) > 0$ and $c^* \neq \underline{c}$, a_1 will reduce $\underline{c}$.

If an agent employs an alternate strategy of maintaining a constant reputation, such as a_1

playing P whenever $c^* \geq c$, keeping $\underline{c}$ constant when a_1 expects a_2 to also play P. However, the discounted future utility for a given reputation and time is lower when maintaining a constant reputation when compared to the corresponding optimal strategy we present. The difference between those two strategies increases as an agent's discount factor decreases.

By relaxing an assumption of the OPT agents we can attain a subgame perfect Bayes-Nash equilibrium with higher utility than the strategy of spending reputation to attain short-term utility. Consider two agents with static reputations playing in a long-running equilibrium. Each agent knows the other's discount factor instead of optimistically assuming that the other has the maximum discount factor given the current information. If a_1 plays N while a_2 is expecting mutual P actions, then a_2 loses utility as well. Agent a_2 can sanction a_1 in retaliation by playing N when a_1 expects both to play P. If both agents know the other will retaliate against spending a reputation for utility gain, then agents will be incentivized to maintain their reputations and achieve higher payoffs. We investigate these strategies further in the next section.

4.5 Discount Factor Replacement

Here, we allow agents to have more complex beliefs about other agents' discount factors. To make the model more flexible, we open an agent's belief of a discount factor to include an adaptive distribution of values, still using the expected value to model an opponent's strategic interactions. Agents may start off believing that other agents' discount factors are low. As agents interact, an agent can optimally change its expected value for another's discount factor based on a process of discount factor replacement.

Each agent $a \in A$ has a constant discount factor replacement probability, λ_a. The discount factor is changed to a new value based on a Bernoulli process [1] with a probability of λ_a of changing between each game, and we shall refer to each discount factor change as a *replacement*. By using a discount factor replacement process, we are able to model agents' type changes over longer durations of time. Agent replacements have been shown to be an effective tool for modeling how agents change over time [Mailath and Samuelson, 2006]. The replacements may be due to a change in the market, agent's ownership, information, or other factors in a dynamic environment.

We make a few assumptions about the discount factor replacement process. First, we continue the assumption that the discount factor is a private value, and must be discovered by other agents. By this assumption, an agent would prefer other agents to overestimate its own discount factor in order to gain utility at the expense of other agents. Second, we

[1] As the Bernoulli process is the discrete counterpart to the Poisson process, the task of transitioning the model from synchronously repeated games to a continuous time process is mostly straightforward. Using a Poisson process, λ_a becomes the discount factor replacement rate.

assume that the discount factor replacement probability is public. Due to the memoryless nature of the replacement process, knowing the last time an agent was replaced does not matter. The actual replacement process in a variety of real-world applications may be arbitrarily complex while still allowing others to estimate the replacement probability. For example, if a firm is performing poorly because of a low discount factor, the firm may have a higher probability of being replaced than a profitable firm with high discount factor. While extending the model to allow agents to model opponents' private dynamic replacement probabilities may have some interesting implications, we focus here on the base model with constant replacement probabilities and leave such extensions of the model for future work.

Finally, we assume that agents are unaware when their discount factor changes, and at any moment each agent assumes that it will continue to have its current discount factor indefinitely. This assumption may be easily relaxed by using $\gamma_a \cdot (1-\lambda_a)$ in place of γ_a as the effective discount factor throughout our model and also use the observation that a discount factor cannot be more than $\gamma_a \cdot (1-\lambda_a)$. While this final assumption is not critical to our central arguments, and our model easily supports its relaxation, this assumption allows more clarity in representation and discussion.

4.5.1 Accounting for Observation History

We denote one of agent a_1's observations, $i \in I_{a_1}$, as the tuple $(a_i, a'_i, \gamma^*_i, t_i)$, where a_i is the agent that made the observation, a'_i is the agent the observation is made about, γ^*_i is the observation range, and t_i is the time of the observation. Given observation event i after playing a game against a'_i, the observed discount factor is above or below a constant b_i, denoted as $\gamma^*_i = [b_i, 1)$ or $\gamma^*_i = [0, b_i]$. We expand the use of the symbol γ^* from Theorem 4, when used with an observation subscript, to represent this range indicated by the observation instead of the indifference point. The two possible observations are $\gamma_{a'_i} \geq \underline{b}_i$ and $\gamma_{a'_i} \leq \overline{b}_i$, where we use a line above or below b and other variables to indicate whether they are an upper or lower bound. Consider the case observation i at time t_i shows $\gamma_{a'_i} \leq \overline{b}_i$. We use T to denote the present time. The cumulative distribution function (CDF) of $\gamma_{a'_i}$ at the present time given an upper-bound observation i for agent a may be written as a function of an input discount factor, x, as $\overline{F}_i(T, x)$. The CDF, $\overline{F}_i(T, x)$, is related to the PDF of $\gamma_{a'_i}$ given this event, $\overline{f}_i(T, x)$, in the usual fashion as $\overline{F}_i(T, x) = \int_0^{\overline{b}_i} \overline{f}_i(T, x)dx$.

Upon immediate observation of i, agent a'_i will not be able to change type, so $\overline{F}_i(t_i, \overline{b}_i) = 1$. Similarly, when the observation is no longer relevant, knowing nothing about agent a'_i's discount factor distribution entails using the corresponding maximum entropy distribution. If an agent has a priori knowledge of the distribution of discount factors, then the model may be adjusted to accommodate the prior probability. Given an upper and lower bound, the maximum entropy

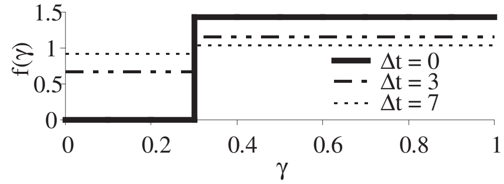


Figure 4.2: Graph of discounted PDF of discount factor, given $\lambda = .7$, and observation $\gamma \geq .3$.

distribution is uniform. Therefore, the infinitely discounted distribution of agent a_i''s discount factor must yield $\overline{F}_i(\infty, \bar{b}_i) = \bar{b}_i$. We may discount the observation given the discount factor replacement parameter as $\overline{F}_i(T, \bar{b}_i) = 1 - (1 - \lambda_{a_i'}^{T-t_i})(1 - \bar{b}_i)$. This makes the other region $1 - \overline{F}_i(T, \bar{b}_i) = (1 - \lambda_{a_i'}^{T-t_i})(1 - \bar{b}_i)$. We divide each region by its corresponding length to find the discounted PDF for the case of $\gamma_{a_i'} \leq \bar{b}_i$ as

$$\overline{f}_i(T, x) = \begin{cases} \left(1 - (1 - \lambda_{a_i'}^{T-t_i})(1 - \bar{b}_i)\right)/\bar{b}_i & \text{if } x \leq \bar{b}_i, \\ 1 - \lambda_{a_i'}^{T-t_i} & \text{if } x > \bar{b}_i. \end{cases} \tag{4.13}$$

Similarly, the PDF for $\gamma_{a_i'} \geq \underline{b}_i$ is

$$\underline{f}_i(T, x) = \begin{cases} \left(1 - (1 - \lambda_{a_i'}^{T-t_i})\underline{b}_i\right)/(1 - \underline{b}_i) & \text{if } x \geq \underline{b}_i, \\ 1 - \lambda_{a_i'}^{T-t_i} & \text{if } x < \underline{b}_i. \end{cases} \tag{4.14}$$

Figure 4.2 shows an example of this PDF discounting.

To find the expected value for agent a_i''s discount factor, we can use the definition of conditional probability to combine historical probabilities. The PDF of $\gamma_{a_i'}$ given the observation history is equal to the intersection of all of these probabilities divided by the total probability of all the intersections, because we are given that $\gamma_{a_i'}$ is chosen from the intersection of these probabilities. Using the function f_i to denote the proper $\overline{f}_i$ or $\underline{f}_i$ as appropriate for the observation, we can thus find the expected value of $\gamma_{a_i'}$ for the current time T from the set of observations I as

$$E(\gamma_a | T) = \int_0^1 x \frac{\prod_{i \in I, a_i' = a} f_i(T, x)}{\int_0^1 \prod_{i \in I, a_i' = a} f_i(T, y) dy} dx. \tag{4.15}$$

Figure 4.3 depicts the PDF from three combined observations as

$$f_{a,I}(\gamma) = \frac{\prod_{i \in I, a_i' = a} f_i(T, \gamma)}{\int_0^1 \prod_{i \in I, a_i' = a} f_i(T, y) dy}. \tag{4.16}$$

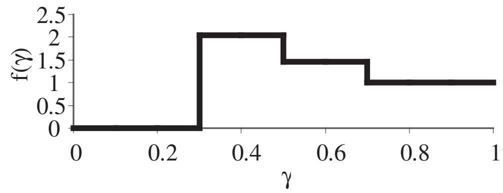


Figure 4.3: Graph of combined discount factor PDFs, given $\lambda = .7$, and observations $\gamma \geq .3$ at time $T - 0$, $\gamma \leq .7$ at $T - 4$, and $\gamma \geq .5$ at $T - 5$.

Suppose an agent observes $\gamma_a \geq .5$ at time 0 and then $\gamma_a \leq .4$ at time 4. Because the overlap of these regions is null, the old evidence is no longer valid and that agent a has undergone a discount factor replacement between the two observations. Further, any observations prior to the replacement should be disregarded.

While we assume that an agent's rate of replacement is public, conflicting observations reveal information about the underlying rate, making it possible to discover if it were private. One way to measure the replacement rate is to find the relative entropy gained by each new observation. If the information gain is high after a long history, then it is more likely that a replacement has occurred. Another is to find the posterior probability of both a retainment and replacement of the discount factor for each combination of observations to determine the expected value of the number of replacements for a given time frame. Both of these methods assume that λ_a does not change. In reality, the probability of an agent being replaced may be an arbitrarily complex function, as mentioned earlier. In these cases, domain specific models or machine learning may be more appropriate.

We will refer to the agent type that utilizes the observations discussed in this section to determine other agents' discount factors and other agents' perceptions of its own discount factor as *Adaptive Discount Factor Discovery* (ADFD). By using discounted reputation measurements, agents receive an eventually forgotten punishment for playing below their discount factor. The temporary punishment gives ADFD the same effect of Tit-for-Tat, but takes into account stochastic interactions and differing discount factors.

4.5.2 Dealing With Observation Errors

Consider the probability of making an erroneous observation of agent a' is $p_{a'}$. This could be due to a' attempting but failing to deliver a favor. In looking at the PDF expressed above, the value of the PDF in the region where the discount factor cannot be can be expressed as $0 \cdot \lambda_{a'_i}^{T-t_i} + 1 \cdot (1 - \lambda_{a'_i}^{T-t_i})$. In dealing with the probability, this formula now needs to yield $p_{a'}$

when multiplied by the area $1 - \overline{b}_i$. This formula becomes $\frac{p_{a'}}{1-\overline{b}_i} \cdot \lambda_{a_i'}^{T-t_i} + 1 \cdot (1 - \lambda_{a_i'}^{T-t_i})$. Since the area under the PDF must be 1, we can split the region into two, the lower with area v_1 and the upper with area v_2, with the constraint of $1 = \overline{b}_i v_1 + (1 - \overline{b}_i) v_2$ since the area represents probability. The PDF of a''s discount factor including this error probability becomes

$$\overline{f}_i(T, x) = \begin{cases} 1 - \frac{p_{a'} - (1-\overline{b}_i)}{\overline{b}_i} \lambda_{a_i'}^{T-t_i} & \text{if } x \leq \overline{b}_i, \\ 1 - \left(1 - \frac{p_{a'}}{1-\overline{b}_i}\right) \lambda_{a_i'}^{T-t_i} & \text{if } x > \overline{b}_i, \end{cases} \tag{4.17}$$

and

$$\underline{f}_i(T, x) = \begin{cases} 1 - \frac{p_{a'} - \underline{b}_i}{1-\underline{b}_i} \lambda_{a_i'}^{T-t_i} & \text{if } x \geq \underline{b}_i, \\ 1 - \left(1 - \frac{p_{a'}}{\underline{b}_i}\right) \lambda_{a_i'}^{T-t_i} & \text{if } x < \underline{b}_i. \end{cases} \tag{4.18}$$

This formula may also be written in the other notation (used for favor reciprocation) as

$$f_i(T, \gamma) = \begin{cases} 1 - \frac{p_{a'} - (1 - (\sup \gamma_i^* - \inf \gamma_i^*))}{\sup \gamma_i^* - \inf \gamma_i^*} \lambda_{a_i'}^{T-t_i} & \text{if } \gamma \in \gamma_i^*, \\ 1 - \left(1 - \frac{p_{a'}}{1 - (\sup \gamma_i^* - \inf \gamma_i^*)}\right) \lambda^{T-t_i} & \text{if } \gamma \notin \gamma_i^*. \end{cases} \tag{4.19}$$

4.5.3 Concealing an Agent's Discount Factor

ADFD type agents must deal with replacement of other agents' discount factors. If a_1 is aware of a_2's currently low discount factor, then a_1 may frequently play N to prevent loss of utility to a_2's N actions. When a_2's discount factor is replaced with a higher discount factor, a_1 will still play N for some time until it discovers that a_2's discount factor is now higher. Because the replacement rates and actions between the two agents are visible to a_2, a_2 will realize that a_1 has not figured out a_2's new discount factor. Agent a_2 should not necessarily assume that a_1's N actions indicate that a_1's discount factor is lower than it actually is.

Consider the case where a_1 has additional information about a_2 that a_2 does not know that a_1 has. This information could be related to distribution of discount factors, obtained from other agents, or a priori beliefs. Because a_2 does not know a_1 has this information, a_2 attributes all of a_1's actions to a_1's discount factor.

We introduce a refrain action, R, which allows an agent to abstain from playing in a round. Action R allows an agent to strategically protect itself against an agent with a lower discount factor without giving false indications of its own discount factor. When either agent plays R, both agents receive 0 utility for the round, regardless of what the other played. Because of the replacement rate, playing R will not stop other agents' views of an agent's discount factor from changing. Agents whose expected discount factor is low will gain slightly from playing R due to the forgetting from other agents, and likewise agents that have a high discount factor will receive a small expected loss. However, the change in expected discount factor will be changed by the same rate for all actions, including R, due to aging observations. Action R makes the model fit more realistic situations, because an agent can often choose whether to interact or do business with another agent. The introduction of R makes the P and N actions more akin to actions after entering a contract, whereas R is to not enter the contract.

To evaluate the utility of the different actions, we need to extend some definitions from the simpler models. We update the definition of $\underline{c}$ and $\underline{m}$ to reflect the highest prices at which an agent will play P that corresponds to the expected value of each agent's discount factor. Theorem 4 continues to hold when given an additional condition on the distribution of M as exhibited in Equations 4.8 and 4.9. This theorem allows us to use the function γ^* to map between a_2's expected value of a_1's discount factor, $E(\gamma_{a_1})$, and a_2's perceived maximum price for which a_1 will play P, $\underline{c}$. We also expand the definition of c^* as the inverse of γ^* to find $\underline{c}$ given a discount factor.

Agent a_1 needs to determine what it expects a_2 to play given only the current information a_1 knows about a_2. Each agent knows it will be retaliated against if it plays N below the maximum value where its expected discounted future utility is equal to the value of playing N in the current game. We denote the expected discount factor of a_2 as $E(\gamma_{a_2})$, yielding a result similar to Inequality 4.11, as

$$E(\gamma_{a_2}) = \frac{\underline{m}}{\underline{m} + r_{a_1,a_2} \cdot PE(W - M | M \geq \underline{m} \cap C \leq \underline{c})}. \tag{4.20}$$

If using the m and c values for the current game in place of $\underline{m}$ and $\underline{c}$ in Equation 4.20 results in a discount factor greater than $E(\gamma_{a_2})$, then a_1 expects a_2 will want to play N. However, a_2 will play R because a_2 knows that a_1 expects this and will play R as to not divulge information about its discount factor. Agent a_2's model of a_1 will look like Inequality 4.10 based on a_2's expected discount factor of a_1, $E(\gamma_{a_1})$, as

$$E(\gamma_{a_1}) = \underline{c} / \left(\underline{c} + r_{a_1,a_2} \cdot PE(M - C | C \geq \underline{c} \cap M \leq \underline{m})\right). \tag{4.21}$$

If $E(\gamma_{a_2})$ is sufficiently high and Equation 4.21 does not yield a sufficiently high discount factor, then a_2 expects a_1 will want to play N, but will actually play R to prevent a_2 from learning more about its lower discount factor. If both discount factors are high enough, then both agents want to play P and expect the other to play P. However, in this case, each agent can use information about itself that the other agent does not have.

When deciding to play P or N, a_1 knows playing P will increase a_2's expected value of a_1's discount factor, and playing N will probably decrease it severely. With a_1 knowing its own replacement rate, λ_{a_1}, a_1 can determine a_2's expected value of a_1's discount factor, $E(\gamma_{a_1} | \gamma'_{a_1})$, given the expected value of a_1's discount factor from the new observation γ'_{a_1}. We can write the expected discount factor given a new observation for agent a as

$$E(\gamma_a|E(\gamma'_a)) = E(\gamma_a) \cdot (1 - \lambda_a) + \lambda_a \cdot E(\gamma'_a). \tag{4.22}$$

Note that Equation 4.22 only works for an immediate observation. Equation 4.15 is required for any other time $t_i \neq T$.

To compute the expected impact of utility given an observation, a_1 will need to model how much a_2's beliefs will change of a_1's discount factor. This expected discounted future utility, $V_{ADFD}(\gamma)$, given a_2 believes a_1's discount factor is γ, is

$$V_{ADFD}(\gamma) = \sum_{t=1}^{\infty} \gamma_{a_1}^t \cdot r_{a_1,a_2} \cdot PE\left(M - C|M \leq \underline{m} \cap C \leq c^*(\gamma)\right). \tag{4.23}$$

The c value of the current game indicates the boundary of the observation of a_1's discount factor. Given only the extrema of a probability distribution, the maximum entropy distribution is uniform, so the expected value of the discount factor is the average of the extrema of the observation. Agent a_2 will expect a_1's discount factor to be $E(\gamma_{a_1}|(1+\gamma^*_{a_1}(c))/2)$ if a_1 plays P and $E(\gamma_{a_1}|(\gamma^*_{a_1}(c)+0)/2)$ if a_1 plays N. Due to the discounting of the replacement process, a_2 will expect $E(\gamma_{a_1}|1/2)$ if a_1 plays R.

Now that we have evaluated a_1's expectations of what a_2 will play and a_1's utility for a_2 observing each action, we can express a_1's expected utilities for each of its actions. If a_1 expects a_2 to play R, then a_1 should play P to give a_2 an observation of playing P. If the expectation of a_2 to play R is mutual knowledge, then a_2 may choose to ignore the observation, since a_1 may not have played P if a_2 was playing P. However, a_1 is taking on some risk in playing P, so a_2 may count the observation.

If a_1 expects a_2 to play N, then a_1 should play R to defend itself. However, if a_1 expects a_2 to play P, then a_1 needs to determine which action maximizes its own marginal utility, U_{ADFD}. In steady-state, where the influx and discounting of observations are in balance, this total discounted utility is represented as

$$U_{ADFD}(P) = m - c + V_{ADFD}(E(\gamma_{a_1}|\frac{1+\gamma^*_{a_1}(c)}{2})), \tag{4.24}$$

$$U_{ADFD}(N) = m + V_{ADFD}(E(\gamma_{a_1}|\gamma^*_{a_1}(c)/2)), \quad \text{and} \tag{4.25}$$

$$U_{ADFD}(R) = V_{ADFD}(E(\gamma_{a_1}|1/2)). \tag{4.26}$$

4.6 Agent Communication

Without collusion or side-payments, agents do not have explicit incentives for sharing information with other agents in our model. Cheap talk is non-binding signaling before a game that has no direct effect on utility, but can affect the outcome when agents' incentives are not negatively correlated [Farrell and Rabin, 1996]. While cheap talk has no effect on the outcome of one-shot prisoner's dilemma games, we show how it is useful in our repeated game model.

Communication of observations between agents can reveal conflicting observations, and communication of such information itself can be strategic under certain circumstances. If agent a_3 intentionally offers agent a_1 incorrect information about agent a_2, a_1 may simply expect that the change in γ_{a_2} is due to replacement. However, an accumulation of incorrect communications of γ_{a_2} from a_3 could indicate a replacement rate higher than γ_{a_2}'s replacement rate. Given an observation of a discount factor range, i, of a_2 from a_3, a_1 would have a probability on the truth of i, $P(i)$, as

$$P(i) = \int_{\gamma_i^*} f_{a_2,I_{a_2}}(T,x)dx. \tag{4.27}$$

Note that a_1 can only evaluate the quality of a communicated observation after making new observations and evaluating the posterior distribution of discount factors. The initial evaluation using the prior probability distribution is more of a test of believability.

Agents can use the quality of information given by another agent as a proxy for measuring that agent's discount factor. Agent a_2 wants a_1 to believe that a_2 has a high discount factor for future interactions, either to both play P with a_1 (if a_2 has a high discount factor) or fool a_1 into playing P so that it can play N (if a_2 has a low discount factor). If a_1 is in a game against a_3 using information given by a_2, and a_1 finds a_2's information to be incorrect, it may assume that a_2 was colluding with a_3 via a side-payment. Assuming collusion, a_1 may lower its expected value of a_2's discount factor by assuming a_2 and a_3 are the same agent, or may model more accurately if a_1 has some information about the side payments. A simple but less rigorous alternative is for an agent to sanction another for giving false information.

If agents offer each other side-payments for information about other agents, then strategic interaction plays a role in whether and at what price an agent will purchase information from another agent. Many models of interaction may be applicable. One model would be for agents to bid based on information metrics of their distribution of observations, such as absolute entropy or entropy relative to an announced distribution. For this model to work, an agent would need to calculate its willingness to pay given the amount of information offered by the other agents, based on what it expects to gain from the transaction. Another model would be to align incentives by deferring agents' side-payments of the information they received until after the transaction takes place. As Crawford and Sobel [1982] have shown, when one agent has aligned incentives with another agent that has more accurate information, strategy plays a major role in maximizing utility, particularly in conveying reputation information.

One model is for agents to communicate all information about other agents. If a_3 communicates to a_1 not just the expected value of a_2's discount factor, but every observation, then a_1 can better determine what observations to discard. If a_1's observations indicate that a_2 underwent a discount factor replacement at a given time, a_1 knows to discard a_3's observations before that time. Similarly, a_1 can discard its own observations earlier than a replacement indicated by a_3 provided a_1 believes a_3's reported observations are true.

A succinct communication model is for an agent to ask other agents whether they would trust a specific agent in a hypothetical situation. Consider a_1 asking a_3 whether a_3 would trust a_2 in a_1's current game. Agent a_3 can build its own reputation by answering correctly or refuse answer if a_3 does not have sufficient information. If a_3 answers it would trust a_2 for the given transaction, a_1 can make an observation that both a_2's and a_3's discount factors are sufficiently high. However, if a_3 answers that it would not trust a_2, then a_1 does not have enough information to tell whether a_2's or a_3's discount factor is low. After performing more transactions and receiving more recommendations, a_1 can review the previous recommendations to gain more information about the agents. If a_3 replied it would not trust a_2, but a_1 later learns that a_2 had had a high enough discount factor, then a_1 can make an observation of a_3's discount factor was below that required to perform the given transaction. Similarly, if a_3 had lied that it would trust a_2 even though a_2 had a low discount factor, as mentioned earlier in this section, a_1 could assume the agents were colluding or sanction a_3 by significantly reducing its expectation of a_3's discount factor. We use the simpler sanctioning method in our simulation results.

Three main communication behaviors emerge from parameterizations of our model. First, consider the case when agents are regularly interacting with a significant portion of the population and discount factor replacement is relatively low, denoted as $a, a' \in A : r_{a,a'} \gg \lambda_a$. Because agents' discount factors are relatively stable and each agent is frequently encountered, agents will easily be found out if they give false information. An example of two agents that undergo frequent transactions is an agent that continually procures perishable food from a supplier. An agent can incrementally combine evidence from other agents using Equation 4.27 to determine the probability each observation is accurate.

Second, we consider cases where $\lambda_a \gg 0$ or $\lambda_a \gg r_{a,a'}$, without a large number of agents. This represents agents' discount factors being replaced so frequently that observations would not matter. In such cases, our model is not applicable because determining an agent's reputation would matter little due to its short life.

Finally, we consider sparse agent interactions relative to the discount factor, expressed as

$a, a' \in A : r_{a,a'} \lesssim \lambda_a \ll 1$. This case exemplifies many business-to-business and business-to-consumer interactions where purchasing happens frequently enough that reputation matters, but not so infrequent that firms dramatically change between purchases. In these cases, agents must frequently depend on accurate information from other agents if they are to have an accurate estimate of another agent's discount factor. However, misinformation in this case is hard to detect a priori, and an agent is unlikely to have the opportunity to directly retaliate on an agent that provided false information.

4.7 Simulation Results

We ran simulations with groups of 16 agents to validate that our model yields higher payoff for agents with higher discount factors and to evaluate the impact of basic communication. Each simulation consisted of 100 time steps and an agent replacement rate of .01 to reflect the mean life of an agent. We randomized payoffs and interactions, but used consistent sets of agent discount factors.

Prior to setting up the experiments, we examined the affect of the distributions of C, M, and W with respect to the minimum discount factor required for transactions to occur. Many parameterizations yield results that are not particularly interesting. For example, exceedingly low values of W relative to C and M results in most agents in the buy role without a particularly high discount factor playing N because patience for later payoffs is not rewarded in that case. For simulation, we chose the parameterizations of uniform distributions of $C \in [0, 5]$, $M \in [10, 20]$, and $W \in [80, 120]$ because they offered a full range of agents' behavior.

We used two topologies to represent agent encounters for our experiments. One is a uniform distribution with each r_{a_1,a_2} chosen from the range $[0, 1]$. The other is a randomly constructed small world network. To construct the small world network, we connect two agents with encounter rates set to 1 and individually add agents to the network. Each new agent's encounter rates with every other agent are randomly chosen proportional to the sum of the encounter rates of the agent to which it is connecting, relative to the sum of all encounter rates. In this topology, each agent frequently encounters a small set of agents and occasionally encounters another agent from outside of that set.

We chose four discount factor distributions to examine: uniform, all the same, 4^{th} root of uniform distribution, and 1 minus the 4^{th} root of uniform distribution. The last two distributions allowed us to see the behavior of a group of agents where most have a high and low discount factor, respectively, but with a couple agents on having discount factors on the opposite end of the spectrum. The distribution with all of the same discount factors was to determine whether agents behaved differently in a general population than of a population of

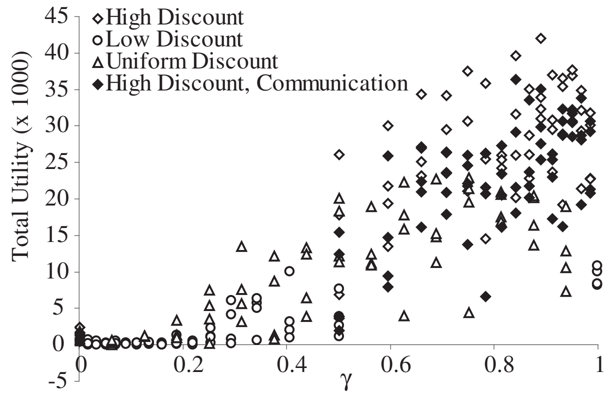


Figure 4.4: Results of simulating 16 games with 16 agents with uniformly distributed encounter rates.

similar agents.

Figure 4.4 shows an example of the overall trend that agents with higher discount factors receive higher utilities which we found across our results. Each data point represents the final utility of a given agent played with uniformly distributed rates of encounter. Those agents that are most patient and have the greatest tolerance for risk are rewarded the highest. The trends depicted remain prominent despite changes in the agents' distribution of discount factors.

Even though our results from small world encounters had 4% more encounters than the uniformly distributed encounters, the final utility of all agents was significantly decreased in all scenarios. The number of times agents played N was almost the same between both topologies, but agents in the small world played P significantly less and R more. Agents protected themselves more frequently in the small world due to the decreased sensitivity of negative observations that low encounter rates bring. While an agent traded primarily with the other few agents that it encountered frequently in the small world, the number of encounters with more rarely encountered agents is still significant.

Communication reduced the overall utility of agents with high discount factors somewhat, as can be seen by the filled-in diamonds in Figure 4.4 when compared to the corresponding open diamonds. In these results, we used the simple heuristic that an agent would communicate results whenever it had made four observations about an agent, following our simple communication method outlined earlier in this section. The additional information gained by communication reduced the agents' speculative beliefs of other agents' distributions based on the maximum entropy distribution. For example, an agent with no a priori beliefs about another agent's discount factor or the distribution of discount factors would initially believe $\frac{1}{2}$. However, we found that effective communication allows agents to better protect themselves in encounters with agents with low discount factors when large utilities are at stake.

4.8 Conclusions

We present a method for measuring agents' discount factors and strategically using those discount factors in stylized market interactions. Our models are rooted in trigger strategies. The refined grim trigger strategies we use in our formulations of our model resemble tit-for-tat once discount factor replacement is added. Trigger strategies are not only practical models of agent interaction, but can be also used to construct some equilibria. The abilities in various populations of agent types that our model offers to protect, adapt, coordinate, and exploit are strongly desirable traits in an effective multi-agent system [Vu et al., 2006]. By demonstrating a Bayes-Nash equilibria of pair-wise reputation and building framework for strategic communication of reputations, we have shown that measuring and modeling other agents' discount factors is a feasible and plausible strategic interaction model for building autonomous agents.

Chapter 5

Evaluating Discount Factors in an Online Market Model

Motivating Example: Online Market

An online auction is a practical motivating scenario for a trust system. As our running example, we outline the general mechanics of this scenario to motivate our results and formally analyze it in Section 5.1. The auction is continually cleared, with buyers choosing which sellers' offers to accept, if any. Exactly one buyer or seller moves at a time. The order of buyers' and sellers' turns are chosen from a stochastic process to simulate realistic market transactions, but each agent gets one turn per unit time. Each agent's goal is to maximize its expected utility and, to account for time preference, is endowed with a privately known discount factor.

Sellers post or update offers to sell items. A seller's costs are private but follow publicly known probability distributions. Cost are initialized before the auction begins. An offer states the asking price and the (true or exaggerated) quality of the item. We define quality as the probability density function (PDF) that an item will irreparably fail as a function of time. The expected lifetime of an item is its mean time to failure (MTTF). Section 5.1.1 considers the cases where a seller can (1) only produce a fixed quality and (2) control the quality of its items.

Buyers see all current offers and choose which and when to accept. When deciding what to purchase, buyers can see the seller's offer as well as a history of "comments" by other buyers about the seller's discount factor, price, and quality. After each transaction, a buyer can post a public comment on the seller. In our formal approach, a comment is a numerical observation of one or more of quality, valuation, time, or discount factor. A comment is formulated as a measurement or inequality, such as "I observed the good offered by agent a at price 5.29 to be of a quality that held up for 1 week of use before breaking" or "agent a's discount factor is greater than 0.60." Sellers can see what price and quality other sellers are currently offering and update their offers accordingly.

Each buyer has its own expected utility gain per unit time for having each additional item, a willingness-to-pay per unit time. Buyers have a price sensitivity with respect to quality, based on the expected useful life of an item coupled with the agent's discount factor and willingness-to-pay. When a buyer makes a purchase, it loses the utility of the amount of the purchase price and gains utility for each unit time that the item is functional.

We note that because the seller controls the price, our model's descending Dutch auction style resembles Craigslist (`http://craigslist.org`) and the retail presence on eBay, where the seller's "minimum bid" is effectively the ask price. This is in contrast to the ascending English auction commonly associated with consumer-to-consumer transactions on Amazon and eBay. We choose to examine the seller-price auction because the analysis yields somewhat simpler results and is therefore easier to discuss in the cases of interest.

5.1 Market Model Examples

This section illustrates examples of how discount factor may be measured and utilized with our example online market. The formalization of the full complex model is beyond the scope of this paper, and we leave extensions involving multiple buyers, sellers, and items simultaneously for future work. We include these basic results to motivate our central thesis of the effectiveness of modeling trustworthiness as discount factors. The first two involve typical trust settings. The third and fourth show how agents can gain knowledge of discount factors outside of something that would normally be measured as trustworthiness while contributing to the agents' knowledge of trustworthiness.

We focus more attention on measuring discount factors than on using them in decision models, as the former has received considerably less attention whereas the latter has been widely used [Dellarocas, 2005, Ely and Välimäkiz, 2003, Hazard, 2008, Jurca and Faltings, 2007, Saha et al., 2003]. Benzion et al. [1989] measure the discount factors of people directly by asking them specific questions. Although it is useful to determine an individual's discount factors from an economic perspective, such measurements may not work in a strategic setting. The literature on measuring private discount factors in strategic interactions is rather sparse. To the best of our knowledge, the following works represent most of what is currently known. The models developed by both Rubenstein [1985] and Güth et al. [2004] yield equilibrium strategies for bargaining between agents when the agents have private discount factors. However, both models require the agents' discount factors to be one of two discrete values. Smith and desJardins [2009] measure the minimal upper bound of an agents' discount factors, although their model requires the assumption that agents only reason with one level of mutual information, rather than assuming agents' actions are common knowledge.

5.1.1 Discount Factors and Production Quality

To demonstrate both how a seller's discount factor can be measured and how a seller may use its discount factor directly in decision making, we employ the frequently studied grim trigger strategy [Axelrod, 2000], where an agent permanently stops interacting with another after a bad interaction. This strategy is typical of some trustworthiness settings, particularly when many other agents supply a substitutable alternative. For example, people may not return after a bad experience at a restaurant or may not purchase a replacement printer from the same manufacturer if their previous printer required frequent maintenance. Agents in these settings would have preferred to have avoided these bad transactions in the first place.

Consider buyer b deciding whether to purchase an item advertised at a high quality from seller s for some specific price. The seller will make $\overline{\pi}$ profit on a low-quality item, and $\underline{\pi}$ profit on a high-quality item, where $\overline{\pi} > \underline{\pi}$. This means $\overline{\pi}/\underline{\pi} - 1$ is the percent increase in profit by selling the low-quality item. Suppose s knows b communicates with a set of other agents, B, that also buy from s, where B is common knowledge. If s is found selling items below its advertised quality, buyers in B will avoid purchasing from the seller, causing the seller to indefinitely lose a total of $|B| \cdot \underline{\pi}$ utility worth of potential profit every time interval.

We assume b would prefer to not buy the item than to pay the current price for a low-quality item. Buyer b can use its knowledge of the seller's discount factor, γ_s, to evaluate whether the seller will produce an item at the advertised high quality. If b believes the seller will produce a high-quality item, then it should proceed with the purchase. The seller will produce the high-quality item if it is more profitable, if

$$\underline{\pi} > \overline{\pi} - \sum_{t=1}^{\infty} \gamma_s^t \cdot |B| \cdot \underline{\pi}. \tag{5.1}$$

In making its decision whether to purchase the item, b will also evaluate (5.1) using its current knowledge of all of the values involved. If b makes the purchase and finds the item to be of high quality, then b additionally learns that $\overline{\pi}/\underline{\pi} - 1 < \gamma_s/(1-\gamma_s)|B|$. If the item were of low quality, then the inequality operator would be reversed. Whereas b may not know the value of $\overline{\pi}/\underline{\pi} - 1$, which is effectively the percent increase in profit, with some reasoning b can still find a discount factor measurement. First, b can use a maximum likelihood estimator based on any other information b has available to find the range in profit, similar to how s may estimate the magnitude of B. Second, b can look at comments and feedback from other agents in B to see what types of products and services were offered to previous agents. If the goods

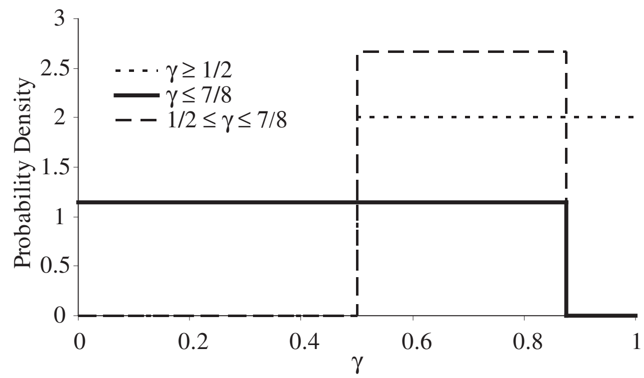


Figure 5.1: PDFs of discount factors given observations.

or services offered by s match, then it is likely that the ratio of profitability is the same, and b can substitute any change of B.

If s is a typed agent, always producing high-quality goods, then buyers' expectation of γ_s will approach the highest possible value given the valuations involved. If the difference in profit between a high-quality and low-quality item ($\overline{\pi} - \underline{\pi}$) is large, then γ_s will be observed to be close to 1. And, a typed agent producing low-quality goods will attain a low γ_s.

Example 1 *Suppose seller s is offering a high-quality item that cost it \$4 at \$5, but could substitute a low-quality item that costs it only \$1. Further suppose it is common knowledge that if the product turns out to be of low quality, the one-time buyer, b, will tell three other agents that each normally purchase one high-quality item per unit time. If b buys the item, then from (5.1), the discount factor where s would be indifferent between offering high and low quality is $\frac{5-1}{5-4} - 1 = \gamma_s/(1-\gamma_s) \cdot 3$, yielding $\gamma_s = 1/2$. Agents observing this transaction would see that b reported $\gamma_s \geq 1/2$ if s provided a high-quality item and $\gamma_s \leq 1/2$ if s provided a low-quality item. Given no other information about s's discount factor and using the maximum entropy distribution (i.e., uniform) yields an expected value of $E(\gamma_s) = 3/4$ if s provides a high-quality item, and $E(\gamma_s) = 1/4$ otherwise.*

To illustrate how trustworthiness can be aggregated, consider another potential buyer, c, reading a comment left by b of obtaining a high-quality item noting $\gamma_s \geq 1/2$, and a comment left by another buyer that $\gamma_s \leq 7/8$. If c believes these comments, c believes $\gamma_s \in [1/2, 7/8]$, with an expected value of $11/16$. Figure 5.1 illustrates the PDFs for this belief. Now, suppose c is deciding whether to buy a different item from s for \$10, and that s must decide between producing a high-quality item at a cost of \$7 or a low-quality item at a cost of \$4. Buyer c will only influence one other agent not to buy from s if it receives a low-quality item. By evaluating and simplifying (5.1) as $\frac{10-4}{10-7} - 1 > \frac{11/16}{1-11/16} \cdot 1$ yields $1 > 11/5$. Because this inequality does not *hold, c concludes s will provide a low-quality item and therefore it should not buy from s.*

5.1.2 Discount Factor and Product Choice

Measuring a buyer's trustworthiness can be important in a number of settings. If the buyer does not pay after the seller delivers the item, then the best the seller can do is refuse to sell to the buyer in the future and warn other sellers about the buyer. This sanctioning is the same as discussed in Section 5.1.1, only with roles reversed. If collusion is policed in the system, but imperfectly so, an untrustworthy buyer would be more likely to collude with other agents because it heavily discounts the utility loss of being caught. Colluding buyers could extort a seller into selling at low price because they could leverage their numbers to produce bad reviews for the seller and thus reduce the seller's future revenue. Whereas other agents may eventually discover the collusion, a large number of bad reviews could still harm some of the seller's future revenue.

We investigate one subtle method of measuring buyers' discount factors. We examine what can be inferred about a buyer's discount factor given its purchasing choice between different items. Because a buyer's valuation is private information, the results here do not give a direct measurement of the buyer's discount factor. However, the results give a constraint between the buyer's valuation and discount factor. These constraints can be used to refine existing information about an agent's valuations and discount factor.

Example 2 *Suppose agent a purchases tires for a fleet of delivery vehicles. If a purchases tires with a mean expected life of 5 years rather than tires with a mean expected life of 10 years for an additional 80% higher cost, another agent b simply cannot infer that the agent has a low discount factor. If b has a belief about a's valuations or current financial situation, b may be able to qualitatively infer that either a has a low discount factor, or a is currently in a difficult financial situation, or some combination of both situations apply. Even though a's actual state remains ambiguous to b, b still knows more about a after having observed a's choice.*

From our motivating example, we assume that the only reliability information provided is mean time to failure (MTTF), which we represent as q. The maximum entropy distribution, assuming discrete time intervals, is the geometric distribution with the cumulative distribution function (CDF) $Q(t) = 1 - (1 - 1/q)^{t+1}$, where the probability that the item will fail at each time step is $1/q$. We represent buyer b's expected utility gain from an item per unit time, that is, its willingness to pay per unit time, as, w_b. The buyer's expected utility of purchasing an

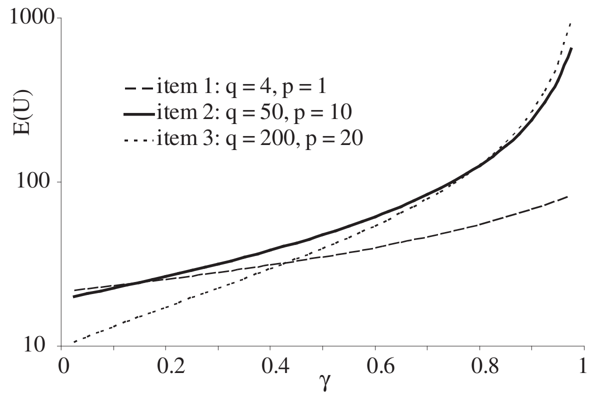


Figure 5.2: Expected utility of three purchases.

item k at price p_k with a failure rate CDF of $Q_k(t)$, $E(U_b(k))$, can be represented as

$$E(U_b(k)) = -p_k + \sum_{t=0}^{\infty} \gamma_b^t (1 - Q_k(t)) w_b. \tag{5.2}$$

Using a geometric distribution of Q_k with mean q_k thus simplifies to $E(U_b(k)) = -p_k + w_b(1 - 1/q_k)/(1 - \gamma_b(1 - 1/q_k))$.

Whereas (5.2) determines a buyer's utility for obtaining an item, it may also be used by other agents to infer information about a buyer's discount factor or willingness-to-pay. Consider a buyer, b, deciding between two items: k_1, at price p_1 with an MTTF of q_1, and k_2, at price p_2 with an MTTF of q_2. Say, b purchases k_1. If item k_1 is universally superior to item k_2, that is, it is cheaper ($p_1 \leq p_2$) and longer-lasting ($q_1 \geq q_2$), then the only information gained by other agents is that

$$E(U_b(k_1)) > 0. \tag{5.3}$$

This information can still be useful because it puts a constraint on b's possible values for its discount factor and willingness-to-pay. If some other agent, a, believes b's discount factor to be particularly low, then a can use this assumption to infer that $w_b \lesssim p$. Alternatively, if a has knowledge of b's willingness-to-pay, a can use this knowledge to gain bounds on b's discount factor by solving Inequality (5.3) for the desired variable. In some cases, such as when an agent has a high discount factor or a willingness-to-pay greater than the ask price, no further information is revealed because the bounds are less restrictive than the domain of the variable.

Now, we consider what information b would have revealed to other agents if k_1 was not universally superior to k_2. In this case, we know that $E(U_b(k_1)) \geq E(U_b(k_2))$. Solving this inequality for either γ_b or w_b can yield zero or one values, and potentially more with distributions other than geometric. Each value is the end of a boundary within which b's variable lies. Solving for these boundaries may be generalized for b choosing between multiple items. By finding the values at which b would be indifferent between each pair of items and then finding the range where k_1 yields the highest utility, another agent can obtain bounds on γ_b or w_b. Figure 5.2 shows an example of choosing between three items, where if an agent knows that $w_b = 30$, then the agent gains the knowledge that $\gamma_b \in [0.14, 0.80]$.

Because b knows its actions are monitored by other agents and b desires to have a perceived discount factor greater than its own, b has an incentive to buy an item that makes it appear as if it had a larger discount factor. Similarly, b may prefer to reveal a lower value for w_b to sellers in order to bring the price down to a lower value faster. Despite these incentives, purchasing actions must be both credible and utility maximizing for b. Except in certain seemingly rare situations, such as where excessive reliance on communication causes b to have an inflated reputation, we have generally found that an agent's optimal strategy is to play in a manner such that other agents will measure its discount factor to be in a truthful range. In our models and previous work [Hazard, 2008], the cost for an agent to over-inflate its reputation typically exceeds the benefit of being able to exploit the reputation in the future, influenced by damage that would be done to its reputation by being inconsistent.

5.1.3 Measuring Discount Factor By Price

A key benefit of using discount factors as trustworthiness is that further information can be obtained in some settings that normally would not involve trustworthiness directly. Suppose a seller, s, will be selling an item in our market model, but has uncertainty about what price it can obtain. We examine what can be learned about a seller's discount factor in a single seller, single item, single impatient buyer market. This simplification yields a negotiation, and if valuations and discount factors for both agents were all public knowledge, the agents could agree on a price without this delay [Rubinstein, 1982].

Example 3 *The website Craigslist (`http://craigslist.org`) is a good example of the scenario we formalize in this section. If an agent is selling a used snowblower in the fair weathered Los Angeles market, information on what price the market will bear would likely be scarce. The seller may believe that a few people might be looking for a snowblower for a distant vacation home in the mountains, for a prop in a movie, or just for spare parts. Using these beliefs, along with the knowledge of what a new snowblower would cost to be shipped to the LA area, the seller might start off at a moderately high price and slowly lower the price if no bids are received. The rate that the seller drops the price can be an indication of the seller's discount factor. Even if the seller undergoes a significant valuation change, such as needing to sell the snowblower*

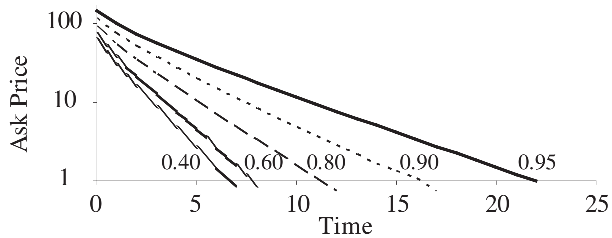


Figure 5.3: Optimal ask price schedule for a seller with different discount factors (0.40, 0.60, 0.80, 0.90, 0.95) with buyer's willingness-to-pay distribution being exponential with mean 50.

because of an unexpected move to a smaller location, examining multiple observations of price drop rates can provide information regarding the valuation changes.

Suppose we have one buyer with a willingness-to-pay of w, drawn from a probability distribution. Further suppose that either the distribution of w accounts for the buyers' discount factors or that the buyers have a low enough discount factor such that w is approximately what they would pay. The seller knows its own discount factor, γ_s. The seller can update its asking price once per unit time, its strategy being to price the item at σ_t at time t, and we denote the complete strategy as $\sigma = \{\sigma_1, \sigma_2, \ldots, \sigma_\infty\}$. The seller's expected utility, $U_s(\sigma)$, can be written as (for notational convenience, we set σ_{-1} to the supremum of the distribution of w)

$$\begin{aligned} U_s(\sigma) = &\quad \gamma_s^0 \cdot P(\sigma_0 \leq w) \cdot \sigma_0 \\ &+ \gamma_s^1 \cdot P(\sigma_1 \leq w \cap w < \sigma_0) \cdot \sigma_1 \\ &+ \gamma_s^2 \cdot P(\sigma_2 \leq w \cap w < \sigma_1) \cdot \sigma_2 \\ &+ \ldots \\ U_s(\sigma) = &\sum_{t=0}^{\infty} \left(\gamma_s^t \cdot P(\sigma_t \leq w \cap w < \sigma_{t-1}) \cdot \sigma_t \right). \end{aligned} \tag{5.4}$$

The seller's optimal strategy is that which satisfies $\operatorname{argmax}_\sigma U_s(\sigma)$. Figure 5.3 shows results of numerical solutions for the seller's optimal ask price at each time given discount factors of 0.40, 0.60, 0.80, 0.90, and 0.95. In this example, the buyer's willingness-to-pay distribution is exponential with mean 50.

To find the seller's optimal strategy analytically, assuming myopic buyers, we can view each σ_t as an independent variable and maximize the expected utility in (5.4). We express the distribution of the buyers' willingness-to-pay by the probability density function (PDF), $v(w)$, and the cumulative distribution (CDF), $V(w) = \int_{-\infty}^{w} v(x)dx$. Because the variables are mutually independent, we can maximize (5.4) by setting $\forall \sigma_t \in \sigma : dU_s/d\sigma_t = 0$. The initial case, $t = 0$, is separate from the general case yielding the equations for $t > 0$ as

$$
\begin{aligned}
0 =& \frac{dU_s}{d\sigma_0} \\
0 =& \frac{d}{d\sigma_0}\left(\sum_{t=0}^{\infty} \gamma_s^t \cdot P(\sigma_t \le w \cap w < \sigma_{t-1}) \cdot \sigma_t\right) \\
0 =& \frac{d}{d\sigma_0}\left((1 - V(\sigma_0))\sigma_0 + \gamma_s(V(\sigma_0) - V(\sigma_1))\sigma_1\right) \\
0 =& \frac{d}{d\sigma_0}\sigma_0 - \frac{d}{d\sigma_0}V(\sigma_0)\sigma_0 + \frac{d}{d\sigma_0}\gamma_s V(\sigma_0)\sigma_1 - \frac{d}{d\sigma_0}\gamma_s V(\sigma_1))\sigma_1 \\
0 =& 1 - (v(\sigma_0)\sigma_0 + V(\sigma_0)) + \gamma_s v(\sigma_0)\sigma_1 \\
\sigma_1 =& \frac{V(\sigma_0) + v(\sigma_0)\sigma_0 - 1}{\gamma_s v(\sigma_0)}
\end{aligned} \tag{5.5}
$$

and

$$
\begin{aligned}
0 =& \frac{dU_s}{d\sigma_t} \\
0 =& \frac{d}{d\sigma_t}\left(\sum_{t'=0}^{\infty} \gamma_s^{t'} \cdot P(\sigma_{t'} \le w \cap w < \sigma_{t'-1}) \cdot \sigma_{t'}\right) \\
0 =& \frac{d}{d\sigma_t}\left(\gamma_s^t(V(\sigma_{t-1}) - V(\sigma_t))\sigma_t + \gamma_s^{t+1}(V(\sigma_t) - V(\sigma_{t+1}))\sigma_{t+1}\right) \\
0 =& \frac{d}{d\sigma_t}\gamma_s^t V(\sigma_{t-1})\sigma_t - \frac{d}{d\sigma_t}\gamma_s^t V(\sigma_t)\sigma_t + \frac{d}{d\sigma_t}\gamma_s^{t+1} V(\sigma_t)\sigma_{t+1} - \frac{d}{d\sigma_t}\gamma_s^{t+1} V(\sigma_{t+1})\sigma_{t+1} \\
0 =& V(\sigma_{t-1}) - (v(\sigma_t)\sigma_t + V(\sigma_t)) + \gamma_s v(\sigma_t)\sigma_{t+1} \\
\sigma_{t+1} =& \frac{v(\sigma_t)\sigma_t - V(\sigma_{t-1}) + V(\sigma_t)}{\gamma_s v(\sigma_t)}.
\end{aligned} \tag{5.6}
$$

We assume a uniform distribution between 0 and $\overline{w}$ with the CDF expressed as $V(w) = \frac{w-0}{\overline{w}-0}$ and the PDF expressed as $v(w) = \frac{1}{\overline{w}}$. Applying (5.5) and (5.6), we find

$$\sigma_1 = \frac{2}{\gamma_{s_1}}\sigma_0 - \frac{\overline{w}}{\gamma_{s_1}}, \tag{5.7}$$

$$\sigma_0 = \frac{\gamma_{s_1}}{2}\sigma_1 + \frac{\overline{w}}{2}, \quad \text{and} \tag{5.8}$$

$$\sigma_t = \frac{2}{\gamma_{s_1}}\sigma_{t-1} - \frac{1}{\gamma_{s_1}}\sigma_{t-2}. \tag{5.9}$$

Applying recurrence relation techniques to (5.9), we introduce the exponential term variable r and solve $r^2 - \frac{2}{\gamma_{s_1}}r + \frac{1}{\gamma_{s_1}} = 0$ to find $r = \frac{1 \pm \sqrt{1-\gamma_{s_1}}}{\gamma_{s_1}}$. We can then apply the two-term linear constant coefficient homogenous recurrence relation to get

$$\sigma_t = \alpha_1 \left(\frac{1+\sqrt{1-\gamma_{s_1}}}{\gamma_{s_1}} \right)^t + \alpha_2 \left(\frac{1-\sqrt{1-\gamma_{s_1}}}{\gamma_{s_1}} \right)^t . \tag{5.10}$$

We can use the initial constants (5.7) and (5.8), to solve for α_2 from $\sigma_0 = \alpha_1 + \alpha_2 = \frac{\gamma_{s_1}}{2}\sigma_1 + \frac{\overline{w}}{2}$ and $\sigma_1 = \alpha_1 \left(\frac{1+\sqrt{1-\gamma_{s_1}}}{\gamma_{s_1}} \right) + \alpha_2 \left(\frac{1-\sqrt{1-\gamma_{s_1}}}{\gamma_{s_1}} \right) = \frac{2}{\gamma_{s_1}}\sigma_0 - \frac{\overline{w}}{\gamma_{s_1}}$ as

$$\alpha_2 = \frac{-\alpha_1(1-\sqrt{1-\gamma_{s_1}}) + \overline{w}}{(1+\sqrt{1-\gamma_{s_1}})}. \tag{5.11}$$

If a seller has not sold an item at the current time, it must drop the price in order to have any chance of selling the item at the next time. The seller will continually lower the price until the item is at the seller's willingness-to-pay. We set the seller's willingness-to-pay or cost of production to 0 for convenience (this may be simply added to all transactions). The seller's asking price therefore should be zero at infinite time. Using (5.10), we find α_1 as

$$\begin{aligned} \lim_{t\to\infty} \alpha_1 \left(\frac{1+\sqrt{1-\gamma_{s_1}}}{\gamma_{s_1}} \right)^t + \alpha_2 \left(\frac{1-\sqrt{1-\gamma_{s_1}}}{\gamma_{s_1}} \right)^t &= 0 \\ \alpha_1 \cdot \infty + \alpha_2 \cdot 0 &= 0 \\ \alpha_1 &= 0. \end{aligned} \tag{5.12}$$

Now we can use $\alpha_1 = 0$ to find α_2 from (5.11) as

$$\alpha_2 = \frac{\overline{w}}{(1+\sqrt{1-\gamma_{s_1}})}. \tag{5.13}$$

By combining the results of (5.10), (5.12, and (5.13), we can now represent the optimal price strategy as

$$\sigma_t = \frac{\overline{w}}{(1+\sqrt{1-\gamma_{s_1}})} \left(\frac{1-\sqrt{1-\gamma_{s_1}}}{\gamma_{s_1}} \right)^t . \tag{5.14}$$

We have found this result to match our numeric results as shown in Figure 5.3.

Note that the optimal ask prices decrease exponentially over time based on the discount factor at a constant rate; we have also found this numerically with an exponential distribution of w. From this information, a buyer could predict a seller's discount factor based on a small number of asks. When not in steady-state, the seller will also need to model its payoff based on its belief of the buyers' beliefs of its discount factor in case any buyers erroneously believe the seller's discount factor is significantly higher or lower than it really is. Nevertheless, this result provides a lower bound on a single seller's discount factor. Cramton [1992] analyzes a

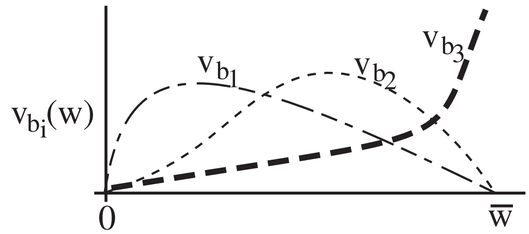


Figure 5.4: Example PDFs of buyers' willingness to pay for three buyers for a given distribution.

similar situation of delay in bargaining, except when the discount factors are publicly known and valuations are unknown.

5.1.4 Single Seller Strategy With Multiple Buyers

To extend the seller's optimal strategy to include multiple buyers we employ order statistics. Order statistics provide the probability distribution of the nth ranked value in the distribution of buyers' willingness-to-pay assuming that there are N buyers. Appendix 7.2 lists the basic order statistics formula and derivations for uniform and exponential distributions.

In the simple case where the seller only has one item, the seller only cares about the buyer with the highest willingness-to-pay, because that is the first agent that will purchase the item with descending prices. The seller has an expected PDF about the distribution of buyers' willingness-to-pay, $v(w)$, and by order statistics knows the PDF of the highest ranked willingness-to-pay, $v_{w_N}(w)$, to the lowest ranked willingness-to-pay, $v_{w_N}(w)$. By replacing the distribution $v(w)$ with the corresponding n^{th} lowest willingness-to-pay distribution, $v_{w_n}(w)$, the seller's optimal strategy can be found using the same manner as described in Section 5.1.3. Figure 5.4 shows an example of the PDFs of buyers' willingness to pay.

If the seller has multiple items, it would like to sell to each buyer at as close to the buyer's willingness-to-pay as the seller's discount factor will allow. When the seller is decreasing the prices of its items, the buyer with the highest willingness-to-pay will purchase the lowest cost item once its price is low enough. If the price of an item is decreased such that the price is below the willingness-to-pay of more than one buyer, then all such buyers will have equal probability of obtaining the purchase due to the stochastic interaction. The remaining buyers will only buy once the items are available below their willingness-to-pay, so the seller's prices should be non-increasing with respect to time. Given these points, a seller's best strategy is to price all of its items the same as it decreases the price.

We can now rewrite Equations 5.5 and 5.6 in terms of a set of N buyers. We define A_{buyer} to be a set of integers indicating buyers sorted by willingness-to-pay, with the index of 1 representing the highest willingness-to-pay and the index of N for the lowest willingness-to-pay. Since we want to maximize the seller's utility, we set $\frac{dU_{s_1}}{d\sigma_t} = 0$ again, this time including all sum over all buyers in the utility. The seller's optimal pricing strategy thus follows

$$\begin{aligned} 0 &= \sum_{n \in A_{\text{buyer}}} \Big(1 - (v_{w_n}(\sigma_0)\sigma_0 + V_{w_n}(\sigma_0)) + \gamma_{s_1} v_{w_n}(\sigma_0)\sigma_1\Big) \\ \sigma_1 &= \frac{\sum_{n \in A_{\text{buyer}}} (V_{w_n}(\sigma_0) + v_{w_n}(\sigma_0)\sigma_0 - 1)}{\sum_{n \in A_{\text{buyer}}} \gamma_{s_1} v_{w_n}(\sigma_0)} \end{aligned} \tag{5.15}$$

and

$$\begin{aligned} 0 &= \sum_{n \in A_{\text{buyer}}} \big(V_{w_n}(\sigma_{t-1}) - (v_{w_n}(\sigma_t)\sigma_t + V_{w_n}(\sigma_t)) + \gamma_{s_1} v_{w_n}(\sigma_t)\sigma_{t+1}\big) \\ \sigma_{t+1} &= \frac{\sum_{n \in A_{\text{buyer}}} (v_{w_n}(\sigma_t)\sigma_t - V_{w_n}(\sigma_{t-1}) + V_{w_n}(\sigma_t))}{\sum_{n \in A_{\text{buyer}}} \gamma_{s_1} v_{w_n}(\sigma_t)}. \end{aligned} \tag{5.16}$$

The seller will need to reevaluate and change its optimal pricing schedule whenever a sale is made, because that buyer will be removed from A_{buyer} and the seller has that more information about the buyers.

5.1.5 Seller Strategy Against Multiple Sellers

The introduction of multiple sellers motivates each sellers to race to the lowest price in order for the buyer to consider purchasing from the seller. Here, we will use M to denote the number of sellers to differentiate it from the number of buyers, N. All single-round games where $M > N$ have a Nash equilibrium of all sellers offering an initial price of 0. This is typical of an oversupplied market, where no sellers can charge for their goods because no buyers would buy at any higher price.

When $M \leq N$, the sellers must not only strategize about price, they must also account for the probability a buyer will choose their item among equally priced items from other sellers. The optimal strategy depends on the density of buyers' on the distribution of willingness-to-pay as well as the price schedule of the other agents. For example, consider two buyers and two sellers. Here we will denote the buyer with the high willingness-to-pay, b_H, as having a valuation w_{b_H} and the buyer with the low willingness-to-pay, b_L, as having a valuation w_{b_L}. We will also denote the seller with the lower discount factor as s_L and the seller with the high discount factor as s_H. Seller s_L will initiate a price schedule that descends more quickly than that of s_H. Figure 5.5 illustrates the situation.

Seller s_H's possible strategies include the following. First, s_H could use its own discount

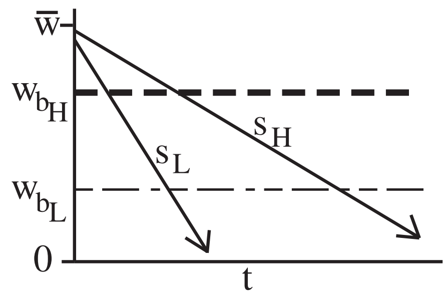


Figure 5.5: An illustration of the optimal price schedule for sellers s_H and s_L competing for buyers with high and low valuations, with expected the values for the buyers' valuations represented as dashed lines.

factor to guide its price schedule to target b_L and ignore s_L targeting b_H. A second strategy is for s_H to use a close price competition against s_L, with each agent continually offering lower prices, with each having a 50% chance of winning b_H by being the first agent to present a price below w_{b_H}. A third strategy is for s_H to employ a starkly more aggressive descending price schedule than s_L, possibly undershooting both b_L and b_H, leaving s_L with a probability of being left with the other buyer. However, by directly competing with s_L's price schedule, if s_H wins b_H, s_H will not have gotten the full price compared to if s_L were not in the market because of the larger jumps in price cuts. However, suppose that s_L and s_H have both just offered an ask price of x with $x < w_{b_H}$, leading to the first purchase. If s_H does not win b_H and $w_{b_L} > x$ is satisfied, then s_H will only receive x instead of some value in the range of $(w_{b_L}, x]$ if s_H had initially chosen to use a price schedule based on its own discount factor.

Despite the complexity of the interactions between sellers, the optimal solution and Bayes-Nash equilibrium is comparatively straightforward. Assume we are operating from the perspective of an agent s_H that has just observed another agent, s_L, which has put a low ask price indicative of a low discount factor. First, we assume that s_L will not change its price schedule in response to observing s_H's ask prices (we will revisit and challenge this assumption shortly). Seller s_H can project s_L's price schedule and figure out the probability that s_L, if acting alone, will win each buyer. The sum of these probabilities will be 1 regardless of assumptions and actual values because $M > N$. From these probabilities, s_H can find the probability that it could win each buyer, n, after s_L has sold its item; if s_L has a probability $P(s_L \text{ sells to } b_n)$, the probability that s_H could sell to this buyer is $1 - P(s_L \text{ sells to } b_n)$. These probabilities can be used to weight each term of the payoffs from Equations 5.15 and 5.16 as

$$\sigma_1 = \frac{\sum_{n \in A_{\text{buyer}}} (1 - P(s_L \text{ sells to } b_n)) \cdot (V_{w_n}(\sigma_0) + v_{w_n}(\sigma_0)\sigma_0 - 1)}{\sum_{n \in A_{\text{buyer}}} (1 - P(s_L \text{ sells to } b_n)) \, \gamma_{s_1} v_{w_n}(\sigma_0)} \tag{5.17}$$

and

$$\sigma_{t+1} = \frac{\sum_{n \in A_{\text{buyer}}} (1 - P(s_L \text{ sells to } b_n)) \cdot (v_{w_n}(\sigma_t)\sigma_t - V_{w_n}(\sigma_{t-1}) + V_{w_n}(\sigma_t))}{\sum_{n \in A_{\text{buyer}}} (1 - P(s_L \text{ sells to } b_n)) \, \gamma_{s_1} v_{w_n}(\sigma_t).} \tag{5.18}$$

By using Equations 5.17 and 5.18, s_H can find the optimal price schedule if it chooses keep a higher price than s_L. In this case, s_H keeps its price above that of s_L, and so s_L's strategy is unaffected by that of s_H, so s_L can operate optimally as if it were the only agent, affirming our assumption of s_H being able to ignore s_L's change in strategy.

Now we consider what happens if s_H instead decides to compete with s_L on price. In order for this to be s_H's optimal strategy, both agents would need to believe that $w_{b_H} \gg w_{b_L}$. By entering a *price competition*, s_H will always ask some ϵ less than s_L's most recent ask price. For competing on price to be profitable, the expected utility of winning w_{b_H}, even with the 50% chance of success, must be greater than or equal to the expected value of obtaining w_{b_L}. This can be expressed as

$$\frac{1}{2} P(s_L \text{ sells to } b_H) \cdot E(v_{w_2}(v)) > E(v_{w_1}(v)). \tag{5.19}$$

Note that this inequality is an approximation because it does not account for the actual selling price due to the rate that the price is dropped. This approximation may be made accurate by subtracting the expected difference between sales price; those terms are omitted here for readability.

The probability of s_L selling to b_H can be evaluated by finding the probability that s_L's price schedule will descend fast enough to cover the valuations of multiple buyers, and the probability that one of the other buyers will accept s_L's offer. This probability will always be greater than $\frac{1}{2}$ in this case, because one of the two buyers will accept the offer. The more patient s_L is and thus the slower the descent of its price schedule, the closer to 1 the probability will be that the buyer with the higher valuation will accept s_L's price.

Because each agent will be evaluating the same expression, albeit for the opposing agent, neither agent will be able to rationally underprice the other without credibly having a lower discount factor. Therefore, s_H's choices are to either to enter a price competition or let s_L have a better chance of winning b_H. However, the price competition will cut s_L's probability of winning the high valuation buyer in half, which will increase the descent of s_L's price schedule. This effect on s_L's price schedule must be taken into account when s_H is deciding whether to enter a price competition.

The distribution of buyer valuations and number of buyers has a major impact on Inequality 5.19. Given a patient s_L, a uniform distribution satisfies the inequality as long as $N > M$, because the worst case scenario is the range of $[0, \overline{w}]$ with two buyers, $E(v_{w_2}(v)) = \frac{2\overline{w}}{3}$ and

$E(v_{w_1}(v)) = \frac{\overline{w}}{3}$, and the ratio decreases as the number of buyers increases. A uniform distribution of buyer valuations will therefore never see direct price competition except when agents have identical discount factors.

Distributions with longer tails tend to encourage direct price competition. The ratio of $\frac{E(v_{w_1}(v))}{E(v_{w_2}(v))}$ for the exponential distribution fails Inequality 5.19 until 4 or more buyers are present, again assuming a patient s_L. The Pareto distribution, a power-law distribution, never satisfies the inequality due to the high drop in the expected values of the order statistics. In such a long-tailed distribution, the two agents would be in price competition except for rare situations when all the buyers have a similar price and the agents have participated long enough to discover that information.

Now, consider a third seller, s_3, entering the market. The first two sellers can be in one of two states, either s_H has consistently higher prices than s_L or they are engaged in a price competition. If neither of the first two agents are engaged in a price competition, then, due to Inequality 5.19, s_3 will not enter a price competition and instead will also use the Equations 5.17 and 5.18 to find its optimal price schedule. This behavior will hold for the remainder of the sellers.

On the other hand, if s_H and s_L are engaged in a price competition, then s_3 must decide if it will join the price competition or find its own price schedule targeted at buyers with lower valuations. However, once b_H is removed from the market, the remainder of the sellers will be competing for the remainder of the buyers. To make the decision about whether a price competition is profitable, we can generalize Equation 5.19 (using the same approximation for readability here), given a set of agents currently involved in a price competition, A_{PC}, to

$$\frac{1}{|A_{PC}|} P(\exists a \in A_{PC} \text{ sells to } b_N) \cdot E(v_{w_N}(v)) \geq E(U_{s_1} |\ b_N \text{ and seller in } A_{PC} \text{ removed}) \quad (5.20)$$

Entering the price competition will again steepen the descent of the price schedule as set by the agent with the lowest discount factor engaged in the price competition, which must be accounted for in the probability that the winner will sell to the highest valuation buyer. The effect on the price schedule is reduced with each additional agent entering the price competition, as the expected value of winning goes from $\frac{1}{|A_{PC}|}\sigma_t$ to $\frac{1}{|A_{PC}|+1}\sigma_t$.

If yet another agent joins the price competition after s_3 with a discount rate lower than any agent in the price competition, the new agent will effectively set the new price schedule, with the other agents matching. Each agent will continually use any new information to reevaluate its position and determine if Inequality 5.20 still holds. The price competition may continue after one of the sellers and the buyer with the highest valuation both leave the market.

5.1.6 Seller Strategy With Multiple Multiitem Sellers

To evaluate the optimal strategy with multiple sellers, each seller potentially having multiple items for sale, the results of Sections 5.1.5 and 5.1.4 can be combined. We again assume more buyers than items, otherwise the only Nash equilibrium becomes to price all items at 0. Because determining the optimal strategy does not reduce to a simple closed form solution, we instead present a process for arriving at the optimal strategy. We again assume myopic buyers.

Suppose seller s_1 is looking to sell some number of items, K_{s_1}. The seller first examines the other agents' sell offers. If s_1 knows that other sellers have additional yet unreported items for sale but have not yet posted an offer, s_1 should assume expected values for corresponding prices for the additional items based on its knowledge of the other sellers. For each number of items that s_1 can sell in the range of $k \in [0, K_{s_1}]$, s_1 must evaluate entering k items in a price competition with the fastest descending price. For each of those possible combinations, s_1 must then evaluate entering the remaining items in every possible combination of entering or not entering price competitions with every remaining descending price, and repeat this process for all price schedules. Given J other price schedules offered by other agents, this results in a total of $\binom{K_{s_1}+J-1}{J-1}$ evaluations, which includes the possibility of entering no price competitions.

To evaluate the optimal price schedule for those items not in a price competition, s_1 can compute the probability that each item will be bought by each buyer based on the other agents' price schedules. Further, s_1 must determine the optimal price schedules for items after price competitions if they do not win in the price competition, based on the probability of not winning the price competition. From these price schedules, s_1 can derive the expected utility. To determine the optimal price schedules, s_1 can compute the probability of winning each buyer and price goods according to multiitem section by using Equations 5.17 and 5.18, except replacing $1 - P(s_L \text{ sells to } b_n)$ with $P(s_1 \text{ sells to } b_n)$.

We note that multiple price wars may be occurring simultaneously, with some price descents steeper than others. For example, several sellers with low discount factors may be in a price competition targeting the highest buyer, whereas several other sellers may be in a price competition targeting a buyer with a lower willingness-to-pay with a more gradual price schedule.

5.1.7 Measuring Discount Factor By Delay

Like the sellers, each buyer has its own discount factor and is trying to maximize its utility. This section also focuses on just one buyer and one seller. We model what the seller can learn about an interested buyer's discount factor and willingness-to-pay, assuming both are constant over time, if the buyer does not purchase at the current asking price, but waits for the seller to lower the price.

A buyer's utility, $U(t)$ is a function of the time it accepts a seller's offer of price σ_t. The buyer's willingness-to-pay, w_b, and discount factor, γ_b, can be used to write its utility as

$$U_b(t) = \gamma_b^t (w_b - \sigma_t). \tag{5.21}$$

The buyer will have the opportunity to continually reevaluate its optimal time to accept the seller's offer, but the optimal absolute time does not change. This can be seen for some time offset, x, as $U_b(t+x) = \gamma_b^{t+x}(w_b - \sigma_{t+x}) = \gamma_b^x \gamma_b^t (w_b - \sigma_{t+x})$. Because the comparative difference between utilities at different times is scaled by the constant based on the time difference, $\gamma_{b_1}^x$, the acceptance time that maximizes utility is the same regardless of when the buyer is reevaluating, making the optimal strategy a subgame perfect solution concept.

When a buyer makes a purchase, the seller observes that at the time of purchase, T, the buyer's utility was the largest. Because the price schedule is strictly decreasing, the decisions at $T-1$ and $T+1$ yield the tightest bounds. The corresponding inequalities are $U_b(T) > U_b(T-1)$ and $U_b(T) \geq U_b(T+1)$. If the seller does not have any information on neither the buyer's discount factor nor the buyer's utility, then the seller only observes a relationship between the two. This observed relationship can be expressed as

$$\frac{w_b - \sigma_{T-1}}{w_b - \sigma_T} < \gamma_b \leq \frac{w_b - \sigma_T}{w_b - \sigma_{T+1}}, \tag{5.22}$$

or alternatively as

$$\frac{\sigma_T - \gamma_b \sigma_{T+1}}{1 - \gamma_b} < w_b \leq \frac{\sigma_{T-1} - \gamma_b \sigma_T}{1 - \gamma_b}. \tag{5.23}$$

The seller can use its beliefs of the distributions of w or γ along with (5.22) and (5.23) to obtain a PDF of the opposite variable, as we will discuss in Section 5.2.

Competition brought by multiple buyers decreases the delays that buyers are willing to incur to wait for reduced prices from sellers. For example, if two buyers are waiting for two sellers to decrease their ask prices, and the buyer with the higher willingness-to-pay waits long enough such that the price falls below the other buyer's willingness-to-pay, then the item may be taken by the other buyer. The first buyer must then wait until the seller with the higher discount factor gradually brings its ask price down. Not only does the delay incur lost opportunity to the buyer with the higher willingness-to-pay, but the seller with the higher discount factor will use smaller price decrements and the said buyer's optimal strategy may include paying a higher price than the first item.

As the number of buyers increases in proportion to the number of items sold, the ability of a patient buyer to successfully employ strategic delay decreases. Having more buyers means that the difference between a buyer's willingness-to-pay and the next highest willingness-to-pay decreases, increasing the chance that a drop in price will bring the item within range of more buyers. In the same way that an excess of supply pushes the price of items to 0, the limit as the number of buyers goes to infinity is that the expected profit of buyers goes to 0. In this case, the market is undersupplied, and even buyers with large discount factors rationally behave as myopic buyers.

5.2 Aggregating Discount Factor Observations

Because our discount factor measurements 1) employ Jeffrey-like probability conditioning by admitting overlapping observations that do not necessarily cover the full probability space and 2) encompass the full probability space under the assumption that the measurement is accurate, we can employ Bayesian inference interchangeably with the principal of maximum entropy, obtaining the same results [Grünwald and Halpern, 2003]. This means we can use the principle of maximum entropy to find agent's initial uninformed beliefs, then use Bayesian inference to update the probability distributions representing agents' beliefs of others' discount factors and willingness-to-pay. These mathematical tools allow agents to aggregate information about other agents' discount factors and valuations from a variety of different measures, including those we discussed in Section 5.1. We generalize the aggregation of beliefs depicted Figure 5.1 across probability distributions and types of observations.

Given no a priori knowledge or beliefs about another agent's discount factor, the maximum entropy distribution is uniform on the range of $[0, 1)$. Suppose agent s observes agent b perform an action that would require b's discount factor, γ_b, to be between 0 and 3/4 inclusive. The cumulative distribution function (CDF)[1] of b's discount factor, as a function of discount factor x, is $F_{\gamma_b}(x) = P(\gamma_b \leq x) = x$, yielding $P(\gamma_b \leq 3/4) = 1$ and $P(\gamma_b > 3/4) = 0$. Using conditional probability, the new CDF in the range of $[0, 3/4]$ becomes $F_{\gamma_b}(x) = P(\gamma_b \leq x | \gamma_b \in [0, 3/4]) = P(\gamma_b \leq x \cap \gamma_b \in [0, 3/4])/P(\gamma_b \in [0, 3/4]) = 4x/3$.

If agent s observes b perform an action, but s can only observe a relationship between b's discount factor and its willingness-to-pay rather than a direct observation of either, s can still gain some information about both of b's attributes. Consider the case in Section 5.1.7, where the observed relation between the willingness to pay and discount factor follow an inequality. We rewrite the relation $\gamma_b \leq (w_b - \sigma_T)/(w_b - \sigma_{T+1})$ in a more general form to encompass other possible observations, dropping the subscripts for convenience, as $\gamma \leq h(w)$. We use the random variable H to represent a random variable on the range of h that is isomorphic to the random variable of the agent's willingness-to-pay. As long as the function h is monotonic,

[1]By standard definition, a CDF is a nondecreasing function with domain $(-\infty, \infty)$ and range $[0, 1]$. If a random variable's domain is a subset of $(-\infty, \infty)$, then the CDF is defined as a piecewise function to yield 0 below the random variable's domain and 1 above the domain.

we can map between the CDF of w, F_w, and the CDF of this transformation, F_H, for some willingness-to-pay of x using h as $F_w(x) = F_H(h(x))$. The probability density function (PDF), f_H, may be found in the usual fashion as $f_H = \frac{dF_H}{dx}$.

Given the relationship $\gamma \leq h(w)$, agent s would like to update its beliefs about the observed agent's γ and w. We use the CDFs F_γ and F_H to denote the current beliefs of γ and w respectively, and the CDFs F'_γ and F'_H to represent the beliefs after the new observation has been taken into account. By the definition of conditional probability,

$$F'_\gamma(x) = P(\gamma \leq x | \gamma \leq H) = \frac{P(\gamma \leq x \cap \gamma \leq H)}{P(\gamma \leq H)} \tag{5.24}$$

and

$$F'_H(x) = P(H \leq x | \gamma \leq H) = \frac{P(H \leq x \cap \gamma \leq H)}{P(\gamma \leq H)}. \tag{5.25}$$

Simplifying, we have

$$F'_\gamma(x) = \frac{\int_{-\infty}^{x} f_\gamma(y) \cdot (1 - F_H(y))\, dy}{\int_{-\infty}^{\infty} f_H(y) \cdot F_\gamma(y) dy} \qquad \text{and} \tag{5.26}$$

$$F'_H(x) = \frac{\int_{-\infty}^{x} f_H(y) \cdot F_\gamma(y) dy}{\int_{-\infty}^{\infty} f_H(y) \cdot F_\gamma(y) dy}. \tag{5.27}$$

After observing an inequality relation between discount factor and a function of willingness-to-pay, (5.26) and (5.27) indicate how an agent's beliefs of another agent should be updated. If the observation yielded an equality relation, such as in Section 5.1.3, similar results can be derived by simply substituting equalities for the inequalities in the initial formulation, leading to the use of PDF functions in place of the CDF (and 1 minus CDF) functions in (5.26) and (5.27).

Example 4 *Agent s is selling an item as described in Section 5.1.7. Buyer b has received extremely accurate information about s from other buyers. However, s has no a priori knowledge about b other than b's willingness to pay follows an exponential distribution with mean of $1.00, yielding $F_w(x) = 1 - e^{-1 \cdot x}$. With no a priori knowledge of b's discount factor, s assumes the maximum entropy distribution, the uniform distribution, yielding $F_\gamma(x) = x$.*

The seller's initial asking price is $1.50. Given b's knowledge of the seller's discount factor, b predicts that the seller's next utility maximizing price will be $1.40. Just as s asks $1.50, b purchases the item, because b would attain more utility by purchasing the item now at $1.40 than waiting for the price to decrease further due to b's discount factor. The seller observes the second half of the inequality expressed by (5.22) as discussed earlier in this section, with $\sigma_T = \$1.60$, and $\sigma_{T+1} = \$1.40$.

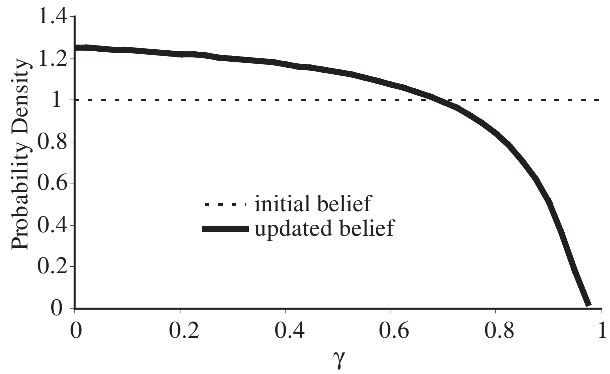


Figure 5.6: PDF of an agent's discount factor before and after an inequality observation.

By updating its knowledge via (5.26) and taking the derivative to convert to the PDF, s's belief of γ_b *is expressed by the PDF* $f_{\gamma_b}(x) = 1.38398e^{0.1/(x-1)}$*, as shown in Figure 5.6. If s makes another observation about b, s will also need to compute the updated PDF for b's willingness to pay, and use both of these functions and combine this with the new observation.*

The aggregation methods presented in this section will work in many situations, as long as the prior beliefs and observations follow the principal of maximum entropy. If noise and error in signaling are introduced, the beliefs will need to account for the probability of error. If an agent's willingness-to-pay or discount factor can change via a certain process, then the distributions must be recomputed as time progresses and the entropy must be increased according to the uncertainty from the process of change.

Chapter 6

Reputation Dynamics and Convergence

6.1 Introduction

Many authors propose desiderata to motivate their trust and reputation systems [Huynh et al., 2006, Kamvar et al., 2003, Zacharia and Maes, 2000]. However, we are unaware of a general characterization of desiderata for reputation systems that are quantitative, objective, and applicable across a wide range of domains. We present four desiderata, focusing on what quantitative properties make one reputation system more effective than another. Devising widely applicable metrics for trust is considered an important open problem [Barber et al., 2003] and is the focus of this work.

The primary purpose of a reputation system is to handle cases of *adverse selection* and *moral hazard* [Dellarocas, 2005]. Adverse selection occurs when agents have limited ability to change, for example, if a peer on a file sharing network supports limited upload bandwidth but wants other agents to believe that it may have ample upload bandwidth. In this case, other agents want to learn which agents have favorable attributes (significant upload bandwidth) so that they can choose agents with whom to interact. Moral hazard arises when one agent must reduce its utility in order to increase another's utility. An example of moral hazard is when one agent buys an item expecting it to be at or above a certain quality, but cannot measure the quality until after the purchase. Here, the seller would face the moral hazard of producing a lower quality item to reduce its costs.

When dealing with rational agents in a pure moral hazard setting, the game-theoretic approach is to devise a folk theorem, possibly modifying the model to achieve desired equilibria. The analogous approach when dealing with pure adverse selection is to use probability and statistics to determine agents' types. Although these approaches are powerful in pure scenarios, most real-world applications do not cleanly fall into one of the pure scenarios. Similarly, a reputation system designed to prevent adverse selection may not work well when faced with a moral hazard setting, and vice versa. A minority of approaches, such as that of Ramchurn et al. [2009] and Smith and desJardins [2009], combine game-theoretic results with uncertainty for specific settings.

Reputation is only meaningful if it can change over time to increase predictive accuracy in cases of adverse selection and to incentivize agents to cooperate in cases of moral hazard. Therefore, we approach reputation from a dynamic systems perspective. As our primary contribution, we motivate and formalize the following quantifiable desiderata.

Monotonicity. Agents who would provide favorable interactions should acquire better reputations than agents who would provide less favorable interactions. For example, a seller who always offers high-quality items at a low price should have a better reputation than an agent who produces defective items that it advertises as being of high-quality (and thus sells at a high price).

Accuracy. Reputation measurements should be accurate regardless of prior beliefs. For example, if a buyer incorrectly believes that a seller produces high-quality items, the buyer should quickly learn an accurate reputation value for the seller.

Convergence. Agents' reputations should converge quickly. For example, it is preferable to be able to learn after a smaller number (rather than a greater number) of interactions whether a seller offers high or low-quality products, regardless of past beliefs, provided the seller keeps to its type.

Unambiguity. An agent's reputation should be asymptotically unambiguous, meaning an agent's asymptotic reputation should be independent of any a priori beliefs about the agent held by some observing agent. An unambiguous reputation system would, as the number of interactions tends toward infinity, always yield the same reputation for a given agent regardless of the specific interactions. Consider two otherwise identical buyers (identical in their valuations for goods of a given quality, utility functions, capabilities, influence over peers, and so on) initially disagreeing about a seller's reputation. Both buyers should converge to an agreement about the seller's reputation after a sufficiently large number of interactions, assuming the seller behaves steadily in the same manner with each buyer.

Our desiderata apply to both adverse selection and moral hazard, with or without the propagation and aggregation of reputation or trust information. The measurements from the desiderata can answer a wide range of questions, such as whether agents would benefit from using a specific system, how stable the system is, and how quickly agents can build up or lose their reputation. Our desiderata evaluate reputation systems as they are situated in a given environment with a given set of agents and their preferences (i.e. utility functions). A given reputation may perform well with respect to our desiderata in one environment and poorly in another environment.

Rather than examine and compare reputation systems against a list of possible attacks, as some researchers have recently done [Huynh et al., 2006, Kerr and Cohen, 2009, Kamvar et al., 2003], we look at general dynamical properties of the system as affected by strategic agents. Our desiderata are useful across many types of reputation systems, regardless of whether the reputation system combines moral hazard with adverse selection, involves interactions in less clearly defined environments, or how difficult it is to solve analytically. Both moral hazard and adverse selection are important aspects of trust. However, most of the attention of related trust and reputation literature has focused on adverse selection, and relatively little work has been done evaluating trust and reputation systems with respect to moral hazard. We therefore focus this chapter primarily on moral hazard to address this gap.

Throughout this chapter, we distinguish two roles that an agent plays in a reputation system. An agent is a *rater* when evaluating others and is a *target* when it is being evaluated. An agent may take on both roles of target and rater simultaneously, but for clarity, we refer to the agents as target and rater in the context of the interaction being discussed.

We apply our desiderata to a diverse group of trust and reputation mechanisms from the literature. Our desiderata require a utility model, so we have chosen reputation systems that either explicitly define agents' utilities or can be augmented with a utility without further significant assumptions. In each case, we pair off a rational target against a rater as defined by the specific trust or reputation mechanism. We primarily focus on the interaction between two agents, but we examine a few larger settings.

We find that the desirable and undesirable behaviors vary across the mechanisms, validating that our desiderata are granular enough to distinguish differences between models. The general mechanism proposed by both Hazard [2008] and Smith and desJardins [2009] exhibits the most favorable results of those studied when faced with pure moral hazard, although this mechanism does not adapt to a continuous range of behaviors as easily as some other systems do. Moral hazard was more prominently considered in the design of this reputation system when compared to the others we examined, so it is not surprising that this reputation system performs best with respect to our desiderata in a moral hazard situation.

We make an additional contribution in this chapter. In order to treat each model as a black box, we present a common conceptual interface for reputation systems. This interface consists of two functions reflecting the two fundamental features of a reputation system.

An update function, used by a rater or central reputation system, which returns a target's new reputation (when participating in the reputation system under consideration) after the target has performed a specified action.

A payoff function, which returns the reward that a target can expect (under the reputation system under consideration) for performing a specified action given its current reputation.

In essence, each reputation system implements the above two functions. In our study we make use of these functions as a abstract interface to uniformly incorporate the various reputation systems.

We find that the main limitations of our methods are the computational complexity of finding the optimal strategy for the strategic agent and applying the model to reputation systems that are tightly coupled with complex interaction systems. The results of our desiderata are sensitive to the environment and our desiderata require an explicit utility model for the agents.

The remainder of this chapter is organized as follows. First, we discuss the related work in comparing reputation systems in Section 6.2. We formalize the agent interactions and desiderata of a reputation system in Section 6.3 and include a discussion on dynamical systems theory. Next, in Section 6.4, we formalize our method of applying our desiderata. We describe a basic interaction model with moral hazard in Section 6.5 and use that model to evaluate and compare reputation systems from related literature. In Section 6.7, we discuss the strengths and limitations of our desiderata, along with details of how they may be applied to different systems. Finally, we present our conclusions in Section 6.8.

6.2 Current Methods of Evaluating Reputation Systems

The ART testbed [Fullam et al., 2005] is a domain-specific problem for the domain of art purchases designed to test reputation systems. ART is useful for comparing reputation systems in a situated environment. However, the ART testbed suffers from some limitations as a general purpose test to compare reputation systems. One limitation is that the ART testbed does not always align incentives between obtaining a good reputation and increased utility [Sen et al., 2006]. The ART testbed also suffers from issues of ambiguity in agent valuations and capabilities, and being limited to a small number of agents [Krupa et al., 2009]. The domain-specific models in ART are both a strength and a limitation. The strength is that ART adds a practical realism to the measure, but the limitation is that the results depend not just on agents' reputation models, but also on how agents model their interactions and the environment outside of reputation. Our methods are domain independent, isolating the dynamics of the reputation system.

Altman and Tennenholtz [2008] take an axiomatic approach to ranking systems. They prove that, in a multiagent system in the context of aggregate ratings, independence of irrelevant alternatives is mutually exclusive with transitivity. An axiomatic system can yield strong proofs, but realistic models or models with complex interactions often preclude strong results with such modeling due to intractability. Our desiderata treat a reputation system as a black box, which extends its applicability into the realm of reputation systems that use complex computations tailored to specific requirements.

Sybil attacks, that is, agents creating pseudonyms in order to artificially manipulate their or others' reputation, are a frequently studied attack on reputation systems. Resnick and Sami [2007, 2008] use an information-theoretic approach to derive worst case bounds on the damage an agent can wreak. Their method, like that of Salehi-Abari and White [2009], limits the amount of influence an agent can wield, striving for resistance to manipulation. While resistance to manipulation is not an explicit dimension in our desiderata. In order to manipulate its reputation, an agent must perform actions to change its reputation from one that represents its actual type to another that offers some advantage. If an agent's behavior leads to those kinds of reputation dynamics, then the reputation system would not perform well against our desiderata. Further, the model proposed by Resnick and Sami does not account for rational agents that include future rewards in their strategy, but rather focuses on Sybil attacks using randomized actions. Conversely, our desiderata focus on temporally strategic agents.

Besides the aforementioned exceptions, the related literature on reputation systems typically compares a performance measure, often utility, of agents under a specific set of defined attacks for each reputation system. Two surveys indicate the widespread use of this technique. Jøsang et al. [2007] enumerate attacks and other problems, as well as corresponding solutions in the literature. Hoffman et al. [2009] compare reputation systems by which particular attacks their systems address.

Of the attacks employed in the related literature, the most common are agents that behave badly a random percentage of the time [Kamvar et al., 2003, Huynh et al., 2006]; build up a reputation by behaving positively and then "spend" it by behaving badly [Srivatsa et al., 2005, Kerr and Cohen, 2009, Salehi-Abari and White, 2009]; open new accounts to reset reputation [Kerr and Cohen, 2009]; launch Sybil attacks [Kerr and Cohen, 2009, Kamvar et al., 2003, Sonnek and Weissman, 2005]; collude with other agents [Kamvar et al., 2003, Sonnek and Weissman, 2005, Srivatsa et al., 2005]; and change behavior based on the value of the transaction [Kerr and Cohen, 2009]. In contrast, instead of devising attacks solely by intuition, we examine the entire strategy space.

Some of the related work empirically compares more than one reputation system, but such studies comprise a small minority of the related work. Kerr and Cohen [2009] and Sonnek and Weissman [2005] compare several systems across a wide range of attacks, having developed reputation systems to address the weakness of others. Of the remaining literature that empirically compares reputation systems, many papers compare three or fewer other systems [Huynh et al., 2006, Salehi-Abari and White, 2009]. We postulate that this is in part due to interoperability difficulties between the reputation systems and how some papers do not adequately specify the relationship between valuations, performance, and reputation, thus requiring major assumptions about each reputation system. We address this challenge by presenting a common conceptual interface for reputation systems and discussing how some reputation systems may be implemented using the interface.

General prescriptive desiderata have also been explored in related work [Dingledine et al., 2000, Huynh et al., 2006, Kamvar et al., 2003, Zacharia and Maes, 2000] to guide interaction design and compare reputation systems. Desiderata for trust and reputation systems are not as straightforward [Dingledine et al., 2000] because trust and reputation are supplemental to *primary interaction mechanisms*. A primary interaction mechanism is one, such as a market, that affects agents' utilities directly. In order for reputation to work, agents must be long lived, ratings must be captured and distributed, and ratings from the past must guide future decisions [Resnick et al., 2000].

6.3 Reputation Dynamics

We represent the attributes of an agent, that is, its *type* including utility functions, valuations, abilities, and discount factors, as $\theta \in \Theta$. An agent may know its type and may keep aspects of their type as private information. The set of all possible agent types, Θ, is dependent upon the system under study. We make no specific assumptions about the space of Θ and simply use θ as a parameter, treating the internals of the reputation system as a black box.

We make no assumptions about how or whether an agent's type can or cannot change over time. When evaluating a reputation system, we hold agents' types constant only to measure the reputation system itself at a point in time. If an agent's type changes faster than the reputation system can measure the new type, then the reputation system will be unable to offer useful information about the agent. A reputation system that performed well with respect to our desiderata would yield reputations that correspond to the agent's current type, even when the type is changing.

The main purpose of a reputation system is to increase the accuracy of beliefs each agent has about each other agent's type. An agent's reputation is a public projection of θ, reflecting the beliefs of other agents about it. This paper focuses on how an individual rater would assess a given target, and how that rating would affect the target's ability to gain utility in the future. We use the term *reputation* because in our analysis it provides the elements of what would be the target's reputation. We denote an individual rater's belief of a target's type generically as $r \in R$. We emphasize that whereas r may include information aggregated from the system or other agents, r is the reputation of a target as viewed by a single rater. The domain of r, R, is defined by the reputation system under examination. The domain may be as simple as a nonnegative scalar or as complex as the complete set of possible interaction histories with all details. For the formalisms in this paper, we assume R to be a normed metric space [Goffman and Pedrick, 1983], whose norm function takes (nonnegative) values that represent the expected utility of an *ideally patient strategic agent*, as we describe in Section 6.3.3. However, all of the metrics and results may be applied using their discrete counterparts. We use the discrete methods when evaluating some existing reputation models.

A target's reputation is computed by measuring outcomes of direct interactions and by obtaining and aggregating other raters' experiences and beliefs. The manner by which a rater updates its ratings of a target drives the dynamics of the reputation system. If a rater a rates a target b as r_t at time t, then after a and b interact at time $t + 1$ (or a learns something about b from another rater), a will rate b as r_{t+1}. For example, suppose a currently believes b's reputation to be r_t, that b sells high-quality products. If a purchases a product from b at time $t + 1$ that turns out to be of low-quality, a updates its belief of b's reputation to r_{t+1}, that b sells low-quality products. Here $r_{t+1} < r_t$. We use r' to indicate the rating after an action or transmission of information has occurred, which is synonymous with r_{t+1} in the case of discrete time.

6.3.1 Constructing the "Next Reputation" Function, Ω

The idea of this paper is to evaluate reputation systems using a consistent methodology as follows. Given a reputation system, first determine the Ω function that maps an agent's current reputation to its next reputation. Once Ω is defined, evaluate properties of Ω to understand key properties of the reputation system, especially with regard to its dynamism and convergence when faced with a rational target.

The target chooses how to behave given the environment, its own type, and the specific reputation system employed. The idea is that the target would behave a certain way, taking its current reputation into account when evaluating its decision. This behavior would cause the rater to assess the target a certain way. Based on the specific reputation system, the rater would adjust the reputation of the target appropriately after an observation or new information. Hence the target's reputation would be mapped from its pre-action value, r, to its post-action value, r', based on the target's type, θ, the parameters of the interaction, $g \in G$, the environment,

$\psi \in \Psi$, and the reputation system, $\xi \in \Xi$. To capture the above intuitions, we define the function $\Omega : \Theta \times G \times \Psi \times \Xi \times R \mapsto R$ that represents how the reputation of a target changes after an interaction. The target's decision process is fully captured within the inputs to Ω.

To enable uniformity in assessing different reputation systems, we assume that the rater is rational and patient, and performs the (typically nonstrategic) actions as prescribed by the reputation system under examination. This means that a rater does not lie about reputations unless it is part of the process of the reputation system being examined. For simplicity, we consider the rater's utility function as a parameter of the interaction. The rater's utility function is largely governed by the payoff function, which is an input to our desiderata, either as prescribed by the reputation system, or as modeled from the interaction environment, or as is used by the actual raters in the system. This is clearly an idealization because in most settings the raters are not strategic agents. However, the idealization systems yields baseline measures of quality and enables us to compare reputation systems.

When making an observation, a rater may also pass information to other agents, either directly or through a centralized mechanism. An agent's reputation can change with respect to a given rater without a direct interaction. Other than evaluation with a couple of reputation systems in Section 6.5.4, we focus on interactions between two agents. Therefore, for clarity and brevity, we do not explicitly model asynchronous agent communication in our formalism. We leave this to be handled by the target's utility function as a change to the environment or as collapsed into an update to the target's reputation with respect to other agents.

Because a target's type includes the target's utility function and decision model, the target's action can be computed from its type and the other parameters to Ω. Therefore, Ω does not require a parameter for the target's action.

A target's decision model must include all actions available to the target. The actions depend on the interaction model employed to evaluate the reputation system. Examples of actions are whether to pay another agent, what quality of item to produce; whether to close the current account and open a new one to reset the agent's reputation; whether to lie when rating another agent; and whether to open pseudonymous accounts controlled by the target itself to manipulate its own reputation (known as Sybil attacks).

Deciding which parts of a reputation system belong in the Ω function and which parts belong in its parameters is fairly straightforward. Anything that is agent-specific, such as valuations, capabilities, initial beliefs of others' reputations (biases), and discount factors should be an attribute of θ. Anything that is common or fixed across all agents, including the processes that define costs and interactions, can be incorporated in the environment, ψ. Attributes which may change from one interaction to another should be specified in g, and the attributes' domains should be specified by the environment. The mechanisms of the reputation system itself should be incorporated into the Ω function.

Throughout this paper, we focus on the process of matching agent types to reputations and how an agent can strategically manipulate a reputation system. When evaluating reputation systems and describing our desiderata, we hold the environment, interaction, and reputation system constant. As the other parameters are held constant, we assume all else remains equal across these interactions, such as the agent relationship topology, valuations, payoffs, game parameters, and probabilities. Our desiderata treat Ω as a black box. For brevity and clarity, we therefore omit the parameters held constant and write a target's reputation update after the target makes a decision as $r' = \Omega_\theta(r)$.

6.3.2 Fixed Points and Reputation Functions

Because reputation systems are supposed to accurately measure targets' reputations, a desirable reputation system should yield stable reputations when the targets themselves remain stable. For example, a desirable reputation system should recognize a seller that provides a good product at a low price with a good reputation. Conversely, an undesirable reputation system would be one where a good seller might receive a good or bad reputation only because of luck or strategic reputation manipulation by other agents. An agent's reputation should follow its type, meaning that a stable agent's reputation should arrive at a fixed point, ideally corresponding to its type.

A fixed point of a function is where the output of the function is equal to the input. Fixed points are a cornerstone of dynamical systems theory [Devaney, 1992]. The properties of fixed points, such as whether and how they attract or repel, govern the dynamics of systems that involve feedback. A reputation is a fixed point if $r = \Omega(r)$, which means that if the reputation were to take the exact value of r, the target's reputation would remain at the same value after subsequent interactions in an unchanging environment.

The set of fixed points of Ω_θ is $\{r \in R : r = \Omega_\theta(r)\}$. We define the function χ, which yields the stable fixed point, if one exists, of a reputation system for a target of type θ, as

$$\chi(\theta) = \lim_{n\to\infty} \Omega_\theta^n(r_{\text{initial}}), \tag{6.1}$$

where Ω_θ^n means that the function Ω_θ is iterated n times. $\chi(\theta)$ depends on r_{initial}, which is the a priori belief that a rater has of a target, given that the rater has no information about the target other than the fact that the target exists. The r_{initial} value is explicitly defined in some systems, and in others it can be assumed to be the expected value over the probability distribution of possible reputations. For example, Sporas defines r_{initial} to be 0, the worst reputation in the domain of $r \in [0, 3000]$ [Zacharia and Maes, 2000]. However, the raters may have differing a priori beliefs or have misinformation about the targets, leading to differing initial reputations. The desiderata of Convergence and Unambiguity, described in Section 6.4, address these challenges by saying that a reputation system should ideally have only one fixed point and the reputation should converge toward that fixed point.

In some reputation systems, the limit expressed by $\chi(\theta)$ may not exist. This can be caused by a lack of fixed points, particularly if the domain of possible reputations includes reputations which are impossible to attain. An example of an impossible reputation is when Ω_θ contains discontinuities with sufficiently large gaps, such as a fixed point that exists at a rating value of 3.5 out of a maximum of 5 when the rating system only permits integer values. The limit expressed by $\chi(\theta)$ may also not exist if the reputation system has a repelling (unstable) fixed point and the reputation never converges to single value. When a target's reputation oscillates around a single value (i.e., the reputation system is Lyapunov stable with a periodic, toroidal, or chaotic orbit), we can use that fixed point as the value for $\chi(\theta)$ to apply our other desiderata, noting the caveat that an agent's reputation will never reach the fixed point, only approximate it. A reputation system could conceivably have multiple fixed points around which a strategic target's reputation will orbit. The appropriate value for $\chi(\theta)$ in this case is unclear and a marked weakness of the reputation system, but we have not encountered this behavior in any of the reputation systems we examined. We further examine repelling fixed points when discussing the Convergence desideratum in Section 6.4.3.

Noise in the environment or stochastic agent strategies can also prevent a reputation system from converging to a fixed point. However, given enough Monte Carlo simulations and analysis, the expected values, moments, and statistical significance can all be propagated through our framework and desiderata. Rather than finding a fixed point, the result will be a stationary stochastic process. For each interaction step, the target chooses its action and then the rater rates. Regardless of whether the rater's observation was correct, the target would have performed the same action because of its future expected behavior. The noise in the rater's observation creates a region about Ω_θ where the agent's next reputation may lie. When noise is present, the expected value of Ω_θ, $E(\Omega_\theta)$, can be used as the next reputation function.

Figure 6.1 shows an example "cobweb" diagram as used in dynamical systems theory [Devaney, 1992] for a reputation system with $R = [\underline{R}, \overline{R}] \in \Re$. Because we apply cobweb diagrams to reputation systems throughout this paper as a basis for discussion, we now briefly describe how to read such diagrams. For simplicity in graphically illustrating concepts, we focus on real scalar reputations and real scalar projections of nonscalar reputations throughout this paper, with $\underline{R}$ representing the worst possible reputation and $\overline{R}$ representing the best possible reputation. The bounds of possible reputation values depend on the reputation system and need not be finite. For our discussions of reputation systems with real scalar values, an unbounded

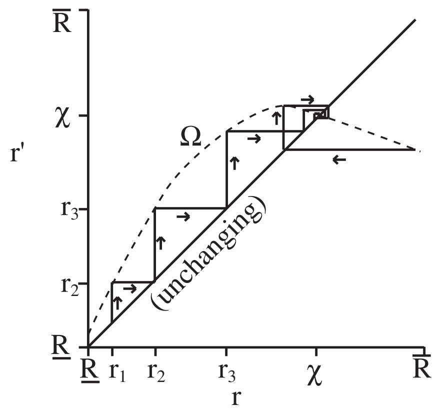


Figure 6.1: Dynamics of a reputation system. The solid horizontal and vertical lines indicate the path of the reputation along the dashed line.

maximum reputation means $\overline{R} = \infty$.

In our application of cobweb diagrams, the horizontal axis represents the target's current reputation over the domain of possible reputation values. The vertical axis represents r', with the dashed line representing the target's next reputation after performing the action as governed by its type, $\Omega_\theta(r)$. The diagonal line represents unchanging reputation and helps identify fixed points. A fixed point exists wherever an Ω_θ function intersects the diagonal line.

Figure 6.1 shows two starting points to illustrate how the reputation changes over time. Suppose a target has a bad reputation, as indicated in this illustration as a low value where the stair-step line starts on the bottom left. What constitutes a bad reputation depends on the specific reputation system (and the associated decision model of the targets), but generally we say a target has a bad reputation if another rater believes the target will likely offer poor-quality products or otherwise behave in an undesirable fashion (we return to this point in Section 6.3.3). The target's subsequent reputation, that is, the target's reputation after performing its next action, is the value on the dashed line above the horizontal position indicating the target's current reputation. This value is then used as input for the next interaction. The target begins with reputation r_1 and its strategy leads it to perform actions that lead its next reputation to be calculated as r_2—and so on, through the series of steps in the diagram. We can find each successive reputation by moving horizontally to the diagonal line and then moving vertically to the new location on the dashed line. In this example, the reputation converges to the (only) fixed point marked by χ on each axis. If the target's reputation somehow becomes higher than the fixed point in this graph, the strategic target would "expend" a small amount of its reputation, for example, by providing poor service. As a result, the target's reputation would be lowered to lie below the fixed point. However, once the reputation is below the fixed point,

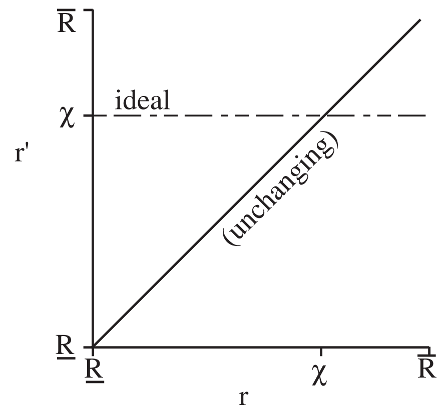


Figure 6.2: An ideal reputation system.

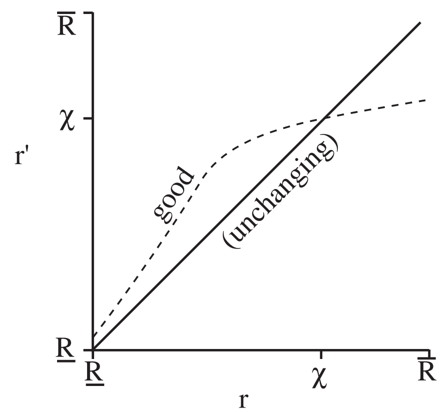


Figure 6.3: A good reputation system.

the target would behave nicely and continue to rebuild its reputation back up to the fixed point. Then it would expend it again, and so on.

Figure 6.2 depicts an ideal reputation system. The horizontal line represents the ideal case as expressed by $\Omega_\theta(r) = \chi$. This represents an ideal reputation measurement system because the reputation is measured accurately in one shot regardless of what the target's previous reputation was. This ideal case is only useful if χ depends appropriately on θ—in other words, if χ accurately reflects the type of the target. A reputation system that always returns the same reputation regardless of behavior may be perfectly precise, but it would be neither accurate nor useful.

The dashed line labeled "good" in Figure 6.3 represents a reputation system that converges to a fixed value regardless of other raters' previous beliefs. Targets whose reputations are greatly undervalued build their reputations slowly, whereas targets with overly inflated reputations slowly converge on an appropriate reputation without their reputation values bouncing around. These dynamics may be observed by using the same stepping method as described for Figure 6.1. Such a reputation system might be indicative of an e-commerce setting where targets with poor reputations charge low prices for decent-quality products and, as they build up their reputation, they begin to charge more for their offered level of quality. In this reputation system, if the target's reputation is overinflated, it may take advantage of the situation by possibly lowering the quality of its product slightly or raising the price, until it achieves its equilibrium fixed point reputation.

6.3.3 Agent Behavior

The key concepts in this paper, particularly the desiderata introduced in Section 6.4, directly apply to any type of agent decision model. One example of a decision model is an agent that plays strategies based on a stochastic process. Another example is of a malicious agent whose utility function increases with the utility loss of another agent. We primarily focus on rational agents, though other agent types may be substituted in computing the desiderata. We discuss other agent types and distributions of fixed agent behaviors to Section 6.8.

When moral hazards exist in an interaction setting, *strategic agents* can be a major threat to a reputation system. A strategic agent will do whatever actions lead toward achieving a goal, and would thus exploit any mechanism or manipulate its reputation if doing so helps achieve the goal. A rational agent is a type of strategic agent that evaluates all possible future actions and payoffs, which often must be approximated due to uncertainty and computational complexity, then chooses the immediate action that will lead it to the largest total payoff (we discuss details of this for our particular experimental evaluation in Section 6.5.1). Although the resilience of a reputation system against strategic agents indicates how well the reputation system may fare in an open real-world setting, much of the related literature on reputation systems does not discuss strategic agents. Of the papers that do discuss strategic agents (e.g., Kamvar et al. [2003]), only a minority formally model strategic agents (e.g., Jurca and Faltings [2007]). A rational agent may maximize its expected utility over its expected lifespan or use intertemporal discounting. Thus a rational agent's Ω function is the path of reputation that maximizes its utility.

To consistently quantify the comparison of reputation values in relation to agent types, we focus on the case when a target is faced by an *ideally patient strategic* (*IPS*) agent. We define an IPS agent as a rational agent that is indifferent to the time of when a specific utility change will occur.

Our motivation for considering IPS agents is as follows. Since the idea of reputation is to help select agents for future interaction based on their expected future behavior, it is natural that we rate targets in a manner that places substantial weight on the future utility of the rater. Specifically, if agent a is interacting with an impatient agent b, then a may perform actions (that affect b) that would be considered socially detrimental to a patient rater. For example, consider two agents in a situation where they can gain utility only by cooperating and offering each other favors. Agent a might not provide a favor to b if a believes that b is not patient enough to return a favor to sustain a mutually beneficial long-term relationship [Hazard, 2008]. In this case, any raters (or centralized rating mechanism) should not necessarily observe a as being impatient or having a low reputation because a is simply protecting itself against b. If b were measuring the reputation of a as the target, b would be unable to distinguish between a myopically greedy target and a target that was simply protecting itself against b's behaviors. By measuring a target against an IPS rater, we can ensure the target does not need to apply any "self-defense" measures because the target has perfect knowledge that the IPS rater will not attempt to take advantage of it for short-term gain. Further, by definition, an IPS agent values a longer running good reputation more than a less patient agent.

A patient rater is also more useful for comparing reputation values than an impatient rater because a patient rater generally can differentiate a larger possible range of behaviors. This notion is supported by the economics literature (e.g., Fudenberg and Levine [1992]). Suppose b is a reference agent, a rater by which we are measuring a property of target a. If b is impatient, then b would attempt to take advantage of a whenever doing so offered a large immediate payoff, regardless of a's type and behavior. Conversely, if b is ideally patient, then b's behavior will reflect b's belief of a's type, providing a measurement of a's type.

Suppose rater b is rational and is interacting with a target a that has type θ_a. Rater b maximizes its total utility, U_b, by controlling its strategy, σ_b, which is a set containing a specific action at each time t, $\sigma_{b,t}$. At each time step, b receives utility $u(\theta_a, \sigma_{b,t})$, which is a function of b's strategy, b's type, and a's type, from which a's optimal strategy may be derived. For an IPS rater, the function u should be chosen to represent typical agents in the system, that is, to represent average valuations and capabilities, or be endowed with capabilities and valuations the designer feels represent a good benchmark for the system. In this paper, we use the same valuations across all agents, including IPS agents.

Definition 4 *We define an* ideally patient strategic agent *(*IPS *agent), b, as having an infinite time horizon such that b maximizes its average expected total utility, $E(\overline{U}_b(\theta_a))$, as a function of any agent a's type, θ_a, as the time horizon, represented by the discount factor γ, goes to infinity as*

$$E(\overline{U}_b(\theta_a)) = \max_{\sigma_b} \lim_{\tau \to \infty} \frac{1}{\tau} \sum_{t=0}^{\tau} u_{\theta_b}(\theta_a, \sigma_{b,t}) = \max_{\sigma_b} \lim_{\gamma \to 1} (1-\gamma) \sum_{t=0}^{\infty} \gamma^t u_{\theta_b}(\theta_a, \sigma_{b,t}). \quad (6.2)$$

We use an IPS agent's utility function, given an environment, interaction, and so on for ordering agent types by preference. If an IPS rater b prefers target a to target c, that means $E(\overline{U}_b(\theta_a)) > E(\overline{U}_b(\theta_c))$. In the case of simple favor transactions with pure moral hazard, the IPS rater prefers targets with higher discount factors because such targets may yield higher payoffs. For example, the IPS agent may achieve higher payoffs with a patient agent via a trigger strategy, where both agents would follow some schedule of actions and be punished for deviation [Axelrod, 2000], because a patient agent would be willing to sacrifice short-term loss to achieve the long-term gain from the schedule of actions. When agents offer products of differing quality for differing prices with pure adverse selection, the IPS agent prefers agents whose products maximize value over time.

We define the IPS agent's utility function to be the norm function for the metric space comprised of agents' reputations. Thus, if we say that one reputation is better or higher than another, this means that it provides a larger expected utility value to an IPS agent. If the utility function's range includes negative utilities, the utility function may generally be translated to the nonnegative domain, for example, by adding a constant positive value.

Evaluating the average expected total utility of an IPS agent is not necessarily always an easy task. Numerical evaluation methods are useful for approximating the limits. Because the process of backward induction generally does not apply to infinite horizon games, finding the expected utility as $\gamma \to 1$ is a viable approximation as long as the set of interactions is small enough that searching through enough plies of interactions is tractable.

6.4 Reputation System Desiderata

Reputation systems may be useful and effective even if their behaviors are not close to ideal. This section examines what makes one reputation system more desirable than another and what can render a reputation system ineffective. The results of each desideratum are highly dependent upon the interaction environment and utility functions. Therefore, reputation systems must be compared per environment, as a given reputation system may work well in one environment with one set of utility functions and poorly in another.

6.4.1 Monotonicity

Consider the line labeled *good* in Figure 6.3. The strategic target would eventually attain its fixed point reputation. However, if Ω_θ yields similar curves for all θ, a rater would not be able to distinguish among different targets based on variations in their reputation because they would all end up with the same reputation value. This may be acceptable when the target has an extremely favorable type, but if other targets' types yield the same structure, then a

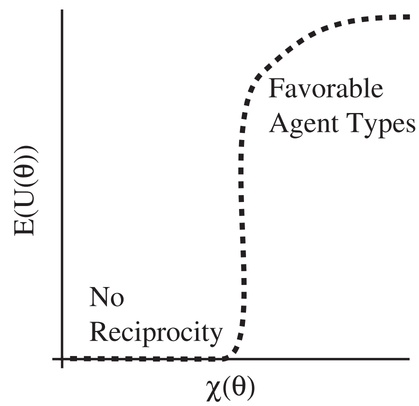


Figure 6.4: Parametric plot of $E(\overline{U}(\theta))$ and $u(\chi(\theta))$ with respect to θ.

strategic target may be able to gain a better reputation than it deserves. This is not to say that a system in which all targets achieve a good reputation is necessarily bad. A mechanism that incentivizes targets to always behave in a socially beneficial manner, regardless of their type, can be desirable. However, if target a has a better reputation than target b, then a trustworthy agent c should expect a to behave at least as well as b in interactions, all else equal, regarding c's own utility. Relating this concept back to the ideal reputation system in Figure 6.2, the horizontal line representing target a's type would be at a more desirable reputation value than that of target b's type.

One reputation is better than another if, with all else equal, the rater expects greater utility interacting with a target with the better reputation. For an IPS agent, c, entering a relationship of repeated interaction with agent a, this utility is a function of the other agent's type, θ_a, $E(\overline{U}(\theta_a))$. A regular rater, however, would not know a's type, but only its reputation, and would only evaluate a single transaction. We write a rater b's utility of entering an interaction with a as $u(\chi(\theta_a))$. The function u is the payoff function that yields the value of a single transaction for a given reputation, which is a property of the reputation system under examination.

Figure 6.4 shows how an IPS agent's utility changes with respect to the fixed point reputations of a one-dimensional agent type. In this example, the IPS agent would not interact with unfavorable agent types because they would try to reduce the IPS agent's utility for their own gain. For some values of θ, an agent may enter a mutually beneficial relationship with an IPS agent, with more favorable agents bringing greater utility to the IPS agent. If this parametric plot were not monotonic, an agent with a high reputation would have a lower expected utility to an IPS agent than an agent with a lower reputation.

Desideratum 1 MONOTONICITY*: If, to an IPS rater c, target a's type is preferable to target*

b's type, then a's asymptotic reputation should be greater than b's reputation. More formally, a reputation system is ***monotonic*** *if $\forall \theta_a, \theta_b \in \Theta : E(\overline{U}_c(\theta_a)) \geq E(\overline{U}_c(\theta_b)) \Rightarrow u(\chi(\theta_a)) \geq u(\chi(\theta_b))$. However, if, c is indifferent across all agent types, that is, $\forall \theta_a, \theta_b \in \Theta : E(\overline{U}_c(\theta_a)) = E(\overline{U}_c(\theta_b))$, then the reputation system is* ***nondiscriminatory****, a generally undesirable subset of the otherwise desirable monotonic property.*

If a reputation system is monotonic, then this means that the expected benefit an IPS agent would receive from another agent can be predicted by the agent's reputation by comparing reputation values, as measured by the utility and agent decision functions. This is important because it means that the reputation system is successful in predicting an agent's behavior, and so other agents do not need to perform expensive computations to extensively evaluate all data external to the reputation system to attempt to predict an agent's future behavior.

6.4.2 Accuracy

As in Figure 6.2, an ideal reputation system would enable a rater to assess a completely unknown target's reputation perfectly after one interaction. The closer Ω is to a horizontal line for one-dimensional reputation measures, represented by $r' = \chi$, the lower the error is between the target's current reputation and the reputation fixed point. We define this error on the domain of possible reputations, R, as follows.

Definition 5 *We define* reputation measurement error*, $\epsilon \in [0, 1]$, at some reputation r for a target of type θ as the distance between a new reputation $\Omega_\theta(r)$ and the asymptotic reputation χ, normalized with respect to the maximum distance between any two reputations, as*

$$\epsilon_\theta(r) = \frac{||\chi(\theta) - \Omega_\theta(r)||}{\max_{x,y \in R} ||\Omega_\theta(x) - \Omega_\theta(y)||}. \tag{6.3}$$

We can aggregate the reputation measurement error into the average error for a given agent type. Agents or reputation systems may have biases. Systematic or a priori biases may be known or unknown with respect to the reputation system or agents involved, and if they are known, they may be used to weight the magnitude of error by the probability of the belief when determining the average error. However, for generality, we make minimal assumptions by taking the maximum entropy approach and using a uniform distribution to weight all possible beliefs.

Definition 6 *We define* average reputation measurement error *(*ARME*), $E(\epsilon_\theta) \in [0, 1]$, as the expected value of reputation measurement error for target type θ across all possible beliefs*

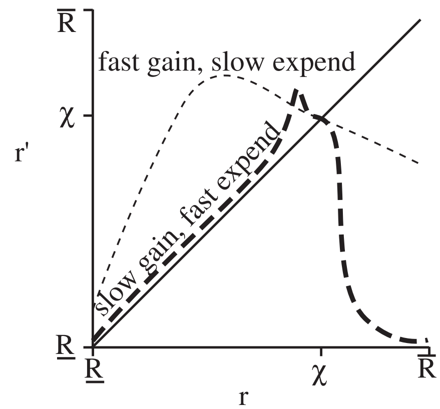


Figure 6.5: Reputation systems with different amounts of error.

of reputation, normalized over all possible reputations, R, with the exterior derivative of R, dr, as

$$E(\epsilon_\theta) = \frac{1}{\int_R dr} \int_R \epsilon_\theta(r) dr. \tag{6.4}$$

Figure 6.5 shows two reputation systems, each with one fixed point and the same derivative at the fixed point. In the reputation system shown by the line labeled *fast gain, slow expend*, targets with low reputations quickly improve their reputation, but the reputation can overshoot and would oscillate as it approaches χ. A reputation system producing the line labeled *slow gain, fast expend* would have targets gain reputation more slowly than *fast gain, slow expend*, and targets that gain overly valued reputations would quickly expend a significant amount of reputation; some targets would cause large oscillations in their reputation, possibly for a significant period of time before their reputation stabilizes, if ever. An example of *slow gain, fast expend* is the recent major Ponzi scheme by Bernard Madoff, where he had gained a strong reputation throughout his career and allegedly used his reputation to build the Ponzi scheme.[1] Qualitatively, the *fast gain, slow expend* reputation system is generally preferable to *slow gain, fast expend* because it is more stable and accurate. The ARME provides a quantitative comparison, yielding a lower error for the *fast gain, slow expend* reputation system.

If a reputation system does not provide an upper bound for reputation, for example, one that simply counts the number of positive encounters, then $\forall r \in R, ||\chi(\theta) - \Omega_\theta(r)|| < \infty : \epsilon_\theta(r) = 0$ because the denominator is ∞. This is problematic because reputation evaluations would have the same error of 0 even if the error should intuitively warrant a nonzero value due to the reputation value typically straying far from the fixed point. One solution is to use the maximum reputation achieved by any agent as the upper bound; this preserves the range of

[1] `http://www.sec.gov/news/press/2008/2008-293.htm`

$\epsilon \in [0,1]$ and thus $E(\epsilon_\theta) \in [0,1]$, keeping reputation systems on the same scale for direct comparison. However, finding the maximum achievable reputation is not always possible or practical. In such cases, the denominator in evaluating $\epsilon_\theta(r)$ may be removed, which will change the range of ϵ and $E(\epsilon_\theta)$ to both be in $[0,\infty)$. The lack of normalization in these cases can limit the usefulness of ARME in comparing dissimilar reputation systems.

Although ARME gives the error for a single target type, an important purpose behind a reputation system is to deal with different target types. One reputation system may yield low error with targets of bad reputations whereas another reputation system may yield low error with targets of good reputation. Further, a system may have mostly good or mostly bad agents, so a reputation system designer should evaluate and compare reputation systems based on the expected mix of target types.

Desideratum 2 ACCURACY*: The average reputation measurement error, $E(\epsilon)$, should be minimized with respect to the believed distribution of target types, represented by the probability density function $f(\theta)$, where $E(\epsilon) = \int_\Theta f(\theta) \cdot E(\epsilon_\theta) d\theta$.*

Accuracy represents a reputation system's resilience to misinformation. If an agent's reputation is significantly incorrect, a reputation system with good accuracy will quickly move the agent's reputation to a value which is more accurate. Accuracy is measured by the average error in a reputation system's evaluations across all possible beliefs given a strategic agent.

6.4.3 Convergence

Whereas ARME gives an indication as to how the reputation system performs across all reputations, it does not give an indication as to how the system performs when a rater's belief of another's reputation is somewhat accurate. To address this situation, we now discuss reputation dynamics around a fixed point.

A fixed point is said to be *attracting* if the dynamical system asymptotically converges to the fixed point when starting near enough to it. A fixed point may also be *repelling*, meaning that the dynamical system diverges from the fixed point unless the current value is at the fixed point. An example of a repelling fixed point is the fixed point of the line labeled *self-affirming* in Figure 6.6. If a reputation system has a single fixed point, then over time the accuracy of a target's reputation increases for an attracting fixed point and decreases for a repelling fixed point. Dynamical systems may also be attracted to or repelled from a periodic cycle of a number of points, or end up chaotic, meaning that the value jumps around within a region in an unpredictable manner [Devaney, 1992].

A fixed point can be attracting on one side and repelling on the other if Ω is tangential to the line $r' = r$ or if the derivative is not continuous at the fixed point. Systems whose

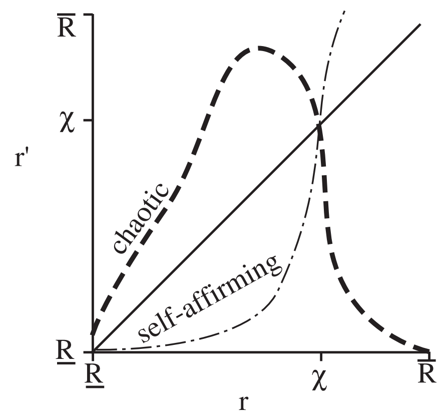


Figure 6.6: Reputation systems with large derivative magnitudes at the fixed point.

derivatives are not continuous at their fixed points can act similar to systems with no fixed points because the reputation moves in only one direction in each case. However, if a target's reputation asymptotically approaches the fixed point from the attracting side but does not cross the boundary of the fixed point, then the system can still exhibit stable reputations.

In Figure 6.6 the line labeled *self-affirming* depicts a reputation system in which a rater's eventual belief of a target is completely dependent on its initial belief due to the repelling fixed point. By tracing the feedback of this function, a reputation below χ would eventually end up at $\underline{R}$ and a reputation above χ would eventually end up at $\overline{R}$. Such a mechanism is not generally useful for measuring reputation, but may nevertheless be useful as an interaction mechanism if

- prior beliefs begin at specified values, e.g., when all agents participating in an online auction automatically start with a neutral reputation;
- better reputations incentivize targets to perform in a more socially beneficial manner, e.g., an online auction that explicitly awards higher payoffs to agents with better reputations; or
- it is otherwise effective in alleviating moral hazard, e.g., a system in which agents with a low reputation are permanently banned.

The curve labeled *chaotic* in Figure 6.6 shows a repelling fixed point that causes a target's reputation to remain persistently unstable. Below the fixed point, the target's reputation grows quickly. Once the target's reputation is above the fixed point, the target's best strategy is to take actions that quickly reduce its reputation. A reputation system exhibiting this behavior would likely be ineffective because a target's current reputation is usually meaningless with regard to its type. An example of such a system is a peer-to-peer file sharing service where agents first must upload content before they can download content. In this case, the agent must first build up its reputation by uploading files to other peers, and then the agent can expend its reputation by downloading. An agent's reputation, that is, the amount of data uploaded or downloaded, functions similar to a currency.

Whether a fixed point is attracting or repelling depends on the derivative at the fixed point [Devaney, 1992]. A fixed point is an attractor if Ω is a local contraction mapping and its Lipschitz constant, the absolute value of the minimum bound of the scaling factor between successive iterations, is less than 1. As we are looking at local dynamics, we can express this constraint on the Lipschitz constant as the maximum component of the gradient as $||\nabla\Omega(r)||_\infty < 1$ at χ, where a target's reputation eventually converges provided no other fixed points exist that change the dynamics.[2] If $||\nabla\Omega(r)||_\infty > 1$ at χ, then the fixed point repels. When multiple fixed points exist, repelling fixed points can create periodic or chaotic dynamics. If $||\nabla\Omega(r)||_\infty \approx 1$ at χ, then the reputations do not change on the fixed point. In this case, a target is incentivized to perform at its asymptotic reputation level, that is, $r \approx \Omega(r)$.

Attracting fixed points need not converge in a stable manner; a negative derivative causes a reputation to oscillate about the fixed point whereas a positive derivative approaches the fixed point from one side. The closer to zero the derivative is, the faster the reputation approaches the fixed point and the quicker the reputation gains accuracy.

Desideratum 3 CONVERGENCE*: At the fixed point,* $\chi(\theta)$*, the sequence of utility maximizing reputation values must be attracting and should converge quickly, that is,* $||\nabla\Omega(r)||_\infty\big|_{r=\chi(\theta)}$ *must be less than* 1 *and should be minimized.*

A system with good convergence means that the reputation system will yield stable reputations for agents that are internally stable, and that the error in any given agent's reputation from where it should be will decrease over time. This is measured by the rate of convergence of a strategic agent's reputation given its optimal strategy.

6.4.4 Unambiguity

Although any number of fixed points may exist for a given target type in a given reputation system, the ideal number is one. If zero fixed points exist, then the reputation values themselves are asymptotically meaningless. In order for a system to have no fixed points, one of a couple of specific situations must occur. One is if the reputation is unbounded such that a target can attain an arbitrarily high reputation and the reputation remains high even if the target behaves in a manner that should yield a low reputation. An example of a reputation system yielding

[2] For a scalar reputation, this can be expressed more simply as $|\frac{d\Omega}{dr}| < 1$.

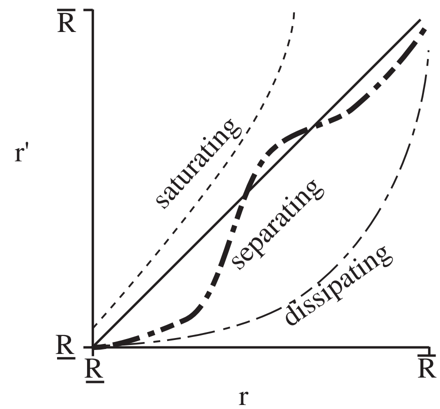


Figure 6.7: Reputation systems without meaningful fixed points.

this behavior would be one where only positive encounters were recorded; because negative encounters are ignored, a target could simply provide enough positive experiences to build a good reputation and provide many negative experiences to boost its own profit. Another case where fixed points may not exist is when Ω is discontinuous, such as a reputation system where agents are incentivized to oscillate between very good and very bad reputations.

When targets' reputations are unbounded and the mechanism has no fixed points, all targets could end up with an unboundedly growing reputation, as in the aforementioned case represented by the *saturating* line in Figure 6.7. If, for all target types, Ω is completely below the diagonal except for the lowest reputation value, as shown by the line labeled *dissipating* in Figure 6.7, then all targets would eventually end up with the worst possible reputation. The dissipating case is similar to the saturating case except that a target's reputation continually decreases. Each target's optimal strategy is to always reduce its reputation, leaving the reputation system meaningless outside of a target's a priori reputation. Because a target's reputation and thus payoff are both guaranteed to continually decrease, a reputation system with such dynamics is generally a poor choice from the standpoint of mechanism design. One real-life example of such a situation is certain vendors at tourist traps. If they provide low-quality goods, tourists do not buy from them, so they increase the sales pressure. At some level of sales pressure, enough tourists do buy from such vendors just to get the vendors to stop trying to sell to them, further incentivizing the vendors to increase pressure on selling low-quality items.

If multiple fixed points exist, then the fixed point that is asymptotically achieved depends on the rater's initial beliefs, hence the reputation is ambiguous. Consider the line labeled *separating* in Figure 6.7. If the target's reputation is above the middle of the reputation domain then the target's reputation converges to χ. If the target's reputation starts below the middle, then it continually receives a lower reputation until it reaches the lowest possible value. Note that this depends solely on the other rater's initial belief; if rater a believes target b has a low reputation, then such a reputation system exaggerates this incorrect yet self-fulfilling belief. This type of graph might be seen in the following reputation situation. Consider a manager at a business who highly values his initial opinions and does not like to be proven wrong. Further, consider some new employee that will behave in a generally consistent manner (has a fixed type). If the manager believes that the new employee will excel, the manager might give the employee many more opportunities to excel than to an employee who the manager believes will not excel. Because of the positive reinforcement in this situation, the manager's initial beliefs, even if invalid, may become a self-fulfilling prophecy.

Having multiple fixed points is not necessarily a problem for a reputation system. If a target's reputation cannot possibly get to a fixed point, the fixed point is irrelevant. On an online auction site, for example, the reputation dynamics of targets with low reputations do not matter if the site bans a seller's account if the seller's reputation drops below a certain threshold. If the reputation system depicted by the aforementioned *separating* line from Figure 6.7 starts all targets off with the maximum reputation then the targets may not ever reach the lower region because the graph changes shape accordingly with the target's type (separate diagrams could be plotted for all values of θ to demonstrate this, much like Figure 6.4). An example of a reputation system that can exhibit this kind of behavior is one that values a long positive history significantly more than recent actions. For a desirable target type, the system might have a fixed point at a high reputation and another at a low reputation. In order for a target with a desirable type to achieve the lowest fixed point, it might need to irrationally expend effort to make outcomes bad enough to diminish its reputation to the point where it is better to put no effort into the quality of its products. If the target is rational, then this fixed point will not be reached unless the target's actions are at least partially driven by a stochastic process and the target was particularly unlucky, which may happen on occasion within an environment with a large enough number of agents. In a market with significant competition, few sales, and small profit margins on products, a target with a favorable type but low reputation may not find it profitable to expend the effort required to build up to a higher reputation fixed point.

Even if a reputation system has theoretically inaccessible fixed points, in practice it does not necessarily mean that it is impossible for targets to reach this region; errors and unforeseen cases could make it possible. A shipment may be lost by an intermediate party who denies responsibility or a bug in software can cause a rating to be inaccurate with respect to the target being rated. Therefore, it is most desirable for a reputation system to have one fixed point per target type. If exactly one fixed point exists for a given target type, then the fixed point is the target's reputation. The ideally descriptive case is when the mapping between type and reputation is bijective.

Desideratum 4 UNAMBIGUITY*: A target's reputation should be* asymptotically unambiguous*, that is,* $\forall \theta \in \Theta : |\{r \in R : r = \Omega_\theta(r)\}| = 1$.

Agents' reputations in a reputation system that is asymptotically unambiguous are not ultimately dependent on a priori beliefs. In a system meeting this desideratum, an agent can never be permanently stuck with a reputation; if the agent randomly changes its underlying type, the reputation system itself is ergodic. Unambiguity is measured by counting the number of unique fixed points in the reputation path of a strategic agent starting at all possible reputations.

6.5 Empirical Results

We now apply our desiderata to some important reputation systems. We investigate the reputation measurement aspects of each system. For each system, we briefly review the reputation measure it embodies, discuss utility considerations, and then directly evaluate the reputation systems. We evaluate each reputation system using the same range of utility values with a simplified transaction model exhibiting moral hazard.

6.5.1 Experimental Method

We evaluate reputation mechanisms using a simple, stylized interaction mechanism for two reasons. First, we use a simple model to keep the problem of evaluating optimal reputation strategies tractable in order to evaluate the reputation mechanism itself; complex markets can require an intractably large number of evaluations [Fullam and Barber, 2006]. The second reason is that complex markets make it more difficult to isolate the effects of a single target's strategy [Kerr and Cohen, 2009].

In each round of our interaction model, a rational agent begins with a specified reputation. The rational agent begins in the role of target, choosing whether to offer a favor to another agent in the initial role of a rater that is operating using the reputation system being evaluated. If the rational agent offers the favor, it incurs a cost of c to itself and the other agent would receive a benefit of b. These roles are then reversed, where the other agent chooses whether to offer the rational agent a favor with the same payoffs, and the round is concluded. To show that "gains from trade" are usually possible when agents grant favors to one another, we examine these variables in the (partially overlapping) ranges of $c \in [1, 12]$ and $b \in [10, 30]$.

We evaluate each system with respect to the above desiderata against rational targets across the range of possible discount factors. A discount factor is how an agent places less value on future events than on present events. Discount factors arise from combinations of factors such as the uncertainty of a future event occurring and external methods of compounding utility (e.g., investments). Discount factors are widely employed in decision models across economics and artificial intelligence [Dellarocas, 2005, Ely and Välimäkiz, 2003, Hazard, 2008, Jurca and Faltings, 2007, Saha et al., 2003]. We employ the commonly used exponential discounting method. Using this method, each agent multiplies the expected utility of an expected future event by γ^t, where $\gamma \in [0, 1]$ is the discount factor and t is the time of the event relative to the present. Discount factors can directly affect an agent's optimal behavior and thus reputation.

As we discussed in Section 6.3.3, an agent's patience can affect both its behavior and its ability to observe behavior in others. Discount factors are a quantitative measurement of patience. A greedy target might rapidly expend its reputation, whereas a patient target may build and retain its reputation. When evaluating reputation systems, we investigate behavior across the range of possible discount factors.

In our simulations, the possible strategies of a target are a series of binary decisions. That is, each strategy is a sequence such as $\langle \text{favor}, \text{favor}, \text{nofavor}, \ldots \rangle$. We limit the length of the strategies we consider via STRATEGYDEPTH, a parameter of the simulation. We write the set of possible strategies in a regular expression notation as $\{\text{favor}, \text{nofavor}\}^{\text{STRATEGYDEPTH}}$. In our simulations, we set STRATEGYDEPTH such that the 95% of the total utility over the infinite horizon is captured with respect to the agent's discount factor, meaning STRATEGYDEPTH $= \lceil \log(1 - 0.95) / \log(\gamma) \rceil$.

Algorithm 4 ComputeNextReputation(raterModel, target, targetReputation)

bestUtility $\leftarrow -\infty$
nextReputation $\leftarrow$ targetReputation
strategySpace $\leftarrow \{\text{favor}, \text{nofavor}\}^{\text{STRATEGYDEPTH}}$
for all s $\in$ strategySpace **do**
  $\langle$util, r$\rangle$ $\leftarrow$ ComputeUtilityAndReputationFromStrategy(raterModel, target, s, targetReputation)
  if util $>$ bestUtility **then**
    bestUtility $\leftarrow$ util
    nextReputation $\leftarrow$ r
  end if
end for
return nextReputation

To find the optimal strategy for a given discount factor, we compute the utility gained for each possible strategy of the entire tree of the extended form game, as outlined in Algorithm 4. Each time the rational target is given the opportunity to decide whether to offer a favor, both decisions are followed. This algorithm approximates $\Omega_\theta(r)$ to the depth of the game tree as specified by the constant STRATEGYDEPTH. Because of the intertemporal discounting, each successive decision yields less utility, and so the utilities of infinitely long strategies may be approximated when the future expected utility falls sufficiently close to 0 with respect to the payoffs from the target's actions in the nearer future. The overall computation of this Markov decision process is exponential in the number of decisions followed. A rational target's future expected utility for a particular reputation, $U(r)$, can be expressed recursively in its Bellman equation form as

$$U(r) = \max_{\sigma} \left(u(r, \sigma) + \gamma \cdot U(N(r, \sigma)) \right), \tag{6.5}$$

where σ is the agent's action, $u(r, \sigma)$ is the utility it expects to get for a given time step, and $N(r, \sigma)$ is the agent's new reputation after it performs σ. The agent's action will be that which maximizes utility for the current reputation, r, that is, the outermost σ. Algorithm 5, which is used on line 5 in Algorithm 4, evaluates this expression for the model-specific functions GetNextReputation and GetExpectedActionPayoff to find the total utility and next reputation of a target that employs a particular strategy. Lines 5 through 8 of Algorithm 5 compute the cost that the target occurs if its strategy is to offer a favor for the given timestep, and line 14 computes the benefit that the target receives gets from the rater based on the target's reputation. Between these two payoffs, line 11 updates the target's next reputation given its current reputation and most recent action.

The functions GetNextReputation and GetExpectedActionPayoff express the entire functionality of the reputation system, encompassing the effects of multiple agents if applicable. The first function, GetNextReputation, returns the target's next reputation with respect to the rater, updated from its current reputation by the action it performs. The second function, GetExpectedActionPayoff, returns the expected payoff that the target will receive given its reputation and whatever parameters are used to determine the benefit. In our particular evaluation scenario, the payoff is independent of the target's action because the rater does not know the target's action. However, in other situations, GetExpectedActionPayoff may be a function of some information about the target's strategy, for example, if the target's action contains a publicly observable signal such as the fact that a product was shipped via an impartial third party.

It may be possible to analytically solve some of the models for the optimal solution, but others are quite complex. We thus use a brute force analysis because it works across all models. However, because this brute force analysis is costly, we do not explore the region of rational agents with the highest discount factors (above 0.90 for individual agents and above 0.60 for networks of agents). Unless an unforeseen phase change exists in any of the reputation models with discount factors greater than 0.90, we expect our results should be representative of the higher discount factors.

Algorithm 5 ComputeUtilityAndReputationFromStrategy(raterModel, target, targetStrategy, targetReputation)

```
utility ← 0
currentRep ← targetReputation
for timestep = 1 to length(targetStrategy) do
  //target plays its strategy
  if strategy[timestep] = favor then
    //if the target gives a favor on this timestep, it loses some utility
    utility ← utility − target.γ^(timestep−1) · FAVORCOST
  end if
  //rater reacts and plays its strategy according to the model
  //for example, if the target gave a favor above, the rater might respond by raising the target's reputation
  currentRep ← raterModel.GetNextReputation(currentRep, strategy[timestep])
  //depending on the target's updated reputation, the rater would reward it with a FAVORBENEFIT
  //the FAVORBENEFIT would add to the target's utility
  utility ← utility + target.γ^(timestep−1) · raterModel.GetExpectedActionPayoff(currentRep, FAVORBENEFIT)
end for
return ⟨utility, newReputation⟩
```

6.5.2 Choice of Models

Whereas many reputation systems have been proposed and studied [Ramchurn et al., 2004], little work has directly compared their effectiveness in general terms. From the body of literature, we choose systems based on the following criteria.

- Each system measures reputation and does not merely aggregate reputations without specifying how reputation is defined for a given context. Trust propagation is an important topic, but as our reputation measures examine the entire system, agents need some method of measuring trust.
- The reputation as measured by each system either explicitly characterizes the agents' utilities or can be used as a basis for making decisions regarding their interactions.
- The implementation of each system is straightforward and well-defined. This means that we identify papers that provide sufficient information to recreate their model. This also means that we sought models that did not require a large number of abstract measurements and parameters and could be applied to simple interactions without requiring a market.

- The set of systems considered is diverse. To demonstrate the generality of our approach, we consider models based on different principles and philosophies.

Whereas we omitted some models due to the above criteria, this omission does not necessarily mean that our measures cannot be applied to them. Kerr and Cohen's Trunits model [2006] requires a market whereby agents need to have some input or control with respect to their goods' prices. Our simple favor experiment would not adequately explore the Trunits reputation space, but a more complex scenario could meet this need, albeit with the requirement of further computational complexity to evaluate the optimal strategies. Similarly, Fullam and Barber's model [2006] is designed for the complex interactions in the ART testbed [Fullam et al., 2005]. Other models, such as that described by Zhang and Cohen [2007], focus on large-scale aggregation. Many of the models focusing on large-scale aggregation resemble or build upon another model that focuses on individual agents; in Zhang and Cohen's case, their model resembles the Beta model. Sierra and Debenham's information-theoretic model [2005] explicitly uses preferences rather than utilities, and is geared toward richer interactions where agents have many possible actions.

The dynamics of a reputation system are greatly influenced by the relationship between reputations, capabilities, and utilities. If a good reputation is expensive to build and maintain, but the difference in utility between having a bad versus a good reputation is small, then even trustworthy agents would not have an incentive to build up their reputation. This is analogous to diminishing returns seen by a company when improving the quality of a product that already meets the standards expected in the marketplace. For example, if agent a in a peer-to-peer environment is requesting a file transfer from agent b, agent a may not notice any difference in service if b's upload bandwidth is slightly greater than a's download bandwidth versus if b's upload bandwidth is ten times a's download bandwidth. For reputation systems (Beta and Sporas) which do not explicitly provide a utility model, we apply a utility model inspired from empirical results on online auctions.

6.5.3 Applying Desiderata to Existing Systems: Beta Model Example

The basic Beta model reputation system is a good exemplary case to apply our desiderata because the reputation mechanism itself is simple to implement and understand, yet contains a few minor hurdles with respect to applying our desiderata.

The Beta model is a frequently studied and extended reputation measure [Jøsang, 1998, Jøsang and Quattrociocchi, 2009, Teacy et al., 2006, Wang and Singh, 2006, 2007, Paradesi et al., 2009], where agents rate each experience with another agent as positive or negative. Using this method, raters quantize interactions into positive and negative experiences and use a beta distribution to indicate the probability distribution that a target will perform positively in the future. Given a number of positive interactions, α, and negative interactions, β, the expected probability that a future interaction will be positive is $\frac{\alpha}{\alpha+\beta}$, the mean value of the beta distribution. Reputation systems using this approach typically assume that agents are not rational and have an intrinsic probability of performing positively or negatively. A target's reputation is its expected probability of yielding a positive interaction.

The following steps describe the process for applying our desiderata measures to a reputation system. In each step, we use the interaction model specified in Section 6.5.1 with the basic Beta model as an illustrative example.

1. Determine the update function. For the Beta model, the update function is straightforward with respect to our interaction model. A rater rates the target positively if the target offered a favor, or negatively if the target did not. A rating, r, consists of a tuple of two nonnegative integers: the total number of positive interactions, $i_{P,r}$, and the total number of negative interactions, $i_{N,r}$. The update function, n, for the Beta model can be expressed as

$$n(r, \sigma_t) = \langle i_{P,r} + \sigma_t, i_{N,r} + (1 - \sigma_t) \rangle \, , \tag{6.6}$$

where σ_t is the strategy of the target at time t yields 1 if it will offer the favor and 0 if it will not.

When computing an agent's payoff or plotting an agent's reputation using the Beta model, we use the belief of a positive outcome, b_P, as the scalar value of an agent's reputation, as defined by Jøsang [1998]. For a given reputation r, b_P can be expressed as

$$b_P(r) = \frac{i_{P,r}}{i_{P,r} + i_{N,r} + 1}. \tag{6.7}$$

2. Determine the payoff function. Adding utility to the Beta models is relatively straightforward. Because the transactions are quantized as being positive or negative, we assume that each carries a constant utility. As reputation is the probability that interacting with the given agent will generate a positive transaction, the expected utility is simply the probability of each outcome multiplied by the utility of each outcome. From a strategic agent's perspective, the main difference between interacting with a single agent using the Beta model and a population of communicating agents using the Beta model is the number of observations any given target will have.

The exact relationship between reputation and price can be unclear in some contexts [Resnick et al., 2006], but Melnik and Alm [2003] have found a multiplicative relationship with sublinear and superlinear terms between reputation and price on eBay. To explore some reputation systems further, we apply three utility models with respect to reputation. The first is linear, meaning that a perfect reputation yields full utility, a middle reputation yields half utility, and the worst reputation yields no utility. We also investigate a sublinear relationship, where the scalar representation of the reputation, b_P in this case, when normalized to the domain $[0, 1]$ yields a utility $k \cdot b_P^2$, where k is the maximum benefit. A sublinear relationship between reputation and utility means that agents strongly favor those with high reputations. We use the relationship of $k \cdot \sqrt{b_P}$ for a superlinear relationship, which offers significant utility to all agents but those with the lowest reputations.

In our simplified interaction model, the agents alternate in granting favors. One can think of this as alternating delivery of an item by one agent, followed by payment by the other agent. With the linear relationship between reputation and utility, a target with $b_P = 0.25$ would receive half the price for a good than would a target with a $b_P = 0.5$. The utility, u, of a target of type θ for a favor at time t, can be written simply as

$$u(p_B, t, \theta) = \gamma_\theta^t \cdot b_P \cdot \text{FAVORBENEFIT}. \tag{6.8}$$

3. Integrate Update and Payoff Functions. The update function and payoff function can now be integrated into Algorithms 4 and 5, where $n(r, \sigma_t)$ and $u(p_B(r), t, \theta)$ are used for GetNextReputation and GetExpectedActionPayoff, respectively.

4. Run Algorithm 4 over Domain of Reputations. In the basic Beta model, subsequent ratings affect an agent's overall rating less than the previous. We examined a few different numbers of previous observations, but for the results reported in this paper, we used 10 previous observations. This means that we ran Algorithm 4 on each possible reputation with 10 observations, from 10 positive and 0 negative observations, through 0 positive reputations and 10 negative reputations (for models other than Beta, we divided the reputation space into 10–100 points), using a variety of cost, benefit, and discount parameters as described in Section 6.5.1. Using other total numbers of observations to cover the full two dimensions of possible data is a valid approach, but we held the magnitude constant simply to rule out the Beta model's nonstationarity (that the influence of each subsequent action has less effect on the agent's reputation), as explained in Section 6.7.

Algorithm 4 also needs to be run with various discount factors for the strategic target agent. Except when otherwise noted, we ran discount factors from 0.0 to 0.8 in 0.1 increments.

Finally, the entire set of tests needs to be run with various values of FAVORBENEFIT and FAVORCOST to determine how consistently the model behaves across the range of favor sizes. For these values, we chose several combinations across the domains of c and b as outlined in Section 6.5.1.

5. Evaluate Monotonicity.

As our interaction model is focused on moral hazard, an IPS agent would prefer to interact with an agent with a higher discount factor than with an agent with a lower one, as discussed in Section 6.3.3. With a more patient agent, the IPS agent could enter into Nash equilibria in the repeated game that have higher payoffs for both agents. The IPS could do so using a trigger strategy, not unlike the related repeated prisoner's dilemma model.

Given that the IPS agent prefers higher discount factors, we can examine whether more preferable strategic target agents have reputations that yield higher utility to raters in the interaction model. The payoff function maps the target's reputation to its utility. The payoff functions for the Beta models are strictly monotonic (linear, square root, and quadratic). We evaluate these results with respect to the ranges of FAVORBENEFIT and FAVORCOST.

If the rater's expected utilities are nondecreasing with respect to discount factor, then the reputation system is monotonic, as is the case with the Beta model with a superlinear relationship between reputation and price. If the utilities are constant, as is the case with the Beta model with linear and sublinear relationships between reputation and price, then the reputation system is nondiscriminatory. If the rater utilities ever decrease with respect to increasing discount factor, then the system is nonmonotonic. Alternatively, if no meaningful asymptotic reputation exists, then the reputation system cannot be evaluated with respect to monotonicity.

6. Evaluate Unambiguity. We find UNAMBIGUITY by first examining each pair of successive inputs, say r_i and r_{i+1}, to Algorithm 4 for a given agent type (discount factor) and environment (FAVORBENEFIT and FAVORCOST). If the line defined by $r' = r$ is crossed by (or coincides with) any two successive reputation values r_j and r_{j+1} when plotted based on their inputs (r_i and r_{i+1}), then the point of intersection is a fixed point. If zero or multiple fixed points exist (as discovered for different values of i), then the system fails UNAMBIGUITY. Otherwise, we use this unique fixed point value of r when computing the other measures. We note that if insufficient resolution is used in evaluating possible input reputations with Algorithm 4 then additional fixed points may be lost. For our results, we examined higher resolution outputs for subsets of our experimental results to make sure we were not likely missing any, though it is difficult to guarantee this numerically for reputation systems that exhibit noisy results.

7. Evaluate Accuracy. After computing the fixed point to determine UNAMBIGUITY, it is straightforward to calculate ACCURACY by computing the normalized mean absolute distance from each output of Algorithm 4 to the fixed point reputation for each agent type and environment.

8. Evaluate Convergence. Computing CONVERGENCE is also straightforward once the fixed point has been found. The slope may be closely approximated by computing the slope of the line segment between the points immediately surrounding the fixed point (or averaging the two nearby slopes if the fixed point lies on the boundary between two line segments).

6.5.4 Results

Here we discuss the results for each of the models we evaluate. We use our desiderata to compare reputation systems and find out how well they perform when faced with a strategic target agent. In doing so, we also validate that our desiderata are granular enough to distinguish differences between reputation systems, and that our results are intuitive.

Table 6.1 shows a summary of the reputation systems used and how they map into the model-specific functions for finding the next reputation and computing the expected action payoff. Table 6.2 shows a summary of our results discussed in the remainder of this section.

Results on the Beta Models

The Beta model, as described in Section 6.5.3, is the foundation for many approaches to reputation systems, including that proposed by Jøsang's [1998, 2009] *Subjective* model, Teacy et al.'s [2006] *Travos* system, and Wang and Singh's [2006, 2007] *Certainty* model. Most of the differentiation between these models is how they measure and aggregate uncertainty of reputation, but the underlying measurements are the same. We refer to this class of reputation systems as the *Beta* model.

Whereas the Beta models deal with the expected value of the probability that a target is trustworthy, many Beta models also focus on the uncertainty of this rating. This uncertainty is useful for determining whether to interact with a particular agent. Uncertainty can be an important element of decision-making for a risk-averse agent, that is, one who would prefer to avoid transactions that might have a negative outcome, even if the expected value is positive. To evaluate the effect of uncertainty as measured by the Travos and Certainty models, we reduce the utility expected from agents of uncertain trustworthiness. In the case of Travos, we multiply the expected utility by both the probability of a positive transaction and the certainty. For the Certainty model, we simply multiply the expected utility by the agent's belief value, as this accounts for the both probability of a positive transaction and the uncertainty. In both models, certainty is in the range of $[0, 1]$.

The Beta model and the Subjective model exhibit nearly identical results, and so we examine them together. This is to be expected as the Subjective model's belief is $\frac{\alpha}{\alpha+\beta+1}$. We did not examine the Subjective model's trust propagation, as it requires significant additional assumptions about beliefs of other agents' digital signatures, which is not within our present scope.

Table 6.1: Summary of reputation systems evaluated.

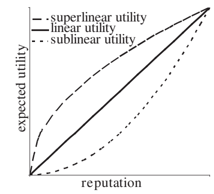
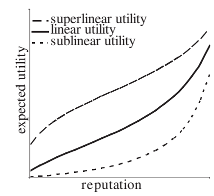
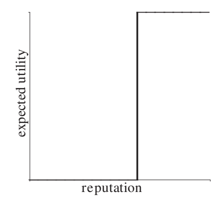
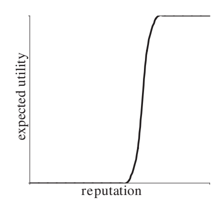
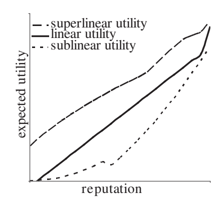
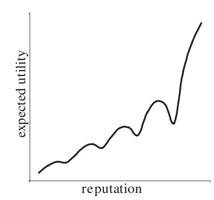
Reputation System	Next Reputation	Expected Action Payoff	Expected Action Payoff Graph
Beta	Increment the positive or negative experience count.	Exponentiate probability of positive outcome to 0.5, 1, or 2, for superlinear, linear, and sublinear, respectively, then multiply by favor value b. The graph shape is due to the straightforward expected value calculations.	
Certainty	Increment the positive or negative experience counts, compute the certainty of information.	Multiply the Certainty model's belief by b. The graph curves are due to the additional factor of uncertainty.	
Discount Factor	Measure discount factor. Update probability distribution using Bayesian inference.	If the discount factor is sufficient to sustain full reciprocity then offer full favor. The graph is a step function produced by the cutoff value of the target's expected discount factor.	
Probabilistic Reciprocity	Add all accumulated favors to compute total balance.	Multiply the favor value by the probability of offering a favor. The graph shape is due to the model's sigmoid function.	
Sporas	Exponentially dampen old rating and combine with new rating.	Use the rating normalized to [0, 1] in place of the Beta model's probability of a positive outcome. The discontinuities in the function's derivative arise due to points when the optimal strategy changes.	
Travos	Increment the positive or negative experience counts. Compute most probable bin in the Beta distribution.	Multiply the probability of a positive outcome by Travos's probability of being in the corresponding bin in the Beta distribution (ρ in their paper). Each nonmonotonicity occurs when the reputation value is near the edge of a bin.	

Table 6.2: Summary of reputation system performances; the values listed are approximate averages across our experiments.

Reputation System	Unambiguity	Monotonicity	Convergence (lower is better)	ARME (Accuracy) (lower is better)
Beta (superlinear)	yes	monotonic	0 and 0.9	0.4
Beta (linear, sublinear)	yes	nondiscriminatory	0.9	0.45
Certainty	no	–	1	–
Discount Factor	yes	monotonic	< 0.1	0.02
Prob. Reciprocity	no	monotonic	no	0.2
Sporas (superlinear, linear)	yes	monotonic	≈ 0	0.3
Sporas (sublinear)	yes	nonmonotonic	no	0.4
Travos	yes	monotonic	0.8	0.2

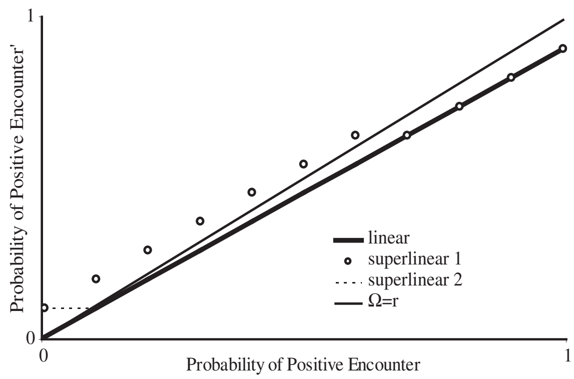


Figure 6.8: Beta and Subjective models.

The quality of the Beta model results varied by the interpretation of the probability of an agent performing positively. Using a linear interpolation of the probability, which is the natural risk-neutral way of modeling utility, led to results where no agents offered any favors and simply spent their reputations. The thick line in Figure 6.8 indicate typical results of such a linear probability-utility relationship, where all targets' reputations converged toward the minimal reputation. The sublinear results were the same as the linear. In this case, the Beta model fails Monotonicity, as all targets' reputations end up the same, which means that the all agents behave the same and that the reputation system cannot differentiate between agents. In the superlinear case, that is, where a target is either risk-seeking or is not harmed as much by negative interactions, the Beta model fares quite well. The superlinear Beta model meets Convergence with positive slopes, either slowly with slopes of 0.9 or at the ideal of 0, and also meets Monotonicity by distinguishing higher values of discount factors. The Beta model's error in Accuracy was mostly independent of the probability-utility relationship and ranged from 0.40 to 0.45. We found that the optimal strategies against the Beta model were to

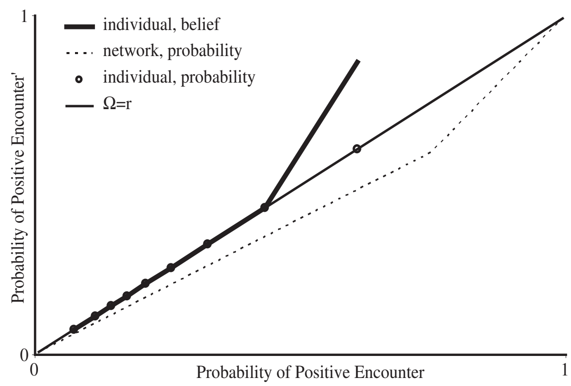


Figure 6.9: Certainty model.

continually build up reputation by acting in a favorable manner for some period and then act in an unfavorable manner afterward. The period in which the agent acted favorably depended on the parameters. When the rational agent was too impatient or the benefits were too low, the rational agent would never behave favorably.

Using the Certainty model's value of belief instead of expected value yields results that are quite different than the basic Beta distribution. Further, the characteristics of the Certainty model became more pessimistic when evaluating against a group of three raters that are communicating and aggregating ratings as opposed to an individual. To measure behavior in this network setting, we had one rational target interact with three initially identical raters, all which employ the model being examined. After each round of interactions, the raters all exchanged information about the target, and the resulting graph is the rating of one of the three raters. The target's possible action space includes all combinations of actions and agents, so a target could concievably act favorably to two agents and use its reputation to exploit a third.

The line labeled *network, probability* in Figure 6.9 shows the typical shape when a target is faced with a network of three raters. As shown by the lines labeled *individual, belief* and *individual, probability*, the targets were not incentivized to change their reputation until it crossed a critical threshold, at which point they would always perform positively. The Certainty model met neither Unambiguity nor Monotonicity, which made it difficult to assess Convergence and Accuracy.

Travos computes uncertainty by subdividing the reputation space into five equal regions, finding the region containing expected probability of trustworthiness, and measuring certainty as the probability that the reputation is within the region. Travos normalizes the magnitude of all reputation information communicated and aggregated to a rater to prevent one rater's rec-

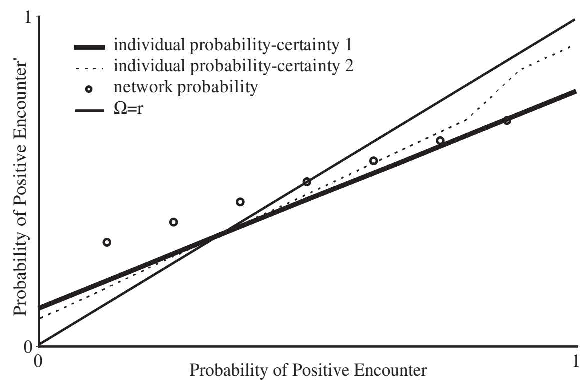


Figure 6.10: Travos model.

ommendation from strongly dominating another rater's recommendation. However, this mechanism also amplifies small numbers of observations, as the aggregation mechanism implicitly assumes a relatively large number of observations. Using Travos's certainty as a multiplicative factor in utility, Travos met MONOTONICITY, but was very close to being nondiscriminatory because most of the parameterizations yielded the same fixed point. The nondiscriminatory behavior is largely due to the normalization methods; Travos was designed for use with a significant volume of transactions rather than the small number of transactions our measures use. However, given that all of the reputations converged to the same point, the fact that Travos generally met the other desiderata does not carry significant weight in evaluating the model with smaller numbers of transactions.

Results on Probabilistic Reciprocity

Sen [2002] proposed the *Probabilistic Reciprocity* model as a way for an agent to experiment with trusting another agent to see if the first agent reciprocates favors back. Each agent keeps track of the total amount of utility spent and gained throughout the history of games between itself and other agents, summing the utilities of the gains and losses as the balance, B. Agents use this balance to adjust their probability of performing a favor to another agent. The probability function is written in terms of the cost of the current favor, c, the expected cost to offer a favor, $E(C)$, as

$$P(\text{offer favor}|B) = 1/\left(1 + \exp(\frac{c - \beta E(C) - B}{\tau})\right). \tag{6.9}$$

The parameters β and τ are tunable cooperation constants. We use balance, B, as an agent's reputation, as this is the only parameter that encodes reputation information.

The Probabilistic Reciprocity model, depicted in Figure 6.11, meets some of the desiderata

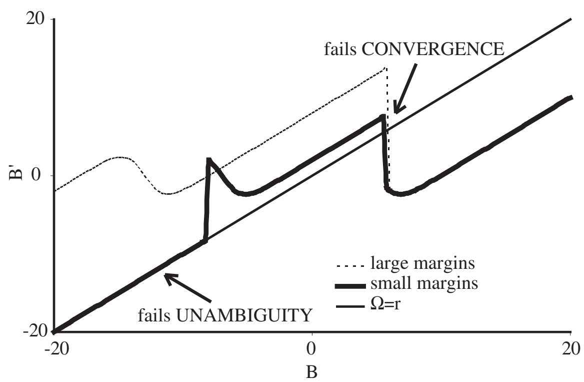


Figure 6.11: Probabilistic Reciprocity model. The model only does not meet UNAMBIGUITY for agents with poor reputations using certain marginal benefits of favors, and although the model does not meet CONVERGENCE, it is Lyapunov stable.

most of the time. Although it does not converge, it does remain Lyapunov stable. The figure shows two examples, one where the benefit of the favors is significantly larger than the costs (large margins: $c = 10$, $b = 18$), and one where the benefit is only slightly larger than the cost (small margins: $c = 10$, $b = 12$). The model generally met MONOTONICITY in every occurrence we examined, excluding ranges of fixed points where an agent's initial reputation is too low, such as the left portion of the line labeled *small margins*. Figure 6.11 shows such a range in the lower portion of the thick line. In these cases, the model fails UNAMBIGUITY because agents would refuse to consider dealing with an agent with a reputation that is too low, leaving its reputation unchanged. The weakest part of the model was CONVERGENCE, as the magnitude of the slope at the fixed point, $|\frac{d\Omega}{dr}|$, was far greater than 1 in all cases, and always negative. This means that an agent's reputation often changes significantly after every successive interaction and never converges. Finally, across our various parameterizations, the ACCURACY of the model was usually around 0.2, but was as low as 0.11 and as high as 0.22. The model's error was lowest when parameterized at moderate to large margins, such as $c = 10$ and $b = 18$, as opposed to those with highest or lowest margins (such as either $c = 10$ and $b = 12$, or $c = 10$ and $b = 30$).

Results on the Discount Factor Approaches

Hazard [2008] and Smith and desJardins [2009] both proposed variations of the *Discount Factor* model, in which agents strategically maximize utility while attempting to discover each others' discount factors. An agent's discount factor is a measure of the agent's patience, weighting how the agent accounts for future utility by an exponentially decreasing function of time. In

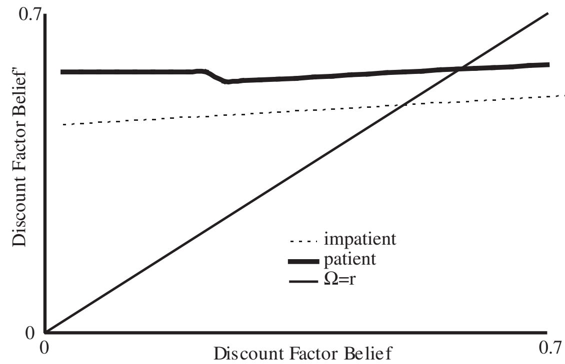


Figure 6.12: Discount Factor model.

this model, the expected value of an agent's discount factor is its reputation. An agent with a discount factor close to 0 would be myopic and greedy, whereas an agent with a discount factor close to 1 would offer favors if it expects the relationship or global reputation from offering a favor to be beneficial to itself in the long run. Like the Probabilistic Reciprocity model, the reputation of the Discount Factor model is explicitly connected with agents' utilities.

Figure 6.12 shows the results of the Discount Factor model. Across all the parameterizations we examined, the results were similar to this graph with all lines of the same shape, the only major variation being the vertical location of the line on the graph. Because the agents in the model are strategic, they choose the optimal strategy that corresponds to their discount factors. Targets cannot credibly maintain an incorrect reputation, and their reputations converge quickly. We found that agents with a higher discount factor always offer better utility to a patient agent, so MONOTONICITY is met. Each agent type also had exactly one fixed point, so UNAMBIGUITY is also met. The model fared well with the CONVERGENCE desideratum, with $\frac{d\Omega}{dr}$ being small and positive, usually less than 0.1. The error was small, and so this model performed well with regard to ACCURACY. Across all our parameterizations, the error was between 0.014 and 0.028.

Results on Sporas

Zacharia and Maes [2000] propose the *Sporas* reputation model which measures targets' reputations according to a specified range, with the rater's reputation influencing the magnitude of the reputation change. This model employs a dampening function that slows the maximum rate at which a target's reputation may change for a given observation as the target's reputation increases. Zacharia and Maes motivate Sporas based on online marketplaces and use continuous reputation values.

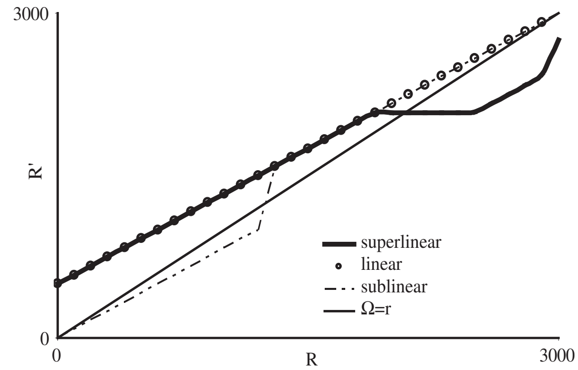


Figure 6.13: Sporas model.

The Sporas model behavior, depicted in Figure 6.13, was remarkably similar to that of the Beta model, even though the Sporas model permitted continuous interactions. Although we were initially surprised at the similarity, both models' reputation computations have a linear term of quality. Sporas model did not meet Convergence in the sublinear case or when the difference between c and b was large. The error in Accuracy for Sporas was slightly better than the Beta model, ranging from 0.20 to 0.45.

6.6 Evaluating Deployed Reputation Systems

The desiderata and analysis techniques we present can also be used to examine behavior in real-world systems that depend on reputation. Measuring how the reputation of a person or a firm changes over time can be a useful indicator as to how a reputation system realistically performs on average.

Whereas eBay is a natural setting in which to investigate online markets due to its size and prominence in the marketplace, it is less well suited to study from a reputation dynamics perspective than other portals because of their own policies of sanctioning sellers with low ratings. EBay's mechanisms strongly favor big sellers who have obtained good ratings,[3] and place restrictions, such as limited payment methods, on sellers that receive low ratings. Virtually all sellers have high ratings, with 95% of sellers having a positive reputation [Rubin et al., 2005]. The fact that eBay is moving toward an enforcement mechanism[4] is not necessarily a bad thing, as enforcement mechanisms can be incentive compatible [Braynov and Sandholm, 2002]. However, this move would render reputation dynamics less important. EBay may have

[3] http://www.businessweek.com/technology/content/jan2008/tc20080129_981043_page_2.htm
[4] http://news.bbc.co.uk/1/hi/technology/7250971.stm

Table 6.3: Example data from an Amazon seller's reputation.

Time	1	2	3	4	5	6	7	8
Rating	r_1	r_2	r_3	r_4	r_5	r_6	r_7	r_8
Rating Value	5	1	5	5	4	4	5	4
Average Rating	5.0	3.0	3.7	4.0	4.0	4.0	4.1	4.1

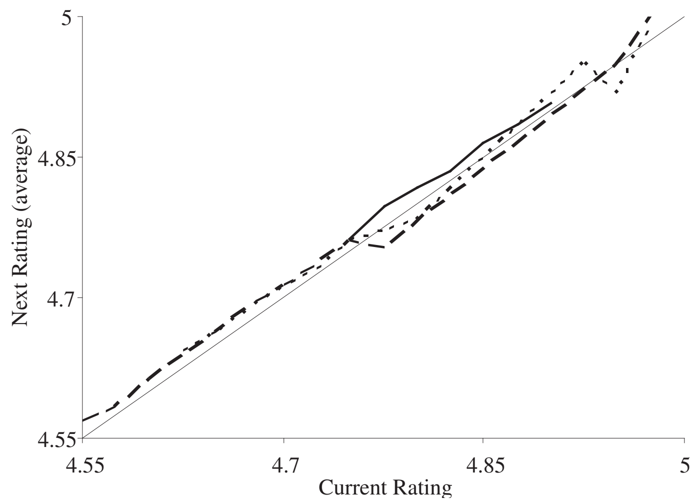


Figure 6.14: Averaged reputation dynamics of sellers who were selling the *Slumdog Millionaire* soundtrack with 50 to 500 previous sales of any item.

moved toward the enforcement method because their earlier policies and reputation mechanisms had difficulties dealing with strategic agents [Khopkar et al., 2005].

In contrast, Amazon (`http://amazon.com`) does not use such enforcement methods such that the buyers rely primarily on sellers' reputations. Amazon's reputation system averages users' ratings, where the expected value of a reputation matches the Beta model as described in Section 6.5.4. To examine Amazon's reputation system dynamics, we gathered the feedback ratings of a number of sellers. To select typical sellers, we chose two best-selling products of different kinds: the "Apple iPod touch 8 GB (2nd Generation)" and the "Slumdog Millionaire" soundtrack CD. We studied feedbacks received by sellers who had fewer than 500 reviews. As on eBay, typical sellers on Amazon do experience a selection bias toward higher reputations; buyers often avoid a seller with a poor reputation or few ratings. Table 6.3 shows an example of the data that was collected for an iPod seller.

To reduce the noise of the ratings in order to see the overall dynamics, we group the

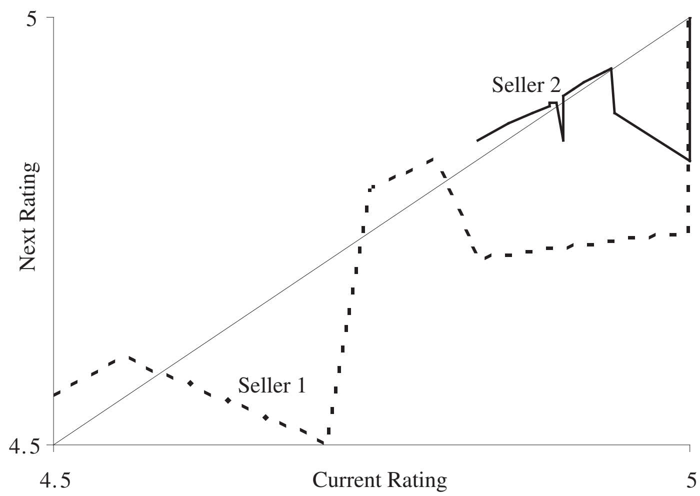


Figure 6.15: Reputation dynamics of low-volume sellers (less than 30 previous sales) who were selling Ipods.

reputations into regions of 0.025 stars, find the average, and only retain regions that consisted of three or more data points. Figure 6.14 shows the resulting next reputation function for several typical sellers. We find that many of the sellers' reputations converge on fairly linear trajectories, such as the solid thicker line above the diagonal line in the figure, each converging at a different slope toward a different intersection of the diagonal line. A number of sellers' reputations span larger ranges, staying close to the diagonal line as shown by the dashed lines. Given the averaging reputation mechanism used by Amazon, we expected these graphs to appear similar in shape to the Beta model in Figure 6.8, which they do. From these results, we can see from the graphs that the UNAMBIGUITY, MONOTONICITY, and CONVERGENCE desiderata were approximately met given the noise. We can tell that the reputation system appears to measure some attribute of the sellers because different sellers' reputations converged at different points. Because of these differing convergence points, the reputation system may meet MONOTONICITY.

The feedback history of sellers chosen from the iPod group are approximately bimodally distributed; the sellers are either high-volume sellers with tens of thousands of feedback entries or casual, low-volume sellers with less than a hundred. Figure 6.15 shows the reputation dynamics of two typical sellers from this group. Because Amazon's system uses an average rating, but does not dampen ratings over time, each feedback rating has greater influence on a user's overall reputation when a seller has few ratings. Given that the number of ratings was low, the randomness in people's experiences such as perceived quality and items damaged in transit may have added noticeable noise to the ratings. However, a number of sellers' reputations converged in a straight line over a short range, such as Seller 2 on the graph.

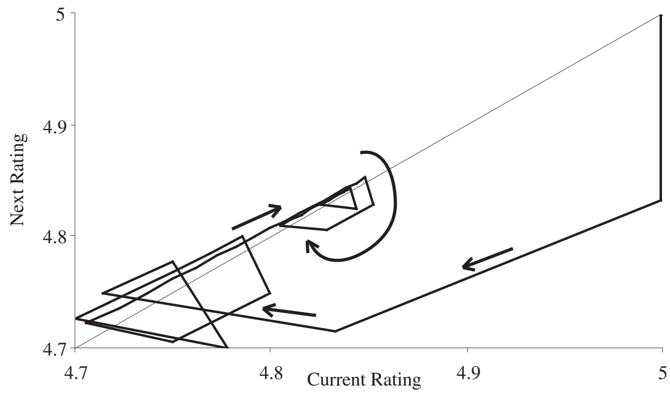


Figure 6.16: Temporally connected reputation dynamics graph of one seller who was selling the *Slumdog Millionaire* soundtrack.

The sellers of the "Slumdog Millionaire" soundtrack tend to have more sellers with feedback entries numbering in the hundreds to low thousands. These longer histories show many reputations that started at one reputation and slowly converged toward another, as shown in Figure 6.16. In this figure, the data points are connected chronologically, as opposed to being sorted by the values on the horizontal axis, in order to give a different sense of convergence given the noise in the system. The reputation starts in the upper right-hand corner after a single positive feedback and slowly spirals toward an asymptotic reputation around 4.83. Before the seller's reputation converges on this value, its reputation hovers between 4.7 and 4.8. We see this secondary attracting region for a number of the sellers, which can be due to changing market conditions, seller behaviors, buyer behaviors, and rating norms.

Determining whether the reputation system is accurately measuring a seller's type is difficult when people are involved instead of computational agents, as reputation involves many psychological aspects and societal norms. The focus of this paper is on computational agents. The results of live reputation systems with people on Amazon do show similar characteristics to our theoretical models with rational agents. We can at least conclude that our desiderata and methods of analysis are worth further investigation in practical online systems where people rely upon reputation for making real-life business decisions.

6.7 Discussion

The primary purpose of a reputation system is to provide information to agents about other agents with the goal of improving social welfare. This goal assumes that if agents know which other agents are trustworthy and which agents are likely to defraud, then the agents' utilities would be improved as agents would self-select transactions with preference toward trustworthy agents. Whereas a possible design goal is to increase trustworthy agents' utility the most, the primary goal is to inform agents and reduce uncertainty with regard to some interaction mechanism.

If an interaction mechanism has the property of incentive compatibility, then all strategic targets would always play honestly according to their valuations. Although incentive compatible systems may have use for multiagent learning, for example to determine which agents might receive the most benefit from which products, such systems have little need for an additional system to measure the reputations of targets. If a reputation measurement system were added in order to measure targets' reputations for use in an additional context or situation, incentive compatibility may no longer hold. Whereas our desiderata would work in measuring reputation in an incentive compatible mechanism, the measurements may have reduced relevance. Alternatively, if the agents in a system exhibit highly specific behavior and are not strategic, then our desiderata would need to be modified to use the specific behavior in place of the strategic behavior. Such a system can arise when interactions are only permitted via proxy agents, that is, targets that have a predefined behavior that act based on a specific set of parameters.

Our desiderata are useful measures for how well a reputation system will hold up against strategic attacks. For example, Kerr and Cohen [2009] outline a number of possible ways that an agent could strategically improve its utility by being dishonest in a reputation system. Their "reputation lag attack," achieved by a target alternating between honest and cheating periods, is applicable when a reputation system that fails to meet CONVERGENCE because a target can exploit oscillations of its reputation. Similarly, their "value imbalance attack," achieved by a target being honest with low-cost goods and dishonest with high-cost goods, and "reentry attack," where an agent continually opens new accounts to dishonestly use a new untainted reputation, both indicate that a reputation system has poor ACCURACY. A reputation system designed for high ACCURACY would recognize dishonest targets more quickly. In general, our results are consistent with those of Kerr and Cohen; they find that the Travos, Beta, or Certanty models all can be effectively exploited by various strategies.

When analyzing pure moral hazard situations, the resulting Nash equilibria are often mixed strategies, where a target chooses its actions stochastically based on some distribution. If mixed strategies are necessary or desirable for a particular reputation mechanism, then the reputation system should somehow recognize when a target is employing a mixed strategy. Detecting whether a mixed strategy is being employed has some uncertainty to it, as it must be done statistically within some bounds of confidence. If the reputation system does not collapse a mixed strategy into a single reputation value, then the expected value of an agent's next reputation can be used in place of Ω_θ.

One difficulty in evaluating a reputation measure is if the measurements are nonstationary, meaning that the reputation measures themselves somehow change over time. Nonstationarity can arise if it becomes increasingly more difficult or easy to change a reputation when further measurements are made, as is the case when interactions are aggregated over the entire lifetime of a target without any sort of dampening—i.e., without weighting old interactions less than recent ones. Whereas some reputation systems, such as Amazon Marketplace, Travos, and Certainty, employ nonstationary measures, such reputation systems must be used with caution because the difficulty of a target changing its reputation becomes increasingly difficult as a function of the target's age, as even the oldest interactions count as much as recent ones.

Our desiderata do not always indicate that one reputation system is the best one for a particular situation. The choice of which reputation system to employ comes down to trade-offs. For example, one system may offer better Convergence whereas another may offer better Accuracy. Having good Convergence means that the given system quickly reaches an equilibrium where the reputation is close to the actual value, but this can be misleading in cases when the reputation dynamics change rapidly close to the fixed point. A system's having good Accuracy means that it corrects a target's reputation to achieve a reasonably accurate value quickly, even if the initial reputation is far off. However, raters may be able to only discern a small amount of information from each transaction in some interaction mechanisms, and so the interaction model may be detrimental to Accuracy. If a system does not exhibit Unambiguity, but the unreasonable fixed points are impossible to reach by the path a target's reputation takes, then those unreasonable fixed points may be ignored. However, if unforeseen events, such as a software glitch, incorrectly push a target's reputation into these regions, then the ignored fixed points become extremely important and can possibly have major negative impacts to the reputation system as a whole.

Reputation systems may work better in one domain than another. A reputation system may work well in the case of adverse selection, but perform poorly in the case of moral hazard. The effectiveness of a reputation system may change drastically even with different parameters in the same environment, even if the only different parameter is the topology of agent relationships. Therefore, when applying our desiderata to a reputation system, they should be applied to a setting as close to the actual environment as possible. If parameters of the environment or interactions are known to change quickly or drastically, then the desiderata should be employed across the range of environments and interactions. One reputation system may perform well in a certain niche case, but may perform poorly across the full range of interactions.

6.8 Conclusions

The four desiderata we present measure how well a mechanism fares at measuring the reputation of a strategic agent. Of the systems we examined, the one that takes moral hazard as the primary environment in designing the system, the Discount Factor model, fared the best when evaluated in a moral hazard situation. The results from applying our desiderata were granular enough to differentiate reputation systems.

Our desiderata may also be evaluated with respect to a distribution of agent behaviors. We choose to use the IPS agent as the default behavior, as this is a practical worst-case scenario for designing a reputation system. At the risk of underestimating the agents' strategic capabilities, reputation systems may be evaluated with boundedly rational agents or a distribution of fixed behaviors to see how various reputation systems may perform in realistic environments.

Although many currently proposed reputation systems do not significant computation to determine agents' reputations, low computational complexity is desideratum in many situations. Adding low computational complexity to our set of desiderata would be useful if more computationally complex reputation systems are proposed and the issue becomes a concern.

Whereas our desiderata are useful for ensuring that reputation systems are useful to the agents, strictly following the desiderata is not always in the best interest of the party that implements the marketplace or mechanism—as opposed to the agents who interact with each other in context of the marketplace. A firm setting up an online auction that profits from each transaction has an incentive to maximize targets' reputations, that is, maximize Ω, such that agents perform transactions before the agents realize that not all are trustworthy. However, such a practice is not sustainable, and so a firm looking for long-term profits would need to ensure that the reputation system is useful. Such a firm would need to make trade-offs among short-term profit, long-term profit, and the various desiderata.

We explore some real-world data from Amazon. We find that the general shape of the dynamics agree with that of the underlying Beta model, given the noise in the system. The purpose of our desiderata is for assessing and designing reputation systems in multiagent systems with rational agents. Our empirical results suggest that they may be applicable to real-world settings with people, but further study is required to determine whether artifacts from human factors differ significantly from rational agents in such reputational settings.

Our desiderata are by no means exhaustive and may be modified or extended for domain-specific purposes. They make a good start toward a general framework for directly comparing the effectiveness of different reputation systems in a specified situation.

Chapter 7

Conclusion

The three overarching contributions of this work are demonstrating how important agents' rationality is when evaluating trustworthiness, showing that patience in the form of intertemporal discounting is isomorphic to trustworthiness under general assumptions, and providing a methodology to measure the dynamics of a reputation system. These three themes are interwoven throughout the chapters.

From Chapter 4, discount factors measure patience and are thus very useful strategically. An agent may be simultaneously rational and impatient, that is, have a short expected life where every game has some probability of being its last. Given rational agents, discount factors and valuations are the agent's primary endogenous attributes affecting reputation.

Strategic models of trust such as the ones we present in Chapters 3, 4, and 5 are required in open agent communities if the strategies are to be evolutionary stable, that is, resilient to invasion by undesirable strategies. While the models we present do not encompass all possible favor scenarios, they provides a foundation from which to build.

In Chapter 6, we present the first methodology for evaluating broad classes of reputation systems, particularly when faced with strategic agents. We employ dynamic systems theory to find what behaviors a reputation system incentivizes, and use our methodology to compare a variety of reputation systems from the related literature. We find that our discount factor model of trustworthiness performs well.

7.1 Broader Implications

We focus on strategic agents because they maximize their own utility and are thus generally more attractive to users. For example, a business would tune the decision models in one of its webservice products to maximize profit, or a user of a peer-to-peer file sharing service might attempt to change the peer-to-peer client software to achieve faster downloads. An agent with a high reputation may have considerably greater ability to cause harm to other agents than an agent with a low reputation, enabling a malicious agent to strategically build up reputation to maximize the harm it causes. A variation on our measures from Chapter 6 would be to use a strategically malicious agent, whose utility is a function of the loss of other agents. Many applications of reputation systems involving businesses and consumers, particularly those where an autonomous agent is acting on behalf of the firm or individual, will be faced against rational agents. However, a strong case may be made for modeling with strategically malicious agents for use in social networks, in businesses that might expect malice from extortionists or angry customers or competitors, or in using reputation as a basis for finding and tracking terrorists.

Seemingly terrible decisions, such as investing in a Ponzi scheme, can even be a rational strategy under certain economic situations where investors know they will likely be bailed out [Bhattacharya, 2003]. That the firm could strategically build its reputation and cause the economic loss that it did is an indication that something should be improved either in the interaction or reputation mechanism; a better reputation system might have prevented some of the financial damage from occurring in the first place.

Our desiderata from Chapter 6 are measured against a rational agent that would take advantage of any weaknesses of the reputation system, obtaining conclusive results for a reputation system intended for human involvement requires a sizable controlled experiment. Such empirical results would need to deal with the significant noise in the system and would require sufficient data to conclude that a fixed point of a reputation is a stationary process. Evaluating currently deployed systems with respect to our desiderata, although a significant undertaking, would contribute to the understanding of our desiderata, reputation systems, and also the rationality of human behavior.

The exact method that people use for discounting is still uncertain [Rubinstein, 2003], as is the notion of what an optimal discount factor is [Weitzman, 2001]. Our work offers many new methods to measure discount factors in different environments and enact upon knowledge of agents' discount factors. These results are beneficial not only to artificial intelligence, but also to other fields where self-interested agents may behave strategically, such as psychology, economics, and decision sciences.

7.2 Direct Future Work

Collusion, side-payments, and Sybil attacks (using many pseudonyms to boost or reset reputation) are exceptions when agents may appear to not act individually rational. Our desiderata from Chapter 6 can be adapted to measure the reputation dynamics given a certain number of colluding raters attempting to boost the reputation of one scamming agent. To extend these desiderata, the colluding raters should be treated as one agent in terms of utility. The reputation of the scamming rater, that is, the colluding agent with the reputation inflated by the other colluding raters, can then be used directly in the desiderata. In this way, our desiderata could be extended to operate on an organizational level.

The biggest weakness of our desiderata list is the computational complexity required to model reputation aggregation across a large number of agents and against strategic agents with high discount factors. Because we simply exhaust all possible actions, the number of states that must be computed is exponential with respect to the number of communicating agents and their actions. Solving a specific reputation system behavior against a strategic agent may be feasible with a simple reputation system and lead to an efficient solution, but large and complex reputation systems, particularly those without closed form solutions and highly domain-specific features (i.e., those having complex relationships between the reputation system and the interaction model), exacerbate the matter. Graphing Ω can offer insight into the dynamics of a reputation system, but visualising Ω may be nontrivial for systems that employ reputations of high dimensions that do not collapse easily to a scalar value. Determining how to adapt our desiderata or methodology to work well in complex scenarios is an interesting and useful avenue for future work.

Although agents' discount factors sometimes approximate agents' behavior in finite-horizon games, rational strategy can sometimes diverge sharply between the two settings, such as in the classic iterated prisoner's dilemma. We have focused on discounted infinite horizon games because, few real-world interactions in our highly connected world, particularly in e-commerce, are accurately modeled as isolated finite-horizon from the agents' perspectives. Though perhaps less practical, applying or adapting our desiderata for reputation systems with finite-horizon games is still an interesting problem to further our understanding of reputation systems.

Investigating more complex favor interaction models is an open area for further work. Agents could choose which agent to ask a favor, and use and learn joint probability distributions between additional parameters. This follows from many e-commerce situations, where agents can choose to procure goods or services from a number of vendors. Consider an agent that purchases a few boxes of printer paper on a monthly basis. The agent is faced with the problem of adverse selection in choosing from which vendor to buy. The vendors will have different prices, and each may have high or low quality paper and may deliver on-time or late. This situation can be abstracted into the widely studied multiarmed bandit problem [Kelly, 1981, Vermorel and Mohri, 2005], where an agent must make repeated discrete choices between transactions with different uncertainties. Conversely, in a situation where agents are acting more as peers, transactions are more equally important for each agent and moral hazard is significant. Examples of this case are agents offering significant bandwidth in peer-to-peer file sharing and when an agent is procuring a costly good or service. Whereas the multiarmed bandit problem has been studied with a single adversarial agent [Auer et al., 1995], our discount factor framework can apply in more complex settings.

When one agent chooses another agent when entering a reciprocal relationship, the second agent can learn information about the first agent's valuations, capabilities, and intertemporal discounting simply from the knowledge that the first agent decided to enter the relationship. Consider an online store that sells luxury watches. An agent that chooses to purchase from a luxury watch selling agent likely has different valuations, capabilities, or discounting than an agent that would choose to buy from a regular agent. The work of Chapter 5.1 can be expanded to determine what agents can learn from other agent's choices.

What would make a rational agent lie when rating another agent? Detection of subtle lies is difficult and thus the marginal cost of a small lie may seemingly be worth even a small gain. Further investigating the relationship between discount factors and incentives to lie, along with effective detecting and sanctioning methods is a further area for future work.

Much work remains to strategize about and measure discount factors and to apply this framework to various problem domains. However, strategic interactions with private discount factors appears to be a way of bridging the gap between trust and game theory.

REFERENCES

Tony Addison and S. Mansoob Murshed. Credibility and reputation in peacemaking. *Journal of Peace Research*, 39(4):487–501, 2002.

Alon Altman and Moshe Tennenholtz. Axiomatic foundations for ranking systems. *Journal of Artificial Intelligence Research*, 31:473–495, 2008.

Donovan Artz and Yolanda Gil. A survey of trust in computer science and the semantic web. *Web Semantics: Science, Services and Agents on the World Wide Web*, 5:58–71, 2007.

P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire. Gambling in a rigged casino: The adversarial multi-armed bandit problem. In *Proceedings of the 36th Annual Symposium on Foundations of Computer Science*, page 322, Washington, DC, USA, 1995. IEEE Computer Society. ISBN 0-8186-7183-1.

Robert Axelrod. On six advances in cooperation theory. *Analyse & Kritik*, 22:130–151, July 2000.

Rina Azoulay-Schwartz and Sarit Kraus. Stable repeated strategies for information exchange between two autonomous agents. *Artificial Intelligence*, 154(1-2):43–93, 2004. ISSN 0004-3702. doi: http://dx.doi.org/10.1016/j.artint.2003.04.002.

Dipyaman Banerjee, Sabyasachi Saha, Prithviraj Dasgupta, and Sandip Sen. Reciprocal resource sharing in P2P environments. In *Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 853–859, Utrecht, Netherlands, July 2005.

K. Suzanne Barber, Karen Fullam, and Joonoo Kim. *Challenges for Trust, Fraud and Deception Research in Multi-agent Systems*, volume 2631, chapter 2, pages 167–174. Springer Berlin / Heidelberg, 2003.

Michael R. Baye and Dennis W. Jansen. Repeated games with stochastic discounting. *Economica*, 63(252):531–541, 1996.

Uri Benzion, Amnon Rapoport, and Joseph Yagil. Discount rates inferred from decisions: An experimental study. *Management Science*, 35(3):270–284, March 1989.

Utpal Bhattacharya. The optimal design of ponzi schemes in finite economies. *Journal of Financial Intermediation*, 12(1):2–24, 2003.

Andreas Birk. Learning to trust. *Trust in Cyber-societies : Integrating the Human and Artificial Perspectives*, LNAI 2246:133–144, 2001.

Sviatoslav Braynov and Tuomas Sandholm. Contracting with uncertain level of trust. *Computational Intelligence*, 18(4):501–514, November 2002.

Chiranjeeb Buragohain, Divyakant Agrawal, and Subhash Suri. A game theoretic framework for incentives in P2P systems. In *Proceedings of the Third International Conference on Peer-to-Peer Computing*, pages 48– 56, 2003.

Cristiano Castelfranchi. Silent agents: From observation to tacit communication. *Lecture Notes in Computer Science*, 4140:98–107, 2006.

Cristiano Castelfranchi. Modelling social action for AI agents. *Artificial Intelligence*, 103(1-2): 157–182, 1998.

Vincent Conitzer and Tuomas Sandholm. Computational criticisms of the revelation principle. In *Proceedings of the 5th ACM conference on Electronic Commerce*, pages 262–263, 2004.

Vincent Conitzer and Tuomas Sandholm. Complexity of (iterated) dominance. In *Proceedings of the 6th ACM conference on Electronic commerce*, pages 88–97, Vancouver, Canada, 2005.

Peter C. Cramton. Strategic delay in bargaining with two-sided uncertainty. *The Review of Economic Studies*, 59(1):205–225, January 1992.

Vincent P. Crawford and Joel Sobel. Strategic information transmission. *Econometrica*, 50(6): 1431–1451, 1982.

Herbert A. David and H. N. Nagaraja. *Order Statistics*. Wiley, 3rd edition, August 2003.

Chrysanthos Dellarocas. Reputation mechanism design in online trading environments with pure moral hazard. *Information Systems Research*, 16(2):209–230, June 2005.

Chrysanthos Dellarocas. Reputation mechanisms. In Terrence Hendershott, editor, *Economics and Information Systems*, volume 1 of *Handbooks in Information Systems*, chapter 13, pages 629–660. Elsevier Science, 2006.

Morton Deutsch. *The Resolution of Conflict*. Yale University Press., 1973.

Robert L. Devaney. *A First Course in Chaotic Dynamical Systems: Theory and Experiment*. Westview Press, Boulder, Colorado, October 1992.

Roger Dingledine, Michael J. Freedman, and David Molnar. Accountability measures for peer-to-peer systems. In Andy Oram, editor, *Peer-to-Peer: Harnessing the Power of Disruptive Technologies*, chapter 16, pages 271–340. O'Reilly, February 2000.

Jeffrey C. Ely and Juuso Välimäkiz. Bad reputation. *The Quarterly Journal of Economics*, 118 (3):785–814, March 2003.

Joseph Farrell and Matthew Rabin. Cheap talk. *The Journal of Economic Perspectives*, 10(3): 103–118, 1996.

Drew Fudenberg and David K. Levine. Maintaining a reputation when strategies are imperfectly observed. *The Review of Economic Studies*, 59(3):561–579, 1992.

Karen K. Fullam and K. Suzanne Barber. Learning trust strategies in reputation exchange networks. In *Proceedings of the 5th International Conference on Autonomous Agents and Multiagent Systems*, pages 1241–1248, Hakodate, Japan, May 2006.

Karen K. Fullam, Tomas B. Klos, Guillaume Muller, Jordi Sabater, Andreas Schlosser, Zvi Topol, K. Suzanne Barber, Jeffrey S. Rosenschein, Laurent Vercouter, and Marco Voss. A specification of the agent reputation and trust (ART) testbed: experimentation and competition for trust in agent societies. In *Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 512–518, New York, NY, USA, 2005. ACM. ISBN 1-59593-093-0. doi: http://doi.acm.org/10.1145/1082473.1082551.

Brian P. Gerkey and Maja J. Matarić. Sold!: Auction methods for multirobot coordination. *IEEE Transactions On Robotics And Automation*, 18:758–768, 2002.

Casper Goffman and George Pedrick. *First Course in Functional Analysis*, chapter 2, page 71. AMS Bookstore, 1983.

Philippe Golle, Kevin Leyton-Brown, and Ilya Mironov. Incentives for sharing in peer-to-peer networks. In *Proceedings of the ACM Conference on Electronic Commerce*, pages 264–267, October 2001.

Ben Groom, Cameron Hepburn, Phoebe Koundouri, and David Pearce. Declining discount rates: The long and the short of it. *Environmental & Resource Economics*, 32:445–493, 2005.

Peter D. Grünwald and Joseph Y. Halpern. Updating probabilities. *Journal of Artificial Intelligence Research*, 19:243–278, 2003.

Werner Güth, Sabine Kröger, and Hans-Theo Normann. Durable-goods monopoly with privately known impatience: A theoretical and experimental study. *Economic Inquiry*, 42(3): 413–424, 2004.

Christopher J. Hazard. ¿Por favor? Favor reciprocation when agents have private discounting. In *AAAI Workshop on Coordination, Organizations, Institutions and Norms (COIN)*, pages 9–16, Chicago, Illinois, July 2008.

Ramón Hermoso, Holger Billhardt, and Sascha Ossowski. Integrating trust in virtual organisations. In Pablo Noriega, Javier Vázquez-Salceda, Guido Boella, Olivier Boissier, Virginia Dignum, Nicoletta Fornara, and Eric Matson, editors, *Coordination, Organizations, Institutions, and Norms in Agent Systems II*, Lecture Notes in Computer Science, pages 19–31. Springer, Berlin / Heidelberg, August 2007.

Kevin Hoffman, David Zage, and Cristina Nita-Rotaru. A survey of attack and defense techniques for reputation systems. *ACM Computing Surveys*, 41(1):1–31, 2009.

Daniel Houser and John Wooders. Reputation in auctions: Theory and evidence from eBay. *Journal of Economics & Management Strategy*, 15(2):353–369, 2006.

Trung Dong Huynh, Nicholas R. Jennings, and Nigel R. Shadbolt. An integrated trust and reputation model for open multi-agent systems. *Journal of Autonomous Agents and MultiAgent Systems*, 13(2):119–154, 2006.

Harvey S. James, Jr. The trust paradox: a survey of economic inquiries into the nature of trust and trustworthiness. *Journal of Economic Behavior & Organization*, 47:291–307, 2002.

Audun Jøsang. A subjective metric of authentication. In *Proceedings of the 5th European Symposium on Research in Computer Security*, pages 329–344, London, UK, 1998. Springer-Verlag.

Audun Jøsang and Walter Quattrociocchi. Advanced features in Bayesian reputation systems. In *Trust, Privacy and Security in Digital Business*, volume 5695 of *Lecture Notes in Computer Science*, pages 105–114. Springer Berlin / Heidelberg, 2009.

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, March 2007.

Radu Jurca and Boi Faltings. An incentive compatible reputation mechanism. In *Proceedings of the IEEE International Conference on E-Commerce*, pages 285–292, Newport Beach, CA, 2003.

Radu Jurca and Boi Faltings. Obtaining reliable feedback for sanctioning reputation mechanisms. *Journal of Artificial Intelligence Research*, 29:391–419, August 2007.

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, Budapest, Hungary, 2003.

F. P. Kelly. Multi-armed bandits with discount factor near one: The Bernoulli case. *Annals of Statistics*, 9(5):987–1001, 1981.

Reid Kerr and Robin Cohen. Modeling trust using transactional, numerical units. In *Proceedings of the 2006 International Conference on Privacy, Security and Trust*, pages 1–11, New York, NY, USA, 2006. ACM. ISBN 1-59593-604-1. doi: http://doi.acm.org/10.1145/1501434.1501460.

Reid Kerr and Robin Cohen. Smart cheaters do prosper: Defeating trust and reputation systems. In *Proceedings of the Eighth International Conference on Autonomous Agents and Multiagent Systems*, pages 993–1000, 2009.

Tapan Khopkar, Xin Li, and Paul Resnick. Self-selection, slipping, salvaging, slacking, and stoning: the impacts of negative feedback at eBay. In *Proceedings of the 6th ACM Conference on Electronic Commerce*, pages 223–231, Vancouver, BC, Canada, 2005.

Yann Krupa, Jomi F. Hubner, and Laurent Vercouter. Extending the comparison efficiency of the ART testbed. In *Proceedings of the First International Conference on Reputation - Theory and Technology*, pages 186–199, Gargonza, Italy, 2009.

Stephan Lewandowsky, Michael Mundy, and Gerard P. A. Tan. The dynamics of trust: comparing humans to automation. *Journal of Experimental Psychology: Applied*, 6(2):104–123, 2000.

Cuihong Li, Bin Yu, and Katia Sycara. An incentive mechanism for message relaying in unstructured peer-to-peer systems. In *Proceedings of the Sixth International Joint Conference on Autonomous Agents and Multi-Agent Systems*, pages 259–266, 2007.

George J. Mailath and Larry Samuelson. *Repeated Games and Reputations: Long-Run Relationships*, chapter 18, pages 568–579. Oxford University Press, New York, 2006.

Mikhail I. Melnik and James Alm. Does a seller's eCommerce reputation matter? Evidence from eBay auctions. *The Journal of Industrial Economics*, 50(3):337–349, 2003.

Lik Mui, Ari Halberstadt, and Mojdeh Mohtashemi. Notions of reputation in multi-agents systems: a review. In *Proceedings of the First International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 280–287, New York, NY, USA, 2002. ACM. ISBN 1-58113-480-0. doi: http://doi.acm.org/10.1145/544741.544807.

Sharon Paradesi, Prashant Doshi, and Sonu Swaika. Integrating behavioral trust in web service compositions. In *Proceedings of the IEEE International Conference on Web Services*, pages 453–460, Los Alamitos, CA, USA, 2009. IEEE Computer Society. ISBN 978-0-7695-3709-2. doi: http://doi.ieeecomputersociety.org/10.1109/ICWS.2009.106.

Ariel D. Procaccia, Yoram Bachrach, and Jeffrey S. Rosenschein. Gossip-based aggregation of trust in decentralized reputation systems. In *The Twentieth International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1470–1475, Hyderabad, India, January 2007.

Sarvapali D. Ramchurn, Dong Huynh, and Nicholas R. Jennings. Trust in multi-agent systems. *The Knowledge Engineering Review*, 19:1–25, 2004.

Sarvapali D. Ramchurn, Carles Sierra, Lluis Godo, and Nicholas R. Jennings. Negotiating using rewards. In *Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 400–407, New York, NY, USA, 2006. ACM. ISBN 1-59593-303-4. doi: http://doi.acm.org/10.1145/1160633.1160703.

Sarvapali D. Ramchurn, Claudio Mezzetti, Andrea Giovannucci, Juan A. Rodriguez-Aguilar, Rajdeep K. Dash, and Nicholas R. Jennings. Trust-based mechanisms for robust and efficient task allocation in the presence of execution uncertainty. *Journal of Artificial Intelligence Research*, 35:119–159, 2009.

Paul Resnick and Rahul Sami. The influence limiter: Provably manipulation resistant recommender systems. In *Proceedings of the ACM Conference on Recommender Systems*, pages 25–32, Minneapolis, MN, 2007.

Paul Resnick and Rahul Sami. The information cost of manipulationresistance in recommender systems. In *Proceedings of the ACM Conference on Recommender Systems*, pages 147–154, Lausanne, Switzerland, 2008.

Paul Resnick, Ko Kuwabara, Richard Zeckhauser, and Eric Friedman. Reputation systems. *Communications of the ACM*, 43(12):45–48, 2000. ISSN 0001-0782. doi: http://doi.acm.org/10.1145/355112.355122.

Paul Resnick, Richard Zeckhauser, John Swanson, and Kate Lockwood. The value of reputation on eBay: A controlled experiment. *Experimental Economics*, 9(2):79–101, 2006.

Shai Rubin, Mihai Christodorescu, Vinod Ganapathy, Jonathon T. Giffin, Louis Kruger, and Hao Wang. An auctioning reputation system based on anomaly detection. In *12th ACM Conference on Computer and Communications Security*, pages 270–279, New York, NY, USA, 2005.

Ariel Rubinstein. "economics and psychology"? the case of hyperbolic discounting. *International Economic Review*, 44(4):1207–1216, November 2003.

Ariel Rubinstein. Perfect equilibrium in a bargaining model. *Econometrica*, 50(1):97–109, January 1982.

Ariel Rubinstein. A bargaining model with incomplete information about time preferences. *Econometrica*, 53(5):1151–1172, September 1985.

John Rust and Christopher Phelan. How social security and medicare affect retirement behavior in a world of incomplete markets. *Econometrica*, 65(4):781–831, 1997.

Jordi Sabater and Carles Sierra. Review on computational trust and reputation models. *Artificial Intelligence Review*, 24(1):33–60, September 2005.

Sabyasachi Saha, Sandip Sen, and Partha Sarathi Dutta. Helping based on future expectations. In *Proceedings of the Second Intenational Joint Conference on Autonomous Agents and Multiagent Systems*, pages 289–296, Melbourne, Australia, July 2003.

Amirali Salehi-Abari and Tony White. Towards con-resistant trust models for distributed agent systems. In *International Joint Conference on Artificial Intelligence*, pages 272–277, 2009.

Tuomas Sandholm and Robert H. Crites. Multiagent reinforcement learning in the iterated prisoner's dilemma. *Biosystems*, 37:147–166, 1995.

Tuomas Sandholm and Victor Lesser. Equilibrium analysis of the possibilities of unenforced exchange in multiagent systems. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 694–701, Montreal, Canada, 1995.

Sandip Sen. Believing others: Pros and cons. *Artificial Intelligence*, 142(2):179–203, December 2002.

Sandip Sen and Dipyaman Banerjee. Monopolizing markets by exploiting trust. In *Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 1249–1256, Hakodate, Hokkaido, Japan, 2006.

Sandip Sen, Indranil Goswami, and Stéphane Airiau. Expertise and trustbased formation of effective coalitions: an evaluation of the art testbed. In *Workshop on Trust in Agent Societies at The Fifth International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 71–78, 2006.

Carles Sierra and John Debenham. An information-based model for trust. In *Proceedings of the 4th International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 497–504, New York, NY, USA, 2005. ACM. ISBN 1-59593-093-0. doi: http://doi.acm.org/10.1145/1082473.1082549.

Munindar P. Singh. An ontology for commitments in multiagent systems: Toward a unification of normative concepts. *Artificial Intelligence and Law*, 7(1):97–113, March 1999.

Michael J. Smith and Marie desJardins. Learning to trust in the competence and commitment of agents. *Autonomous Agents and Multi-Agent Systems*, 18(1):36–82, February 2009.

Jason D. Sonnek and Jon B. Weissman. A quantitative comparison of reputation systems in the grid. In *Proceedings of the 6th IEEE/ACM International Workshop on Grid Computing*, pages 242–249, 2005.

Mudhakar Srivatsa, Li Xiong, and Ling Liu. Trustguard: Countering vulnerabilities in reputation management for decentralized overlay networks. In *Proceedings of the 14th International Conference on World Wide Web*, pages 422–431, 2005.

Chandrasekar Subramaniam and Michael J. Shaw. A study of the value and impact of B2B e-commerce: The case of web-based procurement. *International Journal of Electronic Commerce*, 6(4):19–40, 2002.

W. T. 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. ISSN 1387-2532. doi: http://dx.doi.org/10.1007/s10458-006-5952-x.

Joannès Vermorel and Mehryar Mohri. Multi-armed bandit algorithms and empirical evaluation. In *Proceedings of the 16th European Conference on Machine Learning*, volume 3720 of *Lecture Notes in Computer Science*, pages 437–448, Porto, Portugal, 2005. Springer.

Thuc Vu, Rob Powers, and Yoav Shoham. Learning against multiple opponents. In *Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multi-Agent Systems*, pages 752–759, 2006.

William E. Walsh and Michael P. Wellman. Decentralized supply chain formation: A market protocol and competitive equilibrium analysis. *Journal of Artificial Intelligence Research*, 19: 513–567, 2003.

Yonghong Wang and Munindar P. Singh. Trust representation and aggregation in a distributed agent system. In *Proceedings of the 21st National Conference on Artificial Intelligence*, pages 1425–1430, July 2006.

Yonghong Wang and Munindar P. Singh. Formal trust model for multiagent systems. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence*, pages 1551–1556, Hyderabad, India, 2007.

Martin L. Weitzman. Gamma discounting. *The American Economic Review*, 91(1):260–271, 2001.

Michael P. Wellman. Market-oriented programming: Some early lessons. In Scott Clearwater, editor, *Market-Based Control: A Paradigm for Distributed Resource Allocation*, chapter 4, pages 74–95. World Scientific, River Edge, NJ, 1996.

Joseph M. Whitmeyer. Effects of positive reputation systems. *Social Science Research*, 29: 188–207, 2000.

Steven Willmott, Monique Calisti, and Emma Rollon. Challenges in large-scale open agent mediated economies. In *Agent-Mediated Electronic Commerce IV. Designing Mechanisms and Systems*, volume 2531 of *Lecture Notes in Computer Science*, pages 131–150. Springer Berlin / Heidelberg, 2002.

Jianzhong Wu and Robert Axelrod. How to cope with noise in the iterated prisoner's dilemma. *The Journal of Conflict Resolution*, 39(1):183–189, March 1995.

Bin Yu and Munindar P. Singh. Distributed reputation management for electronic commerce. *Computational Intelligence*, 18(4):535–549, November 2002.

Bin Yu, Munindar P. Singh, and Katia Sycara. Developing trust in large-scale peer-to-peer systems. In *2004 IEEE First Symposium on Multi-Agent Security and Survivability*, pages 1–10, 2004.

Giorgos Zacharia and Pattie Maes. Trust management through reputation mechanisms. *Applied Artificial Intelligence*, 14(9):881–907, October 2000.

Jie Zhang and Robin Cohen. A comprehensive approach for sharing semantic web trust ratings. *Computational Intelligence*, 23(3):302–319, 2007.

APPENDICES

APPENDIX A: Order Statistics

In this appendix, we present derivations of order statistics as may be found in a standard textbook on order statistics [David and Nagaraja, 2003]. The PDF of the ith lowest value in a probability distribution of N agents represented by PDF f and CDF F, f_{X_i}, is expressed as

$$f_{X_i}(x) = \frac{\Gamma(N+1)}{\Gamma(i) \cdot \Gamma(N-i+1)} \left(F(x)\right)^{i-1} \cdot \left(1 - F(x)\right)^{N-i} \cdot f(x). \tag{1}$$

The first term of this function may also be written as $\frac{\Gamma(N+1)}{\Gamma(i) \cdot \Gamma(N-i+1)} = \frac{N!}{(i-1)!(N-i)!}$. The expected value of the i^{th} order statistic can be written as

$$E(f_{X_i}(x)) = \int_{-\infty}^{\infty} \frac{\Gamma(N+1)}{\Gamma(i) \cdot \Gamma(N-i+1)} \left(F(x)\right)^{i-1} \cdot \left(1 - F(x)\right)^{N-i} \cdot f(x) \cdot x dx. \tag{2}$$

For the uniform distribution, the i^{th} order statistic may be expressed simply as

$$E(f_{X_i}(x)) = \frac{i}{N+1}. \tag{3}$$

The order statistic PDF for the uniform distribution is the Beta distribution, commonly represented as $V(i, N-i+1)$ when on the range of $[0, 1]$. The general PDF of the uniform distribution can be expressed on the range of $[a, b]$ as

$$f_{X_i}(x) = \frac{\Gamma(N+1)}{\Gamma(i) \cdot \Gamma(N-i+1)} \frac{\left(\frac{x-a}{b-a}\right)^{i} \cdot \left(\frac{x-b}{a-b}\right)^{N-i}}{x-a}. \tag{4}$$

The order statistic PDF of the exponential distribution about mean k can be expressed as

$$f_{X_i}(x) = \frac{\Gamma(N+1)}{\Gamma(i) \cdot \Gamma(N-i+1)} k e^{-kNx} \cdot (e^{kx} - 1)^{i-1}. \tag{5}$$