Representational Faithfulness in Accounting: A Model of Hard Information

by Michael Kirschenheiter[1]

Purdue University

Most Current Revision: May 2006

Current Date: May 16, 2006

[1] Address correspondence to Michael Kirschenheiter (email: mkirsche@purdue.edu), Purdue University, Krannert School of Management, 403 W. State St., Room 548 Krannert Building, West Lafayette, IN 47907-2056. I would like to acknowledge helpful suggestions from my colleagues, especially Dennis Caplan, Bjorn Jorgensen and two anonymous referees, on this and previous drafts of this paper.

Representational Faithfulness in Accounting: A Model of Hard Information

ABSTRACT: This study models representational faithfulness as “hard” information, or information that has a meaning upon which everyone agrees. In contrast to prior research, I show how "honest disagreements" may arise if we replace the assumption that individuals’ information partitions are common knowledge with the weaker assumption that the language they use is common knowledge. I start from the usual approach where a person's knowledge is modeled as a partition of the set of states of the world. I show that a language is "soft" if and only if it is not isomorphic to a partition of the set of states of the world. This indicates that the standard approach to modeling knowledge may represent an incomplete characterization, since, in a world with soft information, a complete description of knowledge requires the specification of the message set as well as the information partitions of the relevant individuals. I also show how the hard/soft criterion is different from comparing information partitions on their fineness and from the concept of common knowledge. Next, using the probability distribution defined over the set of states of the world, I construct a measure of relative hardness. I show that harder information systems are more informative in terms of Blackwell’s measure of relative informativeness. Also I show that relative hardness can be measured using the entropy of the underlying conditional probability distribution, providing a link between relative informativeness and the entropy measure.

Key words: Reliability, hard information, common knowledge, and entropy.

1. Representational Faithfulness in Accounting: Introduction

Consider the situation of an auditor who acts as an expert witness. The outcome of a court case often pivots on the testimony of an expert, so both sides to a dispute often bring their own experts. These experts auditors are sworn to tell the truth and are provided the same data; yet they often disagree. On the one hand, theoretically speaking, such "honest disagreements" seem surprising. While it may be possible that these experts disagree because they communicate strategically, economic theory has shown that the notion of an "honest disagreement" between experts is impossible.[2] On the other hand, practically speaking, such disagreements are expected and occasion no surprise when they occur. Further, the accounting profession has long recognized the possibility of honest disagreements. The conceptual framework in accounting recognizes that disagreement may exist about what financial information “purports to mean.” According to Statement of Financial Accounting Concepts (SFAC) #2 on the Qualitative Characteristics of Accounting Information, the requirement that accounting information should represent what it purports to represent is referred to as “representational faithfulness.” The objective of this paper is to develop a formal model of representational faithfulness.

The notion that perfectly observed information can have different meaning for different people has been defined in accounting theory as "soft" information. I use the term “hard” information to denote information that has the same meaning for all people, so that it possesses representational faithfulness, and the term “soft” information to denote information that is not hard. I show that the approach usually adopted in economic theory for expressing how economic players know things presupposes that all economic information is perfectly hard. In this manner, I show that honest disagreements do not arise when information is hard, but that they may arise

[2] Strategic communication refers to possibly dishonest communication between individuals with conflicting objectives. The objective of this paper is to investigate the possibility of honest disagreements, so non-strategic communication is assumed. This objective is not trivial. For example, Aumann (1976) shows that if an event is common knowledge to experts with identical priors, then “agreeing to disagree” is impossible. The objective here is to see when, if ever, agreeing to disagree may indeed be possible.

when information is soft. I argue that the reason prior research finds that honest disagreements are impossible is that this research assumes that what each individual can know is common knowledge.[3] If we assume instead that the language individuals use, and not each individual's potential knowledge of the world, is common knowledge, then honest disagreements are possible. Further, I show that allowing information to lack representational faithfulness requires that the language used by decision-makers provide decision-useful information separate from the underlying facts that are being communicated. It is almost as if how we say something is as important as what we say. Also, I provide examples that demonstrate how both the notion of common knowledge and the fineness criterion used in past literature to compare information systems are distinct from the criterion developed to assess the hardness of an information system. Last, I extend the model to develop a measure of relative hardness. I show how relative hardness is related to Blackwell's notion of decision-useful information and to the concept of entropy well established in information theory.[4]

2. Background and Prior Literature

The objective of this paper is to formally model and analyze the characteristic of representational faithfulness. Building this model led me to connect this research to other fundamental game theory research that effectively broadened the initial objective considerably. To appropriately convey where this paper fits in the literature, it is necessary to clarify the narrowness of the original research objective, in particular, to show what the original work did not do, and then to show how the initial work led to a broader research agenda.

---

[3] Sunder (2002) discusses the pivotal role that the common-knowledge assumption plays in accounting research.

[4] Anctil, Dickhaut, Kanodia and Shapiro (2003) develop and test experimental a model of information transparency and coordination and what I refer to as Blackwell's notion of relative informativeness they refer to as the "Blackwell fineness" criterion. Their measure of transparency is close to my measure of hardness, but they do not explicitly identify or distinguish the set of purported meanings of reports.

In SFAC #2, FASB argued that “better” accounting information is that information which readers of accounting information find more useful in making their decisions. In assessing the decision-usefulness of accounting information, FASB identified two primary aspects of accounting information: its relevance and its reliability. I ignore relevance in this paper and focus instead on analyzing reliability.

SFAC #2 decomposes reliability into two components, representational faithfulness and verifiability, and states that a third component, neutrality, interacts with these two. FASB refers to the characterization that a reliable measure will represent “what it purports to represent” (par.59) as “representational faithfulness.” SFAC #2 goes onto describe representational faithfulness as “the correspondence or agreement between a measure or description and the phenomenon it purports to represent” (par. 63). FASB describes verifiability as “the ability through consensus of measures to ensure that information represents what it purports to represent, … without bias or error” (glossary). While the conceptual framework describes intentional bias as a factor in the verifiability and neutrality of a measure, and pertains to representational quality, FASB distinguish representational quality from representational faithfulness. I also ignore representational quality and bias and focus only on analyzing the representational faithfulness of accounting information.

I purposely ignore reporting bias. People will likely communicate strategically, and prior literature has demonstrated that the strategic nature of the communication will significantly alter the game (see for example, Crawford and Sobel, 1982, Gigler, 1984 and Fischer and Stocken 2001). While my analysis can be extended to a strategic setting, intentional bias in reporting is not a factor in representational faithfulness. While analyzing reporting bias is extremely important, I believe that we need to model representational faithfulness separate from reporting bias if we are to understand how these aspects of information differ. I justify the narrowness of my definition of hard information because the focus of my research is not on representational quality (par 59) or on intentional bias in reporting, but on representational faithfulness. To address this research objective, I assume throughout this study that the decision-makers (abbreviated as DMs) communicate non-strategically.

Accounting research has directed attention at the notion of hard information since at least the early seventies when Ijiri (1975) emphasized this notion in arguing that accountability was a primary purpose of accounting measurement.[5] For Ijiri "the lack of room for disputes over a measure may be expressed as the hardness of a measure" (p. 36). He offers cash balances as an example of a hard measure and goodwill as an example of a soft measure. Gjesdal (1981) defines a soft information-reporting system as one in which the underlying information is unverifiable while the report is jointly observed and contractible. Penno (1990), Penno and Watts (1991) and Arya, Fellingham and Glover (1995) have also addressed hardness of accounting information, where the unverifiability of the information drives the definition of softness. These papers do not all use the term "hard information" in the same way. However, all of these papers are capturing some aspect of the notion, stated by Ijiri, that a hard measure "is one constructed in such a way that it is difficult for people to disagree" (p. 36).

My approach differs from those discussed above in that they all focus on the fact that soft information is susceptible to manipulation; my approach focuses on the fact that soft information may be interpreted differently by different people without relying on differences in motivation. As I discussed earlier, this is consistent with the research objective of modeling representational faithfulness, as opposed to modeling bias or representational quality. My approach also indicates that verifiability may be a more complicated notion than it at first appears. For example, my modeling of representational faithfulness shows that we may need to distinguish verifying facts by observing them from verifying what individuals know concerning these facts.

This issue with the meaning of “verifiability” relates directly to the work on common knowledge, especially Aumann (1976). An event is called common knowledge to two players if

[5] Ijiri (1975), page ix. Hong and Page (2001) present a related model of incomplete perspectives outside accounting, applying it to problem solving by teams of agents.

player one knows it, player two knows player one knows it, player one knows player two knows player one knows it, etc. Aumann (1976) showed that if two experts have a prior probability on some event occurring, and these probabilities are common knowledge, then they must be equal. My research objective is to consider how individuals interpret soft information. I show that experts may disagree about such information, which seems clearly to contradict Aumann's result. However, as Aumann points out, his result assumes that the information that each player can know (i.e., their information partitions) is common knowledge. I argue below that this assumption implicitly requires that information is hard, and therefore we need to relax it to allow for the possibility that information may lack representational faithfulness. This is where my initial narrow research topic broadened; this broadening requires elaboration.

Aumann (1976) defined common knowledge using an assumption commonly called the Common Prior Assumption, or CPA. [6] Loosely stated, the CPA says that differences in probability arise solely from differences in information; people who have the same information will assess any event as having the same probability of occurring. Aumann builds both his result in Aumann 1976 and the result (see Aumann 1987) that every Bayesian rational equilibrium is equivalent to a correlated equilibrium, upon the CPA. Challenging the CPA is not done lightly, nor was it the original objective of this research. However it appears unavoidable, so let's discuss the CPA in more detail.

As Aumann points (page 7, Aumann 1987) the CPA does not imply all players have the same subjective probability, but that all subjective probabilities differ only due to differences in information. He goes onto argue (ibid, pages 13-15) that the CPA is used because, unlike preferences, subjective probabilities are not individual, and also because the CPA enable researchers to "zero in on purely informational issues" in our models. Earlier (ibid page 9),

[6] Harsanyi (1968) showed that a game of incomplete information is equivalent to a game of imperfect information with nature making the initial move if and only if the CPA holds; hence this led Aumann to call the CPA the "Harsanyi doctrine" (see page 7 in Aumann 1987).

Aumann argues that his assumption that the information partitions of the individual players is common knowledge is in fact a theorem or tautology, and not an assumption. Further, he argues that this is part of the model and that the "situation with priors is similar". Aumann (1998) expands upon these arguments in defense of the CPA in response to Gul (1998), which argued that the CPA was an assumption, not a result, and that it could be relaxed. Morris (1995) provides perhaps the most thorough and eloquent discussion of the arguments for and against relaxing the CPA. He identifies 3 broad arguments for the CPA but concludes that the arguments either fail outright or hold limited sway, so that relaxing the CPA should produce useful research. Recent research (e.g., Morris, 1994 and Morris and Shin, 2002 and 2005, and Halpern 2000) relaxes the CPA to good effect.[7] I do not argue why relaxing the CPA is valid; instead, I argue below that relaxing the CPA is necessary for us to consider situations that lack representational faithfulness. In effect, assuming the CPA holds implies we assume representational faithfulness holds.

The research in this paper is also related to a line of behavioral accounting research that addresses the form managers use to present their information. This line of research has shown that the form of the information being communicated may affect how users of the information interpret it. The different types of communication that have been studied are quite varied, and include the form taken by earnings forecast (Hirst, Koonce and Miller, 1999), environmental disclosures (Kennedy, Mitchell and Sefcik, 1998) and comprehensive income disclosures (Hirst and Hopkins, 1998), among others. My finding that soft information involves a language that itself affects how DMs interpret the information is consistent with this research. Hence, the modeling of hard information in this paper complements this research, and in so doing, offers a

---

[1] Halpern (2000) investigates two characterizations of the CPA, including one based on a logical language, and shows they differ over infinite spaces. Morris (1994) investigates how "no trade" theorems under heterogeneous beliefs, Morris and Shin (2002) look at the social value of public information under heterogeneous beliefs, and Morris and Shin (2005) construct a set of "interaction games" that generalize different types of heterogeneity.

way to connect this behavioral research to an analytical framework.

Other related research also stretches the range of possible decision-making criteria. In a closely related study, Stecher (2005) constructs an economy with subjective informational assumption and shows that the welfare theorems will not hold in this economy. Gilboa and Schmeidler (1995) offer a decision theory that is "Case Based," where the decision-maker uses a similarity function to organize his decisions prior to taking his actions. Hong and Page (2001) show that behavioral diversity can arise due to the type of game ensemble a player has historically played. The current paper brings the language that DMs use into the framework for describing how DMs know and how they communicate their knowledge, so complements these other studies as well.

3. Research Methodology and the Basic Model

3.1 An Example of the Hard/Soft Information Distinction

To convey what is meant by hard or soft information, consider an example of valuing a piece of equipment for financial statement purposes using historical cost or fair value accounting. For the purposes of this example, assume that fair value accounting requires the manager to identify a comparable asset and then report the fair market value of this comparable asset. Suppose a manager purchases the equipment for a cost $10 and estimates that the equipment has a life of two years and no salvage value. At the end of year one, the manager would report a historical cost for this asset of $5 under straight line depreciation, and any auditor would agree with this valuation. Next, suppose the manager reports $5 under fair value accounting as well. Because this is based on identifying a "comparable" asset, there is ambiguity in the report, even though the actual amount is the same as it is under historical cost. An auditor auditing the fair value amount would need to acquire estimates of the "comparable" asset, and then assess whether the $5 was a reasonable valuation for the equivalent asset. In fact, had the manager reported $6 or $4, it is often the case that the auditor would still have agreed with the valuation precisely because there is ambiguity in the application of the term “comparable.”

This example highlights two points. First, a soft measure does not depend on the presence of incentives. The presence of a disagreement in the preceding example does not necessarily imply that reports are being purposely distorted as a result of differing incentives; the DMs may honestly disagree. Nor does verifying the process necessarily resolve the disagreement. Verification may clarify that a disagreement exists, but resolving the disagreement requires that the DMs agree on what is meant by a comparable asset. This leads to the second point.

This study focuses not on what a DM observes, but on how he communicates what he observes and how this communication is interpreted. The example illustrates that the hardness of information is intimately related to the language or message space involved. Changing the set of messages or changing the manner in which the DMs associate their knowledge with the messages can affect whether the information is hard or not.

At first glance, one might think that fair value accounting merely adds noise to the report, but I believe the difference between historical cost and fair value accounting in the example is subtler. The difference actually relates to how the inference process of the auditor, or any financial statement reader, differs from that of the manager, or financial statement preparer. To see this, I first introduce the formal notation and the basic model. Then I reconsider the example of asset valuation to clarify the subtlety that I think actually exists.

3.2 Basic Model:

Let (Ω, p) be a finite probability space and let S and G be partitions over the states of the world Ω presenting the information partitions of the manager and the auditor, respectively. For each state, $\omega \in \Omega$, let $s(\omega)$ and $g(\omega)$ denote the element of S and G, respectively, that contains ω. Suppose that these two players communicate using a language or message set denoted as M, where the manager can issue message $m_n \in M$. I assume that the message set has cardinality N, denoted as $|M| = N$, so that $M = \{m_n\} \bigcup_{1 \le n \le N}$. To issue the message, the manager uses his knowledge of the current state of the world. We can describe the process of issuing a report as a mapping from the manager's information partition into the set of possible messages, while a similar mapping exists for the auditor. Denote these mappings as $\sigma : S \rightarrow M$ and $\gamma : G \rightarrow M$, and call them the signaling functions of the manager and auditor, respectively, so that, if the current state of the world is $\omega \in \Omega$, then the manager reports $\sigma(s(\varpi)) = m_n$. Using the notation introduced above, define $\chi = (\Omega, p, S, G, M, \sigma, \gamma)$ as the "language-form" that represents the accounting method being employed. To model representational faithfulness, I need to specify what each message "purports to be." I do this by introducing the notion of an "anchor" set.

A1 (Assumption of an Anchor Set): For each language form $\chi = (\Omega, p, S, G, M, \sigma, \gamma)$, assume there exists a set, $X^M \subset \Omega$, with $X^M = \{x_n^M\} \bigcup_{1 \le n \le N}$ and where for each element, x_n^M , the following holds: $\sigma(s(x_n^M)) = m_n = \gamma(g(x_n^M))$

I call x_n^M the anchor of message m_n, or alternatively, I say that message m_n is anchored at x_n^M . The cardinality of the anchor set equals the cardinality of the message set, so that each message is anchored at one, and only one, anchor.

The anchor set specifies a set of states of the world about which the manager and auditor can agree when using message set M . The introduction of anchors accomplishes two tasks. First, it ensures that the two players can communicate at some basic level. For example, when one player says "black" the second player knows that he does not mean "white." Second, in order to formally model representational faithfulness, I need to specify a set of states of the world that describe the phenomenon that the messages "purport to represent." The anchor set fills this role.

As stated earlier, I assume that the DMs communicate non-strategically. Specifically, I assume both players can communicate, that is, can write and read the language, and do it honestly. The following assumptions on the signaling functions ensure that this is the case.

A2 (Signaling Functions Assumptions): For any language form, $\chi = (\Omega, p, S, G, M, \sigma, \gamma)$ assume the following holds for the signaling function of the manager, $\sigma : S \to M$:

i) $\sigma : S \to M$ is surjective or onto, or $\forall m_n \in M, \exists s \in S \ni \sigma(s) = m_n$.

ii) $\sigma : S \to M$ is honest, or $\forall \varpi \in \Omega$, $\varpi \in m_n \Rightarrow \sigma(s(\omega)) = m_n$.

Analogous conditions hold for the signaling function, $\gamma : G \to M$, of the auditor.

Assuming the information partitions of each DM form the domain of their signaling function ensures that they can write in the message set. Condition i), the assumption that the functions are surjective, ensures the DMs can read in this message set while condition ii) ensures that they will not purposely report dishonestly. I use assumptions A1 on the anchor set and A2 on the signaling functions to specify what is meant by a representationally faithful accounting method. I simplify the subsequent analysis by referring to a representationally faithful message as "hard," and a message that lacks representational faithfulness as a "soft" message.

3.3 Definition of the Hard/Soft Information Distinction:

To motivate the hard information definition, reconsider the example comparing two accounting methods for reporting asset value, historical cost and fair value accounting. We use language forms, $\chi_H = (\Omega, p, S, G, M_H, \sigma_H, \gamma_H)$ and $\chi_F = (\Omega, p, S, G, M_F, \sigma_F, \gamma_F)$, to represent these methods, where the subscripts denote accounting under the historical cost and fair value methods, respectively. This means that $m_{H,n} \in M_H \subseteq \Re^+$ and $m_{F,n} \in M_F \subseteq \Re^+$ denote the report or message issued by the manager under historical cost and fair value accounting, respectively. Denote the message mappings of the manager and auditor under historical cost accounting as $\sigma_H : S \to M_H$ and $\gamma_H : G \to M_H$, respectively. Analogous mappings exist that describe how reports are issued under fair value accounting.

What prevents disagreements when the manager reports under historical cost, and what makes historical cost "hard," is that, in every state of the world, the auditor always reports the same message as the manager. More formally, suppose the manager "knows" the element of his partition that contains the state of the world, $s(\omega)$ and reports $\sigma_H(s(\omega)) = m_{H,n} \in M_H$. Seeing the historical cost report $m_{H,n}$, the auditor uses the inverse of her message mapping, $\gamma_H^{-1} : M_H \to G$, to infer the element of her information partition that contains the current state of the world. Under a hard reporting system such as historical cost is assumed to be, the auditor will infer the element $g(\omega)$, that is, we will have $g(\omega) \in \gamma_H^{-1}(m_{H,n})$. Under fair value accounting this may not be true, that is, the auditor may infer that a different element of her information partition describes the current state of the world. Formally, this means that $g(\omega) \notin \gamma_F^{-1}(m_{F,n})$ may hold. This intuitive notion of hard information leads to the following formal definition.

Definition of Hardness: A language form $\chi = (\Omega, p, S, G, M, \sigma, \gamma)$ is hard if for all messages $m_n \in M$ and for all states $\omega \in \Omega$ we have $m_n = \sigma(s(\omega)) \leftrightarrow m_n = \gamma(g(\omega))$.

As noted earlier, I rule out the possibility of strategic message choice via assumptions A1 and A2. This means that, by assumption, hardness is not a choice variable, but an exogenous characteristic of the language form. I will sometimes say a message set is hard, with the understanding that this means the language form is hard.

My definition of hard information coincides with the requirements for information to be hard laid out by Ijiri (1975). Ijiri describes hard measurement as the "processing of verifiable facts by justifiable rules in a rigid system which allows only a unique set of rules for a given situation" (page 36). The rules are the signaling functions, which are unique for each DM. The fact is the state of the world that is observed by the DMs. My definition ensures that, under a hard language form, these facts can be communicated so that each DM "knows" the fact in the same way. However, under a soft language form, as I define it, the facts will not be "verifiable" in the usually sense of this word.

Within the context of my model, a DM can verify a "fact" or event by observing it. Further, one DM can verify that a second DM "knows" this event by observing that the second DM observed the event. However, one DM may not be able to verify that the second DM "knows" the event in the same way as the first DM "knows" it. The DMs know the state of the world through the filter of their own information partition, but understanding what the other DM "knows" requires that this information be communicated and that the message be interpreted. With soft information, DMs interpret this information differently. Different interpretations are possible because we no longer assume the DMs' information partitions are common knowledge. I return to this point when I relate hard information to the notion of common knowledge in Section 4.1 below.

While I define hardness in terms of the signaling functions, we can also speak about the inverse mappings, that is the mappings that take the reports into the information partitions. Denote these inverse mappings as $\sigma^{-1}: M \to S$ and $\gamma^{-1}: M \to G$ for the manager and auditor respectively. I now introduce another definition that proves useful in subsequent analysis.

Definition of Inverse Message Sets: For language-form $\chi = (\Omega, p, S, G, M, \sigma, \gamma)$, define the inverse message sets for the manager and auditor as $S^M = \left\{ \left\{ s_n^M \right\}_{1 \le n \le N} \right\}$ and $G^M = \left\{ \left\{ g_n^M \right\}_{1 \le n \le N} \right\}$, respectively, where $s_n^M = \left\{ s \in \sigma^{-1}(m_n) \right\}$ and $g_n^M = \left\{ g \in \gamma^{-1}(m_n) \right\}$.

This definition introduces message-equivalent signal spaces, which are the set of signals which can be communicated between DMs using a given message space. In this model, one DM's knowledge is knowable by another DM if and only if it can be communicated. Hence, these information partitions are the relevant ones for understanding the information that one DM knows that can be communicated using the given language. The importance of these sets is discussed in more detail below, especially in connection with Corollary 1, where I show that message-equivalent information partitions are identical if and only if the message space is hard.

4. Results

The results are presented in two sections. In section 4.1, I start by presenting the basic representation result concerning hard message sets, and show how soft messages allow for disagreement. In section 4.2, I present definitions of relative hardness, and extend the analysis to show how the different measures of relative hardness can be used

4.1. Basic Results on Hardness

The first question that we wish to answer is whether the hardness of information is related to the message set used. In answering this question, direct message sets are important. A direct message set is defined as $M = \Omega$. I will write $M \leq \Omega$ to indicate that the set of states of the world is a refinement of the message set. The intuition is that hardness and directness are equivalent characteristics, and this intuition is formalized in the following theorem.

Theorem 1: A language-form $\chi = (\Omega, p, S, G, M, \sigma, \gamma)$ is hard if and only if there exists a second language form, $\chi = (\Omega, p, S, G, M', \sigma', \gamma')$, where $M' \leq \Omega$ and M' is isomorphic to M, so that $\sigma^{-1}(m_n) = \sigma'^{-1}(m'_n)$ and $\gamma^{-1}(m_n) = \gamma'^{-1}(m'_n)$ for each $n = 1,\ldots N$ (See Appendix for all proofs.)

Theorem 1 states that an accounting method generates hard information if and only if the language is a direct representation of the states of the world. It is obvious that this condition suffices to ensure hardness, since we assume truth-telling by the DMs. The necessity of this condition is perhaps more surprising. It seems to be in the nature of virtually every language to include some messages that cannot be defined solely in terms of states of the world. For example, consider the formal definition of assets as “future benefits” While the term future benefits can be associated with certain states of the world, it seems impossible to construct an exhaustive list of states such that a future benefit exists if and only if one of these states occurs.

Theorem 1 tells that unless such a list can be constructed, the term future benefits lacks representational faithfulness.

Alternatively, Theorem 1 says that a necessary and sufficient condition for information to be soft is that the language form involves communicating using ambiguous messages in the sense that they have no specific meaning or definition in terms of the states of the world, but are nonetheless decision-relevant.[8] In fact, it is their ambiguity that enables these messages to convey meaning outside of the "knowledge" represented by the set of states of the world. Once we assume that the message set is a coarsening of the direct message set, the possibility that an intelligent and honest player may use his judgement in reporting a signal about the state of the world can be ruled out; he merely repeats the state itself. However, if in our language there is a word that is not definable in terms of specific states of the world and if this word is relevant to some decision, then that decision involves soft information.

Theorem 1 also highlights a relationship among the information partitions and the message space, which is formalized in the following corollary.

Corollary 1: A message space is hard with respect to two information structures if and only if the message-equivalent information partitions are equivalent.

Corollary 1 redefines hardness, so that it merely restates Theorem 1, but in so doing, it highlights how hardness relates to information partitions S and G. First, Corollary 1 tells us that only the message-equivalent sets S^M and G^M matter for determining hardness, not S and G.

[8] Similar to "future benefits," the term "fair market value of a comparable asset" in the example of valuing an asset is an ambiguous yet decision-relevant message if we cannot completely define this term via a list of states of the world. For some assets, such a definition may be possible. For example, the fair value of a used car can be found from its "blue book" value. However, in this case the phrase "fair value of a comparable asset" would be replaced by the phrase "the blue book value of a used car of the same make, model, and year." Using an analogous argument, the "fair value of a common share of IBM stock" may be considered hard information, if it is understood to mean the quote on the NYSE, or it could be considered soft information if this value is based on a subjective assessment by management.

Informally speaking, we do not care what the DMs know, only what they can communicate. Second, it is clear that $M' = S^M = G^M$ is a (weak) coarsening of both S and G. Corollary 1 tells us that the opportunity to communicate knowledge as hard information is restricted to information that can be represented by an information partition that is a common coarsening of the information partitions of the DMs involved. Hence, the DMS may have information partitions that preclude the existence of any non-trivial accounting method that has representational faithfulness.

Two additional points follow almost immediately from Theorem 1 and Corollary 1. These points concern how the hard-soft distinction relates to the fineness criterion and to the notion of common knowledge of events. I discuss each of these in turn.

First based on a casual notion of hardness, one might think initially that hardness is equivalent to the notion of fineness: Corollary 1 shows otherwise. Simple examples demonstrate that having comparable information partitions, that is, being able to rank the original information partitions based on fineness, is neither a necessary nor a sufficient condition for the message set to be hard. First, if the message set is a singleton, so that the same message is always reported, then it is clearly hard. This holds even if partitions S and G are not comparable. Next, suppose $M = \{m_1, m_2\}$ and let $S = \{s_1, s_2, s_3\}$ be a refinement of $G = \{g_1, g_2\}$ where $s_1 = g_1$. and $s_2 \cup s_3 = g_2$ and suppose that $\omega \in s_3$ is the state of the world. Then it is clearly possible that $\sigma(\varpi) = m_1 \neq m_2 = \gamma(\omega)$, so that comparability of the information partitions does not imply hardness. The independence of the hard/soft distinction from the fineness is particularly interesting, as fineness of an information partition is often associated with the informativeness of a signal, in Blackwell's sense. I return to this point when I derive relative measures of hardness in the next section.

The second point following from Theorem 1 and Corollary 1 is that hardness is also independent of the notion of common knowledge. To see this, consider a situation, due to Aumann (1976), where disagreement is impossible. Let (Ω, p) be a finite probability space and let S and G be partitions over the states of the world Ω presenting the information of two players. Let R be the meet or finest common coarsening of these two partitions, written in standard notation as $R = S \wedge G$. For each state of the world, $\omega \in \Omega$, let $s(\omega)$ denote the element of S that contains ω, and let $g(\omega)$ and $r(\omega)$ denote similar relations for those partitions. Aumann proposed the following definition for common knowledge. Call an event, $A \subset \Omega$ common knowledge for the two players at $\omega \in \Omega$ if $r(\omega) \subset A$ holds. Aumann then used this definition to prove that two experts cannot "agree to disagree." Formally, his result is given as follows:

Proposition (see Theorem 1 in Milgrom 1981): Suppose that for some event $A \subset \Omega$ and state $\omega \in \Omega$, it is common knowledge at $\omega \in \Omega$ that $p(A|S) = \alpha$ and $p(A|G) = \beta$. Then $\alpha = \beta$.

This result says that if two experts have a prior probability on some event occurring, and these probabilities are common knowledge, then they must be equal. Theorem 1 seems clearly at odds with this result. We can trace the source of this conflict to what we assume to be common knowledge under each approach.

The approach in this paper conflicts with Aumann's approach because I separate what the players know from what they can communicate. Aumann assumes that the information partitions of the two experts are themselves common knowledge. He acknowledges this fact, saying that "worthy of note is the implicit assumption that the information partitions… are themselves common knowledge" (see page 1237). He claims that the assumption can be made without loss of generality since "included in the full description of the state $\omega \in \Omega$ of the world is the manner in which information is imparted to the two players" (page 1237). In making this assumption, Aumann implicitly assumes the players communicate using a hard language.

I relax the assumption that the information partitions are common knowledge and replace it with the weaker assumption that the language is common knowledge. Each individual can know an event, but the first DM cannot know what the second DM knows, only what the second DM says she knows. By shifting the focus from the common knowledge of an event to the common knowledge of the report of an event, I shift the focus from the players' information partition to their inverse image sets. As Corollary 1 shows, the two DMs always agree about the report of events if and only if the inverse image sets are identical, that is, if and only if the message set is hard. Further, an event may be common knowledge for two players, yet they may disagree in their reporting of this event. Hence, the notion of whether or not the event is common knowledge is distinct from whether or not the report of this event is hard or soft. While the concept of common knowledge remains important, it seems that the latter criterion is the relevant one when investigating whether or not experts will agree on reported information.

The attractiveness of using information structures to describe the distinction between hard and soft information is its generality. However, this approach does not quantify hardness. In the next section, I address the latter difficulty at the cost of some generality.

4.2. Measures of Relative Hardness

This section provides a measure relative hardness. In my definition of relative hardness, I use the probability distribution defined over the inverse image sets of the DMs. Defining hardness based solely on the underlying probability distribution allows a quantification of hardness. Prior to this point in the analysis, hardness has been defined only in terms of sets and functions and only in absolute terms. Showing an equivalent definition in terms of probabilities may convey more intuition about hardness, and it relates the hardness concept to a large body of literature. Further quantifying hardness allows us to speak of relatively harder or softer message sets, increasing the potential usefulness of the concept.

Use the conditional probabilities defined on the inverse messages sets of the two DMs, (i.e., $p\left(s_k^M \mid g_n^M\right)$, where $s_k^M = \sigma^{-1}(m_k)$ and $g_n^M = \gamma^{-1}(m_n)$) to define relative hardness as follows.

Definition of Relative Hardness: Consider two language-forms, $\chi = (\Omega, p, S, G, M, \sigma, \gamma)$ and $\chi' = (\Omega, p, S', G, M, \sigma', \gamma)$ with the same anchor set, so that $\forall n$, both s_n and s'_n are anchored at x_n^M. Define χ as harder than χ' if, for the elements of the inverse image sets, $g_n^M \in G^M$, s_n^M and $s_k^M \in S^M$ and s'^M_n and $s'^M_k \in S'^M$, the following hold:

a. $p(s_n^M \mid g_n^M) \geq p(s'^M_n \mid g_n^M) \geq p(s'^M_k \mid g_n^M)$, and

b. $p(s_k^M \mid g_n^M) \leq p(s'^M_k \mid g_n^M)$, with at least one inequality strict.

In general, the definition of relative hardness introduced above does not completely order a set of standards. We shall call two language forms compatible if they can be ranked using the above definition of relative hardness.

One can understand the intuition behind the definition most easily by recognizing that the definition of relative hardness is analogous to the notion of mean-preserving spreads. One distribution is a mean-preserving spread of a second distribution if the first distribution spreads the probability from realizations closer to the mean to realizations farther from the mean, while keeping the mean the same. In the above definition, the language forms differ only by the signaling function of the first DM. Instead of spreading the probability around the mean, softer information systems spread the probability of each conditional distribution of the first DM's signals around the second DM's signal.

Reconsider the example of valuing an asset used in Section 3 to motivate the definition of absolute hardness. Suppose for simplicity that the asset's value is either low or high, denoted as m_1 or m_2, respectively. When message m_n is reported, the manager and auditor infer the inverse image message denoted as $s_{Y,n}^M$ and $g_{Y,n}^M$, respectively, under accounting method $Y = \{H, F\}$, where as before H represents historical cost and F represents fair value accounting. If, as earlier assumed, historical cost accounting is perfectly hard, this means $p(s_{H,1}^M \mid g_{H,1}^M) = 1 = p(s_{H,2}^M \mid g_{H,2}^M)$, while $p(s_{H,2}^M \mid g_{H,1}^M) = 0 = p(s_{H,1}^M \mid g_{H,2}^M)$. Suppose that fair value accounting is soft, so that under fair value accounting, the inferences of the manager and auditor agree 80% of the time for a low valuation and 75% of the time for a high valuation. This means that when the auditor infers a low valuation, we have $p\left(s_{F,1}^{M} \mid g_{F,1}^{M}\right)=.8$ and $.2=p\left(s_{F,2}^{M} \mid g_{F,1}^{M}\right)$, while when she infers a high valuation, we have $.25=p\left(s_{F,1}^{M} \mid g_{F,2}^{M}\right)$ and $p\left(s_{F,2}^{M} \mid g_{F,2}^{M}\right)=.75$.

In general, a softer language form reduces the probability of a same message outcome, $p\left(s_{n}^{M} \mid g_{n}^{M}\right)$, and raises the probability of a different message outcome. Hence, the perfectly hard message set has $p\left(s_{n}^{M} \mid g_{n}^{M}\right)=1$ for $k=n$ and $p\left(s_{k}^{M} \mid g_{n}^{M}\right)=0$ for $k=1,\ldots,n-1,n+1,\ldots,N$. As discussed below, it seems natural to define a message set of maximal softness as one that has $p\left(s_{k}^{M} \mid g_{n}^{M}\right)=\frac{1}{N}$ for all $n=1,\ldots,N$ and $k=1,\ldots,N$.

I begin the analysis of the results of this section by showing that a harder language form is more informative in Blackwell's sense. This is done in the following theorem.

Theorem 2: Suppose χ is relatively harder than χ' as given in the definition above. Then utility maximizing DMs prefer the relatively harder language form χ to χ'.

Theorem 2 says that a harder message space produces signals that can be used by a DM to increase his expected utility. The intuition follows again by recalling the analogy of relative hardness to the notion of mean-preserving spreads. Just as mean-preserving spreads lower expected utility, so do softer message sets.

A valuable aspect of Theorem 2 is that it says a harder message set produces signals that are more informative, in the Blackwell sense, even though the information partitions of the two DM's may not be comparable.[9] Blackwell defined one signal as being more informative than a

[9] Blackwell (1951) provided necessary and sufficient conditions under which ranking message sets by their informativenes is equivalent to ranking them by fineness. More specifically, Blackwell considered the case where there were two signals, $s \in S$ and $g \in G$, and a decision-relevant variable $x \in X$, and a probability distribution p defined over these variables. He assumed that the signals were noiseless, so that $p(g \mid x)=1$ or $p(g \mid x)=0$ and $p(s \mid x)=1$ or $p(s \mid x)=0$ held for all realizations of the variables. Blackwell showed that S was relatively more informative than G if and only if S was a refinement of

second signal if the first signal enables the DM to obtain a higher expected utility. Theorem 2 may seem surprising, since Blackwell's famous result is often interpreted to mean that information systems can be ranked based on their informativeness only if the signals are comparable.[10] Theorem 2 demonstrates that this is not so if signals are noisy. Informally, a noisy signal is one that provides imperfect information about a decision-relevant variable (see note 7 for a formal definition of noiseless signals). For example, in the earlier discussion of valuing an asset, the decision-relevant information might be the cash flow that the asset will generate. The historical cost of the asset would be considered a noiseless signal if only one historical cost amount is possible for each actual cash flow realization.

In general, we expect most signals in accounting to be noisy signals, and not to provide perfect information. Theorem 2 states that compatible information systems, that is, those that can be ranked by relative hardness, can be ranked in terms of their informativeness even though they may include noisy signals. Hence it offers a way to partially order information partitions and message sets that is qualitatively different than the fineness criteria.

The second approach to measuring hardness uses the concept of entropy.[11] For any arbitrary probability distribution, $p(y_n)$, with finite support $y_n \in Y$ for $n = 1,\ldots,N$, denote the entropy of this distribution as $H(p(y_n))$, where the entropy is given by the following formula

$$H(p(y_n)) = -\sum_{n=1}^{N} \left(p(y_n) \cdot \mathbf{ln}(p(y_n)) \right) \geq 0 .$$

For this second measure of relative hardness, I again use the conditional probabilities defined on the inverse message sets of the two DMs. Using these probabilities, let

$$H\left(S^M \mid g_n^M\right) \equiv H\left(p\left(s_k^M \mid g_n^M\right)\right) = -\sum_{k=1}^{N} \left(p\left(s_k^M \mid g_n^M\right) \cdot \mathbf{ln}\left(p\left(s_k^M \mid g_n^M\right)\right)\right)$$

G. Marschak and Miyasawa (1968), McGuire (1972), Green and Stokey (1977) and Malueg (1985) clarify that assuming noiseless signals is an indispensable condition for refinement to be a necessary condition for relative informativeness. In my model, I relax the assumption that the signals are noiseless.

[10] Demski (1973) applies Blackwell's theorem to accounting using this perspective.

[11] As Khinchin (57) points out, the entropy concept first arose from the attempt to create a theoretical model for the transmission of information.

denote the entropy of the conditional probability distribution $p\left(s_k^M \mid g_n^M\right)$ and let

$$H\left(S^M \mid G^M\right) = \sum_{n=1}^{N} \left(p\left(g_n^M\right) \cdot \left(H\left(S^M \mid g_n^M\right)\right)\right)$$

denote the expected value of the entropy of this conditional-probability distribution. Using these measures, we have the following result that describes relatively hard message sets in terms of the entropy of the conditional distributions.

Theorem 3: Language-form $\chi = (\Omega, p, S, G, M, \sigma, \gamma)$ is hard if and only if $H\left(S^M \mid G^M\right) = 0$. Also, if χ is compatible with a second language form, $\chi' = (\Omega, p, S, G, M, \sigma', \gamma)$, then χ is harder than χ' if and only if $0 \leq H\left(S^M \mid G^M\right) < H\left(S'^M \mid G^M\right)$.

Theorem 3 provides a second and equivalent measure of relative hardness using the entropy of the conditional distributions over the inverse image sets. The previous analogy between mean preserving spreads and relative hardness again conveys the intuition of the result. The entropy of a distribution increases as we spread the probability among all the realizations, reaching a maximum when all realizations are equally likely. Under a softer language form, the probability of the conditional distribution is spread among the unanchored realizations. So, for example, given the auditor infers the inverse image message g_n^M, a softer language form would decrease the probability $p\left(s_n^M \mid g_n^M\right)$ and increase the probabilities $p\left(s_k^M \mid g_n^M\right)$ for $k \neq n$. The entropy of the conditional distribution captures and quantifies the impact of changes in hardness.

Theorem 3 offers additional intuition into the notion of hardness, as well as potential practical benefits. First, the entropy of a distribution is often interpreted to be a measure of uncertainty, so that Theorem 3 shows that harder language forms involve less uncertainty under this interpretation. Second, the entropy measure of relative hardness may be easier to calculate than the Blackwell measure. However, while the ordinal ranking clearly holds, it is not clear what, if anything, can be inferred from the cardinal values of the entropy measure. For example, it is not clear how we should interpret the quantitative difference in hardness between two language forms if $2H\left(S^M \mid G^M\right) = H\left(S'^M \mid G^M\right)$ holds.

Theorems 2 and 3 together provide a link between Blackwell's measure of relative informativeness and the entropy measure; to my knowledge this is the first time such a link has been established. The key to establishing this link is, of course, the construction of the anchor set. The anchors formally establish a set of states of the world that represent what each message "purports to represent." By doing so, we are able to provide a non-trivial role for the accounting notion of representational faithfulness. The importance of building a formal model of representational faithfulness is underscored by Aumann's earlier result demonstrating conditions under which experts could not disagree. As the preceding results demonstrate, experts can disagree if the language they use is soft. Further, the relative level of disagreement can be rank-ordered. Hence, this model offers the potential for a comparative analysis of accounting standards based on their relative representational faithfulness.

5. Summary and Suggestions for Future Research

The objective of this paper was to develop a model of representational faithfulness by formally distinguishing between hard and soft information. The intuition for this distinction and the formal model were presented in Section 3 and I present the result of my analysis in Section 4. Theorem 1 showed that a necessary and sufficient condition for information to be soft was that the information be communicated using a nonsensical symbol. By nonsensical I mean a symbol that cannot be defined in terms of the states of the world, but which was nonetheless relevant to the decision-maker. I also showed that the comparability of the information partitions was neither a necessary nor a sufficient condition for the message space to be hard. Further I showed that whether or not an event is common knowledge does not dictate whether that event could be represented by a hard message.

I also investigated the relationship between hard information as defined on information partitions and the underlying probability distribution, and defined a measure of relative hardness.

Theorem 2 showed that a harder language form would generate a more informative signal, using Blackwell’s notion of relative informativeness. Theorem 3 showed not only that the entropy of the conditional probability distribution offered an equivalent definition of hard information, but also that entropy could be used to rank order the language forms in terms of their relative hardness. Hence, entropy quantified hardness, and provides a quantification of relative informativeness in Blackwell’s sense. Although no specific applications of the hard/soft distinction in information have been formally developed, numerous research questions might benefit from the concept of soft information. I describe a few in more detail.

In SFAC #2, FASB offer the use of fair value or replacement cost for valuing assets as an explicit example of the difficulties faced when accountants attempt to achieve representational faithfulness. However, this is just one example where the representational faithfulness of an accounting construct may be questioned. The procedures for estimating most reserves, such as a reserve on uncollectible receivables, inventory obsolescence, warranties and sales returns, or a reserve for litigation may all be described as examples that require accountants to clarify the representational faithfulness of the accounting information reported. Also the efficacy of standards, both auditing and accounting, might be analyzed using this distinction. In particular, modeling hard information as representational faithfulness may be a first step in constructing an analytical framework that corresponds to the conceptual framework of financial reporting provided in the Statements of Financial Accounting Concepts.

Second, the effect of the hard/soft information distinction on the role of public accountants is particularly interesting. Often research portrays the role of public auditors as verifying unobservable signals and then focuses on possible collusion between the manager and the auditor against the investor, resulting in misrepresentations to the investor. While blatant collusion such as fraud clearly exists, most such collusion seems subtler, and more difficult to analyze within the current auditing models. For example, current approaches do not model well the possibility of honest disagreements between knowledgeable parties, on which most lawsuits rest. The hard/soft distinction in information offers an approach to formalize the demand for expert opinions (including lawyers and economists as well as auditors). It also offers a way to explicitly model judgement or subjective evaluation of information that must be a part of any complete model of auditing. Auditing expertise may be related to soft information. An auditor does more than verify the accuracy of numbers in the financial statements; she also judges whether or not they are accumulated in accordance with GAAP on a consistent basis. If all people interpret information in the same way, then no judgement is required of the auditor: only with soft information does the auditor's judgement and expertise play a role.[12]

Third, in the standard principal-agent model, the incentive problem arises solely because the agent's effort is unobservable by the principal. An equally valid description of the problem might be that the agent's effort is soft information. Thus even if the agent was honest, difficulty in contracting may arise. Further, these two different problems may arise simultaneously. The contract that solves the incentive problem when the action is unobservable may (or may not) be the same as the optimal contract when the outcome of the action is soft information.

The three areas discussed above are only a few of the areas to which the hard/soft information distinction might be applied. Other areas include work on incomplete contracting, valuation, bounded rationality, common knowledge, and the mechanism design literature's analysis of the information requirements for implementation of an equilibrium, to name a few. For example, the hard/soft distinction can be easily extended to the framework of a Bayesian communication game, which has been applied to the principal-agent model (Myerson, (1982)) as well as to the mechanism-design literature (Reiter and Reichelstein, (1988)). Among other results, it is straightforward to show that the revelation principle will fail if information is soft, but holds for hard information. Thus the approach appears very flexible.

[12] See Caplan and Kirschenheiter (2004) for an application of the hardness criteria to auditing expertise.

References

Arya, A., J. Fellingham and J. Glover, "Implicit Incentives and the Contents of the Accounting Library," working paper, (1995).

Aumann, R. "Agreeing to Disagree." *The Annuals of Statistics*, 4, (1976): 1236-1239.

----- "Correlated Equilibrium as an Expression of Bayesian Rationality." *Econometrica*, 55, No. 1 (January, 1987): 1-18.

----- "Common Priors: A reply to Gul." *Econometrica*, 66, No. 4 (July, 1998): 929-938.

Blackwell, D. "The Comparison of Experiments." Proceedings, Second Berkeley Symposium on Mathematical Statistics and Probability, (1951): 93-102. University of California Press, Berkeley.

Caplan, D. and M. Kirschenheiter. "A Model of Auditing Under Bright-Line Accounting Standards," *Journal of Accounting, Auditing and Finance*, 19 (Fall 2004) 4: 523-559.

Crawford, V. P., and J. Sobel. "Strategic Information Transmission." *Econometrica*, 50, No. 6, (November 1982): 1431-1451.

Demski, J. "The General Impossibility of Normative Accounting Standards." *The Accounting Review*, (October 1973): 718-723.

Financial Accounting Standards Board. 1983. *Statements of Financial Accounting Concepts 2*. Stamford, CT: FASB.

Fischer, P and P. Stocken. "Imperfect Information and Credible Communication." *Journal of Accounting Research*, 39, (June 2001): 119-134.

Gigler, F. "Self-Enforcing Voluntary Disclosure." *Journal of Accounting Research*, 32, (Autumn 1994): 224-240.

Gilboa, I. and D. Schmeidler. "Case-Based Decision Theory." *The Quarterly Journal of Economics*, 110, No. 3 (August 1995): 605-340.

Gjesdal, F. "Accounting for Stewardship." *Journal of Accounting Research*, 19, (Autumn 1981): 208-231.

Green, J. and N. Stokey. "Two Representations of Information Structures and Their Comparisons," working paper, Northwestern University, (1977).

Gul, Faruk. "A Comment on Aumann’s Bayesian View." *Econometrica*, 66, No. 4 (July, 1998):

923-927.

Halpern, J. "Characterizing the Common Prior Assumption." *Journal of Economic Theory*, 106, No. 2 (October 2002): 316-355.

Harsanyi, J. C. "Games with Incomplet Information Played by Bayesian Players." Parts I, II, and III. *Management Science*, 14, Nos. 3, 5 and 7, (November 1967 and January and March, 1968): 159-182, 320-334, and 485-502.

Hirst, D. E. and P. E. Hopkins. "Comprehensive Income Reporting and Analysts' Valuation Judgements." *Journal of Accounting Research*, 36 (Supplement 1998): 47-75.

Hirst, D. E., L. Koonce and J. Miller. "The Joint Effect of Management's Prior Forecast Accuracy and the Form of Its Financial Forecasts on Investor Judgment." *Journal of Accounting Research* 37 (Supplement 1999): 101-124.

Holmstrom, B. "Moral Hazard and Observability", *The Bell Journal of Economics*, (Spring 1979): 74-91.

Hong, L. and S. Page. "Problem Solving by Teams of Heterogeneous Agents." *Journal of Economic Theory*, 97, No. 1 (March 2001): 123-63.

Ijiri, Y. "The Theory of Accounting Measurement", Studies in Accounting Research #10, (1975), American Accounting Association, Sarasota, Florida.

Kennedy, J., T. Mitchell, and S. E. Sefcik. "Disclosure of Contingent Environmental Liabilities: Some Unintended Consequences." *Journal of Accounting Research*. 36 (Autumn 1998): 257-77.

Khinchin, A. I. Mathematical Foundation of Information Theory, (1957), Dover Publications, Inc., New York, New York.

Malueg, David A. "Comparing Comparisons of Experiments," working paper No 85-39, University of Iowa, (September 1985).

Marschak, J. and K. Miyasawa "Economic Comparability of Information Systems." Economic Information, Decision and Prediction, (1968), D. Reidel Pub. Co., Boston, Mass.

McGuire, C. B. "Comparisons of Information Structures," chapter 5 in Decision and Organization, C. B. McGuire and R. Radner, editors, (1972), University of Minnesota Press, Minneapolis.

Milgrom, P. "An Axiomatic Characterization of Common Knowledge," *Econometrica*, 49, No. 1, (January 1981): 219-222.

Morris, Stephen. "Trade with Heterogeneous Prior Beliefs and Asymmetric Information." *Econometrica*, 62, No. 6 (November, 1994): 1327-1347.

-----. "The Common Prior Assumption in Economic Theory." *Economics and Philosophy*, 11, (1995): 227-253.

Morris, S., and Hyun Song Shin. "The Social Value of Public Information." *American Economic Review*, 92 (2002): 1521-1534.

-----. “Heterogeneity and Uniqueness in Interaction Games” in The Economy as an Evolving Complex System III, edited by L. Blume and S. Durlauf, Sante Fe Institute Studies in the Sciences of Complexity. New York: Oxford University Press (2005).

Myerson, R. B. "Optimal Coordination Mechanisms in Generalized Principal-Agent Problems." *Journal of Mathematical Economics*, 10, (1982): 67-81.

Penno, M. "Accounting Systems, Participation in Budgeting, and Performance Evaluation." *The Accounting Review*, (April 1990): 303-314.

Penno, M., and J.S. Watts. "An Independent Auditor's Ex Post Criteria for the Disclosure of Information." *Journal of Accounting Research,* 29 (Supplement, 1991): 194-212.

Reiter, S. and S. Reichelstein. "Game Forms with Minimal Message Spaces." *Econometrica*, 56, No. 3, (May 1988): 661-692.

Sunder, S. "Knowing What Others Know: Common Knowledge, Accounting and Capital Markets," *Accounting Horizons*, 16, No. 4, (December 2002), 305-318.

Table 1: Notation

Ω: set of states of the world, with $\omega \in \Omega$ being a specific state of the world.

p: probability distribution defined over set of states of the world, so that $p : \Omega \to [0,1]$.

S: Information partition of player one, called the manager, having elements $s \in S$. Similarly $g \in G$ is the information partition of player 2, called the auditor. I write $S \leq \Omega$ and $G \leq \Omega$ to indicate that both partitions are coarsenings of the set of states of the world.

M: message set, with cardinality N, denoted as $|M| = N$, and composed of messages m_n for $n = 1,..., N$, so that $M = \{m_n\}_{1 \leq n \leq N}$.

σ: Signaling function for player 1 where $\sigma : S \to M$. An analogous signaling function exists for player 2 and is denoted as $\gamma : G \to M$.

χ: Language form, composed of the set of states of the world, probability distribution over this set, information partitions for the players involved, a message set and signaling functions for the players, so that $\chi = (\Omega, p, S, G, M, \sigma, \gamma)$.

X^M: Anchor set for message set M. Anchor sets are subsets of the set of states of the world, or $X^M \subset \Omega$, and have elements x_n^M for $n = 1,..., N$, so that $X^M = \{x_n^M\}_{1 \leq n \leq N}$

APPENDIX

Proof of Theorem 1: The proof utilizes a lemma that is first introduced and proved.

Lemma 1: A language-form, $\chi = (\Omega, p, S, G, M, \sigma, \gamma)$, is hard if $M \leq \Omega$.

Proof (of lemma 1): To show that χ is hard, we need to show $\forall \omega \in \Omega$ and $\forall m_n \in M$ that $\sigma(s(\omega)) = m_n$ if and only if $\gamma(g(\omega)) = m_n$. If $M \leq \Omega$, then $\forall \omega \in \Omega$ there exists a unique message, $m_n \in M$, such that $\omega \in m_n$. By truthful reporting, that is, by condition ii) in assumption A2, we have $\sigma(s(\omega)) = m_n$ if and only if $\omega \in m_n$ while $\gamma(g(\omega)) = m_n$ holds under the same conditions. Since this holds $\forall \omega \in \Omega$ and $\forall m_n \in M$, we have that χ is hard as required, completing the proof of lemma 1

Turning now to the proof of the theorem itself, we begin with sufficiency and follow with necessity. Let M' be defined as in the theorem 1. Since M' is a coarsening of the set of states of the world, or $M' \leq \Omega$, we have that $\chi' = (\Omega, p, S, G, M', \sigma', \gamma')$ is hard. Also, M isomorphic to M' means that $\forall \omega \in \Omega$ we have that $\sigma(s(\omega)) = m_n = \sigma'(s(\omega))$ and that $\gamma(g(\omega)) = m_n = \gamma'(g(\omega))$, which implies that $\chi = (\Omega, p, S, G, M, \sigma, \gamma)$ is also hard.

Let $\chi = (\Omega, p, S, G, M, \sigma, \gamma)$ be hard. Consider the message set $M' = S^M$, where $M' = \{\{s_n^M\}_{1 \leq n \leq N}\}$ and where $s_n^M = \{s \in \sigma^{-1}(m_n)\}$ holds $\forall m_n \in M$. From the functions $\gamma' : G \to M'$ and $\sigma' : S \to M'$ by setting $\gamma^{-1}(m_n) = \gamma'^{-1}(m'_n)$ and $\sigma^{-1}(m_n) = \sigma'^{-1}(m'_n)$ for $n = 1, \ldots, N$. Clearly $M' \leq \Omega$, so the language form, $\chi' = (\Omega, p, S, G, M', \sigma', \gamma')$ is also hard, completing the proof of theorem 1.

Proof of Corollary 1: The proof follows immediately from Theorem 1 once we note that the messages set defined in Theorem 1 insures that $M' = S^M = G^M$.

Proof of Theorem 2: The proof utilizes a lemma that is first introduced and proved.

Lemma 2: Let $\chi = (\Omega, p, S, G, M, \sigma, \gamma)$ and $\chi' = (\Omega, p, S, G, M, \sigma', \gamma)$ be two language forms that have the same anchor set, $\{x_n^M\}_{1 \le n \le N} = X^M \subset \Omega$ where χ is harder than χ' . This means that, for the elements of the inverse image sets determined by each pair of anchors, $x_n^M, x_k^M \in X^M$ the following hold:

a. $p(s_n^M \mid g_n^M) \ge p(s'^M_n \mid g_n^M) \ge p(s'^M_k \mid g_n^M)$, and

b. $p(s_k^M \mid g_n^M) \le p(s'^M_k \mid g_n^M)$,

with at least one inequality strict. χ is harder than χ' implies that there exists a $N \times N$ Markov matrix, B, having elements b_{jk} where $\forall g_n^M \in G^M$, $\forall s_j^M \in S^M$ and $\forall s'^M_k \in S'^M$ the following holds: $p(s'^M_k \mid g_n^M) = \sum_{j=1}^{N} \left(p(s_j^M \mid g_n^M) \times b_{jk} \right)$.

Proof of Lemma 2: Using the assumed existence of an anchor set, we will show that this together with the definition of relative hardness suffice to imply there exists a markov matrix B as described in the lemma. The proof is by construction, and can be informally explained as follows. First I show that B can be written in terms of the elements of the two conditional probability matrices. Second, I show that the elements in the columns of the B matrix sum to one. Third, I show that the elements of B are between 0 and 1. This final step is the complicated one. I use an algorithm where I start by writing each of the equations so that all have zero on the right-hand side but one. The first element is negative but the remaining elements of the B matrix may be either positive or negative. Then I zero out all of the off diagonal elements of B, except for the first element, until only the diagonal elements are left and the first column are left. In this manner, I show that all the elements are positive but less than 1, and this completes the proof.

To simplify the notation, without loss of generalization, let P and P' denote the matrices of conditional distributions, where P is harder than P'. As is usual, denote the nth row and mth column of the P matrix as p_{nm}, and similarly for the P' matrix, so that,

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1N} \\ p_{21} & p_{22} & \cdots & p_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ p_{N1} & p_{N2} & \cdots & p_{NN} \end{bmatrix} \equiv \begin{bmatrix} p\left(s_1^M \mid g_1^M\right) & p\left(s_2^M \mid g_1^M\right) & \cdots & p\left(s_N^M \mid g_1^M\right) \\ p\left(s_1^M \mid x_2\right) & p\left(s_2^M \mid g_2^M\right) & \cdots & p\left(s_N^M \mid g_2^M\right) \\ \vdots & \vdots & \ddots & \vdots \\ p\left(s_1^M \mid x_N\right) & p\left(s_2^M \mid g_N^M\right) & \cdots & p\left(s_N^M \mid g_N^M\right) \end{bmatrix}$$

and

$$P' = \begin{bmatrix} p'_{11} & p'_{12} & \cdots & p'_{1N} \\ p'_{21} & p'_{22} & \cdots & p'_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ p'_{N1} & p'_{N2} & \cdots & p'_{NN} \end{bmatrix} \equiv \begin{bmatrix} p\left(s'^M_1 \mid g_1^M\right) & p\left(s'^M_2 \mid g_1^M\right) & \cdots & p\left(s'^M_N \mid g_1^M\right) \\ p\left(s'^M_1 \mid x_2\right) & p\left(s'^M_2 \mid g_2^M\right) & \cdots & p\left(s'^M_N \mid g_2^M\right) \\ \vdots & \vdots & \ddots & \vdots \\ p\left(s'^M_1 \mid x_N\right) & p\left(s'^M_2 \mid g_N^M\right) & \cdots & p\left(s'^M_N \mid g_N^M\right) \end{bmatrix}.$$

As is standard, the row element indicates the conditioning variable and the column element indicates the realization of the random variable, so for example, $p_{nm} = p\left(s_m^M \mid g_n^M\right)$. This means that for each row in P and P', the sum of the row elements equals one. If B exists, we have

$$P \times B = P'.$$

To show B is markov, we need to show that such a B exists where each element is non-negative and the rows sum to one, or more specifically, where the following two conditions hold,

App.i) $\sum_{k=1}^{N} b_{jk} = 1$ for each j, and

App.ii) $0 \le b_{jk} \le 1$ for each b_{jk}.

As above, the b_{jk} element is in the jth row and kth column of the B matrix.

First, to show that such a matrix exists, it suffices to show that a solution exists for the $N \times N$ linear equations of the form

$$p_{n1}b_{1k} + p_{n2}b_{2k} + \cdots + p_{nN}b_{Nk} = p'_{nk}.$$

By definition of relative hardness, there exist $N \times N$ numbers, ε_{nk}, each less than 1 in absolute value, defined as $p_{nk} = p'_{nk} - \varepsilon_{nk}$, where $\varepsilon_{nn} \ge 0$, with at least one inequality strict, where $\varepsilon_{nk} \le 0$ if $k \ne n$, again, with at least on inequality strict, and $\forall j$, $\sum_{k=1}^{N} \varepsilon_{nk} = 0$. Further, by definition, $\forall k, n$ we have $p_{nn} \ge p_{nk}$ and $p'_{nn} \ge p'_{nk}$, so that in particular, $p_{nn} > 0$. This holds $\forall n$, and in particular for $n = j$. Substituting back into the equation, we have

$$p_{j1}b_{1k} + p_{j2}b_{2k} + \cdots + p_{jN}b_{Nk} = p'_{jk} = p_{jk} - \varepsilon_{jk}.$$

Hence, we can write the element b_{jk} as follows:

$$b_{jk} = \frac{p_{jk} - \varepsilon_{jk} - \sum_{m \ne j}\left(p_{jm}b_{mk}\right)}{p_{jj}}.$$

Since $p_{jj} > 0$, this proves the B matrix exists.

For condition App.i), i.e. to show $\sum_{k=1}^{N} b_{jk} = 1$ for each j, note that we have $j = 1,\dots,N$ equations where $p'_{jk} - p_{jk} = -\varepsilon_{jk}$. Substituting for p'_{jk} from above, each of these equations can be written as follows:

$$
\begin{aligned}
&p_{j1}(b_{11}-1)+p_{j2}b_{21}+\cdots+p_{jN}b_{N1}=-\varepsilon_{j1}\\
&p_{j1}b_{12}+p_{j2}(b_{22}-1)+\cdots+p_{jN}b_{N2}=-\varepsilon_{j2}\\
&\vdots\\
&p_{j1}b_{1N}+p_{j2}b_{2N}+\cdots+p_{jN}(b_{NN}-1)=-\varepsilon_{jN}
\end{aligned}
$$

Summing these equations, we get

$$p_{j1}\left(\sum_{k=1}^{N}b_{1k}-1\right)+p_{j2}\left(\sum_{k=1}^{N}b_{2k}-1\right)+\cdots+p_{jN}\left(\sum_{k=1}^{N}b_{Nk}-1\right)=-\sum_{k=1}^{N}\varepsilon_{j1}=0\,.$$

There are N such equalities, with $p_{jj}>0$ for the jth equation, so that the system of equations is solved by $\sum_{k=1}^{N}b_{jk}=1$ for all j.

For condition App.ii), i.e. to show $0\le b_{jk}\le 1$ for each b_{jk}. First, note that the product of two markov matrices is itself markov. This means that we can focus on the case where P and P' differ only by two elements. Again, without loss of generalization, let $p_{11}>p'_{11}$, $p_{12}<p'_{12}$, and $p_{nj}=p'_{nj}$ for all other elements of the matrices (i.e. for all $n>1$ or $j>2$). Since $p_{nj}=p'_{nj}$ for $n>1$ or $j>2$, the b_{jk} elements of the B matrix for $j>1$ or $k>2$ are given as $b_{jk}=0$ for $j\neq k>2$ and $b_{jj}=1$ for $j>2$. More specifically, the B matrix is given as follows:

$$B=\begin{bmatrix} b_{11} & b_{12} & 0 & \cdots & 0\\ b_{21} & b_{22} & 0 & \cdots & 0\\ b_{31} & b_{21} & 1 & \cdots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ b_{N1} & b_{N2} & 0 & 0 & 0 \end{bmatrix}$$

Hence, we need show condition (ii) holds only for the N elements in the first two columns of B, (i.e. the b_{jk} elements with $k=1$ or $k=1$), or equivalently, to solve the equations for the first two columns in the P' matrix.

To see this, consider the equations for the first column in the P' matrix, denoted as P'^{C}_{1}. The value for each element in this row is given by the following matrix product:

$$P'^{C}_{1} = \begin{bmatrix} p'_{11} \\ p'_{21} \\ \vdots \\ p'_{N1} \end{bmatrix} = P \times B^{C}_{1} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1N} \\ p_{21} & p_{22} & \cdots & p_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ p_{N1} & p_{N2} & \cdots & p_{NN} \end{bmatrix} \times \begin{bmatrix} b_{11} \\ b_{21} \\ \vdots \\ b_{N1} \end{bmatrix}.$$

This produces the following system of N equations.

$$\begin{aligned} & p_{11}b_{11} + p_{12}b_{21} + \cdots + p_{1N}b_{N1} = p'_{11} = p_{11} - \varepsilon_{11} \\ & p_{21}b_{11} + p_{22}b_{21} + \cdots + p_{2N}b_{N1} = p'_{21} = p_{21} \\ & \vdots \\ & p_{N1}b_{11} + p_{N2}b_{21} + \cdots + p_{NN}b_{N1} = p'_{N1} = p_{N1} \end{aligned}$$

Since the elements of the P and P' matrices are given, this is a system of N equations in the b_{jk} unknowns, with $k=1$ and $j=1,...,N$. I develop an algorithm that can be applied to the N elements b_{j2}, $j=1,...,N$, to show that condition App.ii) also holds for these elements as well.

Before proceeding with the algorithm, I rearrange the above system of equations to provide a basis for the algorithm. Subtracting p_{n1} from both sides of the n$^{\text{th}}$ equation, the N equations shown above can be rewritten as the following equations denoted [A.1] through [A.N].

$$\begin{array}{ll} [A.1]: & -p_{11}(1-b_{11}) + p_{12}b_{21} + \cdots + p_{1N}b_{N1} = -\varepsilon_{11} \\ [A.2]: & -p_{21}(1-b_{11}) + p_{22}b_{21} + \cdots + p_{2N}b_{N1} = 0 \\ ([A.3]-[A.N-1]): & \vdots \\ [A.N]: & -p_{N1}(1-b_{11}) + p_{N2}b_{21} + \cdots + p_{NN}b_{N1} = 0 \end{array} .$$

To show condition App.ii) holds, I identify a series of manipulations to the system of equations that adds or subtracts a fraction of one equation to a second equation while ensuring no sign change on the coefficients of the b_{j1} variables. Before we begin, insure the coefficients on each b_{j1} variable, $j=2,...,N$ are non-zero, and then start with the equations having a zero coefficient on the b_{11}

variable (we assumed only A.N meets this condition). The process is given in several steps below.

Step 1: In this step, I add multiples of the different equations to insure that the probabilities in each column are positive above some specified row, and all the diagonal values equal 1. At the end of this step, I will have created a new probability matrix, $P*$, with the following form:

$$P* = \begin{bmatrix} 1 & p*_{12} & p*_{13} & 0 & p*_{15} & \cdots & p*_{1N} \\ p*_{21} & 1 & p*_{23} & 0 & p*_{25} & \cdots & p*_{2N} \\ p*_{31} & p*_{32} & 1 & 0 & p*_{35} & \cdots & p*_{3N} \\ p*_{41} & p*_{42} & p*_{43} & 1 & 0 & \cdots & p*_{4N} \\ p*_{51} & 0 & p*_{53} & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ p*_{N1} & 0 & p*_{N3} & 0 & 0 & \cdots & 1 \end{bmatrix}$$

Find the equation n that has the greatest number of p_{nj} probabilities equal to zero and make this equation A.N. Starting with equation A.N-1, re-order the equations so that the ascending equations all have zeroes only for the same j entries for which A.N has a zero probability. For any equation that has a zero for a p_{nj} probability that is non-zero in A.N, replace this with a new equation having probabilities denoted as $p*_{nj}$ that are non-zero, but where the $p*_{nn}$ probabilities are still maximal. For example, suppose equation A.2 has $p_{23} = 0$. Multiply A.N by $e > 0$ and add this to equation A.2 to give a new set of probabilities, but choose $e > 0$ sufficiently small to insure that $p*_{22}$ is still maximal over all the new $p*_{2j}$ probabilities. That such an $e > 0$ exists can be seen as follows: identify all m such that $(p_{Nm} - p_{N3}) > 0$. This set is non-empty, since $(p_{NN} - p_{N3}) > 0$. Next choose $e > 0$ such that $\frac{(p_{22} - p_{2m})}{(p_{Nm} - p_{N3})} > e$; this insures that $p*_{22} = p_{22} + p_{N3}e$ is maximal. For the final part of this first step, for each equation A.n, $2 \le n \le N$, divide this equation by p_{nn}. We now have a system of equations similar to the following system.

$$
\begin{array}{ll}
[A.1]: & -p_{11}(1-b_{11})+p_{12}b_{21}+\cdots+p_{1n}b_{n1}+\cdots+p_{1N}b_{N1}=-\varepsilon_{11} \\
[A.2]: & -p'_{21}(1-b_{11})+b_{21}+\cdots+p'_{2n}b_{n1}+\cdots+p'_{2N}b_{N1}=0 \\
([A.3]-[A.n-1]): & \vdots \\
[A.n]: & -p'_{n1}(1-b_{11})+p'_{n2}b_{21}+\cdots+b_{n1}+\cdots+p'_{nN}b_{N1}=0 \\
([A.n+1]-[A.N-2]): & \vdots \\
[A.N-1]: & -p'_{(N-1)1}(1-b_{11})+0+p'_{(N-1)3}b_{31}+p'_{(N-1)4}b_{41}+\cdots+p'_{(N-1)N}b_{N1}=0 \\
[A.N]: & -p'_{N1}(1-b_{11})+0+p'_{N3}b_{31}+0+\cdots+b_{N1}=0
\end{array}
$$

In particular, for $n>1$, the coefficient on the b_{n1} variables in equation A.n is 1, while the coefficient on every other b_{n1} variable is positive but less than 1, except for the coefficients on the b_{n1} variables where the coefficient is zero in equation A.N.

Step 2: Steps 2, 3 and 4 together will change the $P*$ matrix into and an identity matrix. I do this by zeroing out all the off-diagonal probabilities, starting with those in the last column. Starting with equation A.N-1, determine if $p*_{(N-1)N}$ is the smallest of the $p*_{(N-1)j}$ probabilities for $j>1$. If so, zero it out by multiplying A.N by $p*_{(N-1)N}$ and subtracting the resulting equation from A.N-1; then proceed to equation A.N-2 and repeat step 2 for that equation. Continue to zero out probabilities on b_{N1} in this manner until all are zeroed out, in which case proceed to Step 4, or until a non-minimal coefficient is found, in which case, proceed to Step 3.

Step 3: Suppose the coefficient on b_{N1} is minimal for each equation A.N-1 up to and including A.n+1, but not on A.n. Find the smallest probability in equation A.n; suppose it is $p*_{nj}$. Then multiply A.N by $p*_{nj}$ and subtract the resulting equation from A.n. All the coefficients remain positive, since we multiplied by the smallest probability in A.n, and the probability on b_{n1}, call it $p**_{nn}$ remains maximal. Divide A.n through by this new probability $p**_{nn}$, so that the coefficient on b_{n1} is now 1. Continue this process until the coefficient on b_{N1} is minimal, and then repeat step 2 for A.n.

Step 4: Having zeroed out the coefficient on b_{N1} for each of the equations A.1 through A.N-1, next turn to the coefficient on $b_{(N-1)1}$ for the equations A.N and A.N-2 through A.1. Using equation A.N-1, perform Steps 2 and 3 in order to zero out these coefficients in the same manner as was done for the coefficients on b_{N1}. Repeat these steps until all the off-diagonal coefficients have been zeroed out on variables b_{j1} for j > 1. Note: the right-hand side of equations A.2 through A.N is still zero to this point. Finally, zero out the probabilities in the first column, those multiplied by $(1-b_{11})$, in equations A.2 through A.N by multiplying equation A.1 by the coefficient on $(1-b_{11})$ and subtracting from each equation A.2 through A.N.

After completing Steps 1 through 4, we are left with N equations having zero coefficients for all off-diagonal elements. The right-hand side of equations A.2 through A.N is either positive or zero depending on whether the coefficient on $(1-b_{11})$ in the original equation was non-zero or zero. The right-hand side of A.1 is negative, but so is the coefficient on $(1-b_{11})$, so that cross-multiplying and solving for b_{11} proves that b_{11} is also positive. As shown above, the sum of the on b_{j1} variables equals 1, so all the b_{j1} variables are between 0 and 1, completing the proof that condition App.ii) holds. This completes the proof of Lemma 2.

Proof of Theorem 2: Using Lemma 2, we have that P' can be formed from P by multiplication by a Markov matrix. Then, applying the sufficiency portion of Blackwell's theorem, we have information structure χ is statistically sufficient for χ', so that for any expected utility maximizing player, conditioning on χ provides a higher expected utility than conditioning on χ'. Hence χ is preferred to χ' by this player, as stated in the theorem. This completes the proof of theorem 2.

Proof of Theorem 3: First, suppose language-form $\chi = (\Omega, p, S, G, M, \sigma, \gamma)$ is hard. Corollary 1.1 tells us that $p(s_k^M \mid g_n^M) = 0$ if $n \neq k$ and $p(s_k^M \mid g_n^M) = 1$ if $n = k$. Since $\forall n$ we have $\mathbf{ln}(p(s_n^M \mid g_n^M)) = \mathbf{ln}(1) = 0$, $H(S^M \mid G^M) = 0$ follows immediately. If $H(S^M \mid G^M) = 0$, then $\forall n$, there exists a single $k = k*$ such that $p(s_{k*}^M \mid g_n^M) = 1$, while for $k \neq k*$, $p(s_k^M \mid g_n^M) = 0$ holds. However, the anchor set ensures, $p(s_n^M \mid g_n^M) > 0$, which implies $k* = n$, implying that χ is hard.

Next, let $\chi' = (\Omega, p, S, G, M, \sigma', \gamma)$ be a second language form and suppose that χ is relatively harder than χ'. From Lemma 2, we know that we can write the conditional probability matrices as $P \times B = P'$, where B is a Markov matrix and P and P' are conditional probability matrices for the language form χ and χ', respectively. Without loss of generality, suppose P and P' differ only on $n = 1$, so that $\forall k$, and for $n > 1$ we have $p(s_k^M \mid g_n^M) = p(s'^M_k \mid g_n^M)$. This means that to show $0 \le H(S^M \mid G^M) < H(S'^M \mid G^M)$, it suffices to show $H(S^M \mid g_1^M) < H(S'^M \mid g_1^M)$ holds. By definition, we have $H(S^M \mid g_1^M) \equiv H(P_1^R)$ and $H(S'^M \mid g_1^M) \equiv H(P'^R_1)$, where P_1^R and P'^R_1 are the first row matrices in the matrices P and P', respectively. Also, from above we have that $P_1^R \times B = P'^R_1$, where B is independent of P_1^R. I next use two properties of the entropy measure. First, entropy is non-negative. Second, for two independent probability distributions A_1 and A_2, $H(A_1 \times A_2) = H(A_1) + H(A_2)$ (see Khinchin (1957), equation 2, page 5). Together these imply the following

$$H\left(P'^{R}_{1}\right) = H\left(P_1^R \times B\right) = H\left(P_1^R\right) + H(B) > H\left(P_1^R\right).$$

This proves that χ harder than χ' implies $0 \le H\left(S^M \mid G^M\right) < H\left(S'^M \mid G^M\right)$.

Next, suppose $0 \le H\left(S^M \mid G^M\right) < H\left(S'^M \mid G^M\right)$ holds. By assumption, χ and χ' can be compared based on hardness, so that, by Lemma 2, we know there exists a Markov matrix B where either $P \times B = P'$ or $P' \times B = P$. The second can be shown not to hold by contradiction. Suppose $P' \times B = P$ holds. This implies $H(P) = H(P' \times B) = H(P') + H(B) > H(P')$, which in turn implies $H\left(S^M \mid G^M\right) > H\left(S'^M \mid G^M\right)$, providing the contradiction. Hence $P \times B = P'$ holds, implying χ is harder than χ', which completes the proof of Theorem 3.